Department of Economics- FEA/USP

Welfare-improving misreported polls

Felipe R. Durazzo
David Turchick

Working Paper Series Nº 2020-10

DEPARTMENT OF ECONOMICS, FEA-USP
WORKING PAPER Nº 2020-10

Welfare-improving misreported polls

Felipe R. Durazzo (felipedurazzo2019@u.northwestern.edu)

David Turchick (dturchick@usp.br)

Abstract:

An often-heard criticism about electoral pollsters is that they might misreport pre-election poll results. We show that this can happen even in the absence of partisan motives, but purely for reputational ones. By underreporting the expected number of supporters of the most preferred candidate, the pollster is able to induce an election result more in line with its report. By doing so, not only victory chances of the most preferred candidate in society rise above 50%, but also total election costs are reduced, thus yielding welfare gains. Our model also allows for the accommodation of both the underdog effect (a feature of pivotal voting models) and the apparently inconsistent bandwagon effect, in the sense that the latter may be an illusion on the part of an observer who disregards the possibility of nontruthful polls. All of these results hold even as the electorate size grows without bound.

Keywords: costly voting, pivotal voting model, pre-election polls, misreporting, bandwagon effect.

JEL Codes: C70, C72, D72, C46.

Welfare-improving misreported polls

Felipe R. Durazzo [a,*], David Turchick [b,†]

[a] University of São Paulo / Northwestern University

[b] University of São Paulo

Abstract

An often-heard criticism about electoral pollsters is that they might misreport pre-election poll results. We show that this can happen even in the absence of partisan motives, but purely for reputational ones. By underreporting the expected number of supporters of the most preferred candidate, the pollster is able to induce an election result more in line with its report. By doing so, not only victory chances of the most preferred candidate in society rise above 50%, but also total election costs are reduced, thus yielding welfare gains. Our model also allows for the accommodation of both the underdog effect (a feature of pivotal voting models) and the apparently inconsistent bandwagon effect, in the sense that the latter may be an illusion on the part of an observer who disregards the possibility of nontruthful polls. All of these results hold even as the electorate size grows without bound.

JEL classification: C70; C72; D72; C46.

Keywords: Costly voting. Pivotal voting model. Pre-election polls. Misreporting. Bandwagon effect.

Acknowledgement 1 *We are grateful to Filipe Campante, Fernanda Estevan, Elisabeti Kira, Daniel Monte and Marcos Nakaguma for their invaluable comments and suggestions. Felipe Durazzo gratefully acknowledges FAPESP for financial support (grants 2016/18987-4 and 2017/12285-0).*

*E-mail address: felipedurazzo2019@u.northwestern.edu.

†E-mail address: dturchick@usp.br.

1 Introduction

The canonical pivotal voting model with two candidates, as in Palfrey and Rosenthal (1985), Ledyard (1984) and, more recently, Börgers (2004), assumes that a citizen participates in an election insofar as there exists a possibility that his/her vote is pivotal, that is, it changes the result of the election. To compute the probability that his or her vote is pivotal, a constituent needs to know the distribution of preferences for each candidate within society. In modern democratic societies, this information is provided by electoral pollsters. Empirical evidence suggests that pre-election polls can affect voter turnout and, ultimately, election results (see, e.g., Bursztyn et al., 2018).

However, little is known about the incentives that pollsters have regarding the release of pre-election poll results. In other words, can we be certain that electoral pollsters are not manipulating poll results? Moreover, could such a strategic behavior harm society?

We remark the importance of these questions by noting that, even if society finds it desirable, it is impossible to fully restrain pollsters from potentially misreporting information. Pollsters can misreport information in several ways, such as rigging the sampling procedure, framing questions in a biased way,[1] marking answers incorrectly or simply discarding part of the sample.

However, the behavior of a pollster might be restrained by reputational concerns: people compare the result of the election with what they were expecting given the information released by the pollster. If the election result is significantly different from what the information provided by the pollster implied it would be, the quality of the pollster might be questioned. This happened, for example, in the infamous "Dewey Defeats Truman" event[2] and after the 2016 U.S. presidential election, when most pollsters predicted a Clinton victory over Trump.[3]

Indeed, the reputation of pollsters varies considerably. For instance, ABC News' project

[1] Bischoping and Schuman (1992) provides experimental evidence that even the pen used by the interviewer to record answers can alter the answers of the respondent.

[2] See http://articles.latimes.com/1998/nov/01/news/mn-38174.

[3] See, for instance, http://www.nytimes.com/2017/05/31/upshot/a-2016-review-why-key-state-polls-were-wrong-about-trump.html.

FiveThirtyEight rates pollsters by "analyzing the historical accuracy of each firm's polls along with its methodology."[4] This rating is then translated into a grade, from A+ to F. Pollsters are graded F if they raise suspicions of "faked polling results."

It would only be natural to assume that better ranked polling firms are able to charge a higher price for their services, whereas lower ranked firms, due to their perceived inaccuracy, will have their market share and future profits hindered (even if not prosecuted and convicted of fraud). Anecdotal evidence comes from Brazil, where polling firms are required by law to register their price. For instance, for the presidential elections of 2018, the average price charged by all 27 pollsters per respondent was 32 reals, while the standard deviation of the price charged was 35 reals. The average price of the five most expensive pollsters was 98 reals. To some degree, this large dispersion in the prices suggests that the perceived quality and reputation of a pollster directly impacts how much a pollster is able to charge.[5]

The effects of pre-election polls on voter behavior have been studied in Taylor and Yildirim (2010a) and Goeree and Großer (2007). Both papers, using a two-candidate costly voting model, have concluded that opinion polls may be harmful to the citizens' expected welfare, for stimulating the "wrong" group of citizens (that is, the group who most likely is the minority) to vote more, while, at the same time, stimulating the group who most likely is the majority to vote less. Moreover, the first effect is stronger than the second one. Therefore, pre-election poll results have two negative consequences: they increase the expected aggregate voting cost and decrease the probability that the most preferred candidate wins the election. In their models, however, poll results are reported truthfully.

We develop a model in which, by formally introducing an electoral pollster, the poll report is endogenously determined and is not necessarily truthful. More precisely, we suppose that the pollster knows the true distribution of the citizens' preferences between the two candidates, but it may report to the public a different distribution. The pollster does not have preferences

[4] http://projects.fivethirtyeight.com/pollster-ratings.

[5] This data is available at http://www.tse.jus.br/eleicoes/pesquisa-eleitorais/consulta-as-pesquisas-registradas.

among the candidates: its objective is to be highly rated, by "getting the election right." Thus we seek to understand how an electoral pollster concerned with its reputation might report the pre-election poll results, given that it knows that they may alter voters' behavior and consequently the election result, upon which the reputation of the pollster is based.

An often-heard criticism about pollsters is that they might misreport to benefit some candidate. Here we show that they have incentives to misreport even in the absence of partisan motives, but in the presence of reputational motives alone. Our main findings are that, unless the expected number of supporters of each candidate is the same, i) for a sufficiently large population, truthful reports do not happen, ii) the pollster always underreports the expected number of supporters of the most preferred candidate, and iii) this misreporting of the pre-election poll results is welfare improving relative to a truthful reporting. The first two of these results imply that, relative to a truthful report, with a misreported poll, citizens in the expected majority group will vote with more intensity, the opposite being true for citizens in the expected minority group. Thus, contrary to what the pivotal voting model with truthful polls would predict, election results are not ties, in expected terms. This is the intuition behind the third of these findings.

Once reputational concerns are brought into the picture, other conclusions of the canonical pivotal voting model can also be revisited. A classical prediction of rational voting models is the underdog effect, where citizens in the minority group vote with more intensity than citizens in the majority group, as the latter have an incentive to free ride (e.g., Ledyard, 1984, and Palfrey and Rosenthal, 1983). This result is present in and is important for the conclusions of Taylor and Yildirim (2010a), Goeree and Großer (2007) and Krasa and Polborn (2008).

An alternative prediction would be the presence of a bandwagon effect, where constituents vote with a higher intensity if they realize they are in the majority.[6] Grillo (2017) shows that it is possible to generate a bandwagon effect in a rational voting model if citizens are assumed to

[6]Besides the mentioned "bandwagon abstention effect," a "bandwagon vote choice effect" could also be considered: some citizens may switch their preference to the candidate that ranks first in the pre-election polls (see Morton and Ou, 2015). In our model, however, this is not a possibility, since the preference of a given citizen regarding the two candidates is held fixed.

be risk averse.[7] In contrast, our work implies that, even without departing from the ubiquitous risk-neutrality assumption, an observer who assumes that polls are always truthful would see data consistent with the bandwagon effect, even though the voters' behavior generates the underdog effect. Thus, we suggest that part of what is usually attributed to the bandwagon effect could actually be an illusion due to misreporting, and not an inconsistency between the classical model and the data.

An unrealistic prediction of rational voting models with fixed and homogenous voting costs is the neutrality result, namely, that both candidates should have equal chances of winning the election in equilibrium, regardless of the expected number of supporters of each candidate. Taylor and Yildirim (2010b) shows that, in small elections, if voting costs are not fixed, but rather independently drawn from a common distribution among citizens, the neutrality result disappears in favor of the candidate with the support of the majority group. However, for large elections (with the number of citizens tending to infinity), they show that a necessary condition for the nonoccurrence of the neutrality result is heterogeneity of the voting cost distributions between the types of constituents. In our model, with fixed (and equal) voting costs, the neutrality result disappears even in large elections, that is, the candidate supported by the majority group wins the election with probability greater than 50%.

After presenting the model in section 2, in section 3 we will characterize the equilibria of the electoral game played by the citizens given the information released by the pollster. Our characterization of electoral equilibria is slightly more general than that previously found in the literature, in that it includes also noninterior type-symmetric equilibria (basically, those in which the supporters of a particular candidate decide on a 0%, or a 100%, probability of casting their votes). Notwithstanding the possible switch between one and another equilibrium type, not only their continuity but also their smoothness is proved, thus allowing us to tackle the relevant comparative statics problems. Since the analysis in the subsequent sections hinges upon the result of the election, the distribution of the difference in the number of votes cast for

[7] Börgers (2004), Goeree and Großer (2007), Krasa and Polborn (2008) and Taylor and Yildirim (2010a and 2010b), for instance, all implicitly assume risk neutrality.

both candidates is also provided (we refer to it as the multinomial difference distribution).

Section 4 presents the utility function of the pollster and characterizes the solution of its optimization problem. Bounds for this solution are also offered, as well as comparative statics. An asymptotic result on the pollster's behavior, alongside the limiting distribution of the difference in the number of votes (a Skellam, or Poisson difference, distribution), are also provided, since they will be key in the welfare analysis of section 5.

In section 5, the welfare comparison between misreported and truthfully reported poll results is done. Our approach differs from other welfare analyses present in the literature, in the sense that it considers an approximate welfare function (as the one in Taylor and Yildirim, 2010a) not to suggest a result about the exact welfare function, but to actually prove such a result (with the aid of several other results from the theory of discrete probability distributions). We show that, contrary to common belief, misreporting actually increases the expected welfare of citizens, relative to a truthful report. This holds even if the electorate size grows without bound. Section 6 concludes.

2 Model environment

The main elements of the model are described below.

There are $n \geq 2$ constituents (to be referred to simply as citizens), two candidates (the Blue party candidate B and the Red party candidate R) and an electoral pollster. The voting cost $c \in \mathbb{R}_+$ is homogenous among citizens. A citizen gains 1 unit of utility if his/her most preferred candidate wins the election, and loses 1 unit of utility if the other candidate is the winner. An election tie is broken by the toss of a fair coin. Voting is voluntary.

The probability that a citizen favors candidate B (or R) is q $(1-q)$. We assume that $q \in [\bar{q}, 1-\bar{q}]$, where $\bar{q} \in (0, 0.5)$.[8] Only the electoral pollster knows the true probability q. The pollster declares to the citizens that the probability that a citizen favors candidate B is p.

[8] One can take $\bar{q}$ arbitrarily close to 0.

Since citizens do not know q, they will use p as an estimate for this parameter in their voting decisions. We say the pollster is misreporting if $p \neq q$.

The citizens are instrumental voters. A citizen will vote if $\Pi \times$ benefit $> c$ and will not vote if $\Pi \times$ benefit $< c$, where Π is the probability of being pivotal in the election and "benefit" represents the benefit associated with being pivotal.

A voter is pivotal if and only if his/her vote creates or breaks a tie. In both situations, the expected increase in utility for voting is 1, therefore benefit $= 1$.[9]

Given the probability p reported by the pollster, the citizens play among themselves a Bayesian Game. We focus on type-symmetric Bayesian Nash equilibria, in which the strategies played by the citizens are homogenous within types (B-citizens or R-citizens) but might differ between these types.

From the point of view of a B-citizen, given that all of the other B-citizens are voting with probability γ and all R-citizens are voting with probability δ, the probability that his/her vote is pivotal is:

$$\Pi_B(n,p,\gamma,\delta) = \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} \binom{n-1}{k,k,n-1-2k} (p\gamma)^k ((1-p)\delta)^k (1-p\gamma-(1-p)\delta)^{n-1-2k} \quad (1)$$
$$+ \sum_{k=0}^{\lfloor \frac{n-2}{2} \rfloor} \binom{n-1}{k,k+1,n-2-2k} (p\gamma)^k ((1-p)\delta)^{k+1} (1-p\gamma-(1-p)\delta)^{n-2-2k},$$

where the first summation refers to the probability of breaking a tie and the second summation refers to the probability of creating a tie. An analogous expression holds for an R-citizen:

$$\Pi_R(n,p,\gamma,\delta) = \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} \binom{n-1}{k,k,n-1-2k} ((1-p)\delta)^k (p\gamma)^k (1-(1-p)\delta-p\gamma)^{n-1-2k} \quad (2)$$
$$+ \sum_{k=0}^{\lfloor \frac{n-2}{2} \rfloor} \binom{n-1}{k,k+1,n-2-2k} ((1-p)\delta)^k (p\gamma)^{k+1} (1-(1-p)\delta-p\gamma)^{n-2-2k}.$$

[9] An implicit assumption is that citizens are risk neutral. Grillo (2017) allows for risk-averse citizens.

An important feature of the model is that the pollster is not ideological: by assumption, its goal is only to "get the election right." That is, the electoral result must, in some sense, be coherent, given the announced probability p. Bear in mind that this is not trivial because the voters' equilibrium is a function of p. Since the pollster's objective function will be known only to itself (it does not play any sort of information game with the voters in our setup), its presentation and discussion will be postponed until section 4.

In summary, the timing of the model is as follows:

- Nature chooses the type of each citizen;
- The pollster discovers the probability q (probability that a given citizen favors candidate B);
- The pollster reports $p \in (0, 1)$;
- Each citizen takes p as given and chooses to vote or to abstain;
- The election happens;
- Each citizen and the pollster receive their payoffs.

3 Electoral equilibrium

This section has two main goals. The first one is to characterize the Bayesian-Nash equilibrium of the voting game given the pre-election poll report p, which is done in Proposition 1, and present the corresponding comparative statics results, done in Proposition 2.

The second one is to show (Lemma 3) that, at least for a sufficiently large population size, the only relevant equilibrium type will be the interior one. Lemma 3 is a key result in this study since it serves as a starting point for the derivation of our main results in the next two sections (Propositions 4 and 7), as well as their asymptotic counterparts later on.

In order to characterize the electoral equilibrium, it will prove useful to define

$$\underline{c}_n(p) = \begin{cases} \Pi_B(n,p,1,1) & \text{if} \quad p \in [0.5,1) \\ \Pi_R(n,p,1,1) & \text{if} \quad p \in (0,0.5) \end{cases} \tag{3}$$

and

$$\overline{c}_n(p) = \begin{cases} \Pi_B\left(n,p,\frac{1-p}{p},1\right) & \text{if} \quad p \in [0.5,1) \\ \Pi_R\left(n,p,1,\frac{p}{1-p}\right) & \text{if} \quad p \in (0,0.5) \end{cases}. \tag{4}$$

In appendix A, it is shown that, unless $p = 0.5$ (in which case $\underline{c}_n(p) = \overline{c}_n(p)$), we have $\underline{c}_n(p) < \overline{c}_n(p)$. Other properties of these two functions are also shown, such as their symmetry relationship: $\underline{c}_n(p) = \underline{c}_n(1-p)$ and $\overline{c}_n(p) = \overline{c}_n(1-p)$ for all $p \in (0,1)$.[10]

The complete characterization of the electoral equilibrium is given in the following proposition.

Proposition 1 *Given $n \geq 2$, $c \in \mathbb{R}_+$ and $p \in (0,1)$, there is one and only one type-symmetric electoral equilibrium (γ,δ). If $c \leq \underline{c}_n(p)$, then $(\gamma,\delta) = (1,1)$. If $c \geq 1$, then $(\gamma,\delta) = (0,0)$. If $c \in (\overline{c}_n(p),1)$, then $(\gamma,\delta) \in (0,1)^2$, $\delta = (p/(1-p))\gamma$ and $\gamma \in (0,\min(1,(1-p)/p))$ solves $\Pi_B(n,p,\gamma,(p/(1-p))\gamma) = c$. For $p \geq 0.5$, if $c \in (\underline{c}_n(p),\overline{c}_n(p)]$, then $\delta = 1$ and $\gamma \in [(1-p)/p,1)$ solves $\Pi_B(n,p,\gamma,1) = c$ ($\gamma = (1-p)/p$ if and only if $c = \overline{c}_n(p)$). For $p < 0.5$, if $c \in (\underline{c}_n(p),\overline{c}_n(p)]$, then $\gamma = 1$ and $\delta \in [p/(1-p),1)$ solves $\Pi_R(n,p,1,\delta) = c$ ($\delta = p/(1-p)$ if and only if $c = \overline{c}_n(p)$).*

Given this proposition, we can define functions γ_B and γ_R that map the vector of parameters (n,c,p) into its corresponding type-symmetric electoral equilibrium. Figure 1 shows the location of all such equilibria for any fixed $n \geq 2$, $p \in \{0.3,0.5,0.7\}$ and for all $c \in \mathbb{R}_+$.

[10] The proofs of all lemmas and propositions of this section can be found in Appendix A.

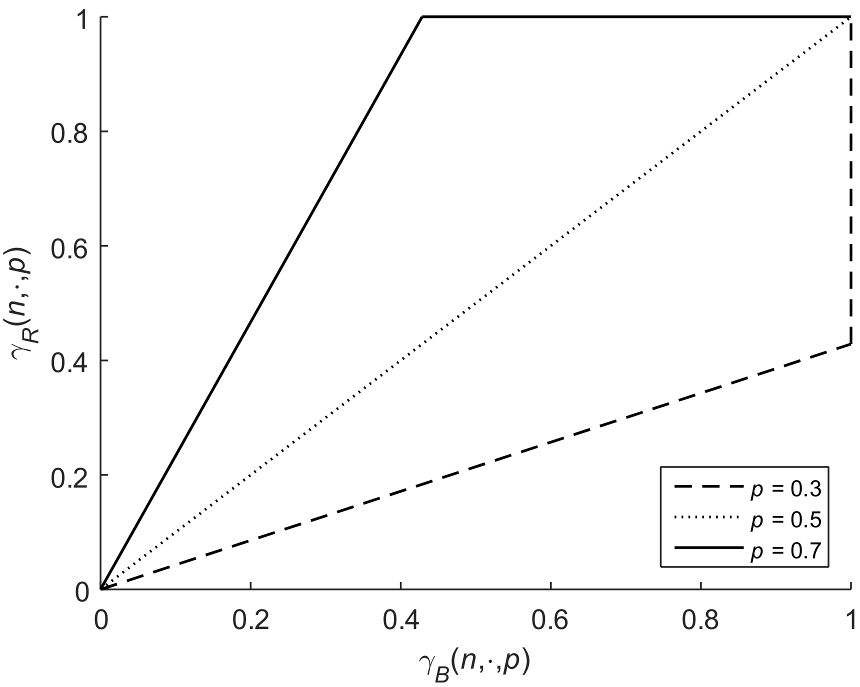

Figure 1

The solid line in Figure 1 is the locus of all points of the form $(\gamma_B(n,c,0.7), \gamma_R(n,c,0.7))$. As seen in Proposition 1, its slope is $p/(1-p) \approx 2.33$ for all c between $\bar{c}_n(0.7)$ and 1, and it becomes horizontal at $(\gamma_B, \gamma_R) = ((1-p)/p, 1) \approx (0.43, 1)$. Similarly, the slope of the dashed line is approximately 0.43 for all c between $\bar{c}_n(0.3)$ $(= \bar{c}_n(0.7))$ and 1, and it becomes vertical at $(\gamma_B, \gamma_R) = (1, p/(1-p)) \approx (1, 0.43)$. By the symmetry of the model, $\gamma_R(n,c,p) \equiv \gamma_B(n,c,1-p)$ (see appendix A), so that the dashed line is just the reflection of the solid line across the 45° line, which, in turn, is the locus of all points of the form $(\gamma_B(n,c,0.5), \gamma_R(n,c,0.5))$.

A basic yet crucial property of interior electoral equilibria is the well-known (see, e.g., Goeree and Großer, 2007, and Taylor and Yildirim, 2010a) neutrality result of the pivotal voting model, according to which the expected turnout for each candidate ($np\gamma_B(n,c,p)$ and $n(1-p)\gamma_R(n,c,p)$) should be equal. This property is stated in the next lemma. In it, we make use of a notation that will be used throughout the paper: $\gamma^*(n,c) := \gamma_B(n,c,0.5)$.

Lemma 1 *Given* $n \geq 2$*,* $p \in (0,1)$ *and* $c \in (\bar{c}_n(p), 1)$*, we have* $\gamma^*(n,c) \in (0,1)$ *and*

$(\gamma_B(n,c,p), \gamma_R(n,c,p)) = (\gamma^*(n,c)/(2p), \gamma^*(n,c)/(2(1-p)))$.

We know from Proposition 1 that, if $c \in (\bar{c}_n(p), 1)$, then $(\gamma_B(n,c,p), \gamma_R(n,c,p))$ is an interior (i.e., $(\gamma_B(n,c,p), \gamma_R(n,c,p)) \in (0,1)^2$) type-symmetric electoral equilibrium. The fact that, in this case, $\gamma_R(n,c,p) = (p/(1-p))\gamma_B(n,c,p)$, as follows from Lemma 1, is the standard neutrality result of the pivotal voting model. The fact that we can write γ_B and γ_R as simple functions of the pollster's choice variable p will be used extensively in the following sections, since it brings analytical tractability to the pollster's problem.

As mentioned in the introductory section, this neutrality result is at odds with electoral data. In the next section we will argue that the apparent discrepancy between theory and data can be better understood if one recognizes that the pollster may have incentives to misreport, that is, honesty is not assumed from the outset.

The comparative statics of the type-symmetric electoral equilibrium with respect to parameters c and p are described in the following proposition. Although not strictly necessary for the remainder of our analysis (see Lemma 3), for completeness' sake, results for the noninterior equilibrium case are also provided.

Proposition 2 *Given $n \geq 2$, the functions $\gamma_B(n,\cdot,\cdot)$ and $\gamma_R(n,\cdot,\cdot)$ are continuous. Furthermore,*

i. γ_B and γ_R are decreasing in c.[11] *Moreover,*

(a) if $p \in [0.5, 1)$, then $\partial\gamma_B(n,c,p)/\partial c < 0$ if $c \in (\underline{c}_n(p), 1)$ and $\partial\gamma_R(n,c,p)/\partial c < 0$ if $c \in (\bar{c}_n(p), 1)$;

(b) if $p \in (0, 0.5)$, then $\partial\gamma_B(n,c,p)/\partial c < 0$ if $c \in (\bar{c}_n(p), 1)$ and $\partial\gamma_R(n,c,p)/\partial c < 0$ if $c \in (\underline{c}_n(p), 1)$.

ii. γ_B is decreasing and γ_R is increasing in p. Moreover,

[11] Throughout the paper, we favor the "in(de)creasing"/"strictly in(de)creasing" nomenclature over the "monotonic nonde(in)creasing"/"in(de)creasing" one.

(a) *if* $p \in [0.5, 1)$*, then* $\partial\gamma_B(n, c, p)/\partial p < 0$ *if* $c \in (\underline{c}_n(p), 1)$ *and* $\partial\gamma_R(n, c, p)/\partial p > 0$ *if* $c \in (\overline{c}_n(p), 1)$*;*

(b) *if* $p \in (0, 0.5)$*, then* $\partial\gamma_B(n, c, p)/\partial p < 0$ *if* $c \in (\overline{c}_n(p), 1)$ *and* $\partial\gamma_R(n, c, p)/\partial p > 0$ *if* $c \in (\underline{c}_n(p), 1)$*.*

The intuition is as follows. Firstly, as voting becomes more costly, the net benefit of voting decreases and thus we should expect a smaller turnout. Secondly, the probability according to which a citizen votes is decreasing in the perceived proportion of citizens of their same type, for if a citizen believes that there are many others who support his or her preferred candidate, then he/she has a larger incentive to free ride, thus avoiding the cost c.

Figure 2 illustrates the first part of this proposition. The graph of $\gamma_B(10, \cdot, 0.7)$ meets the upper horizontal line at $\underline{c}_n(p) \approx 0.07$, while the graph of $\gamma_R(10, \cdot, 0.7)$, at $\overline{c}_n(p) \approx 0.32$. Figure 3 illustrates the second part of this proposition.

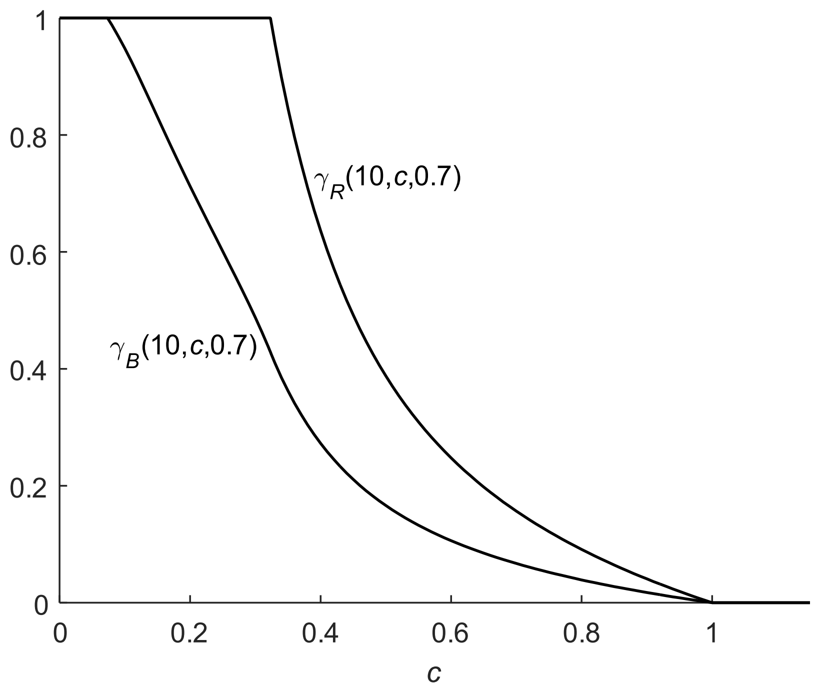

Figure 2

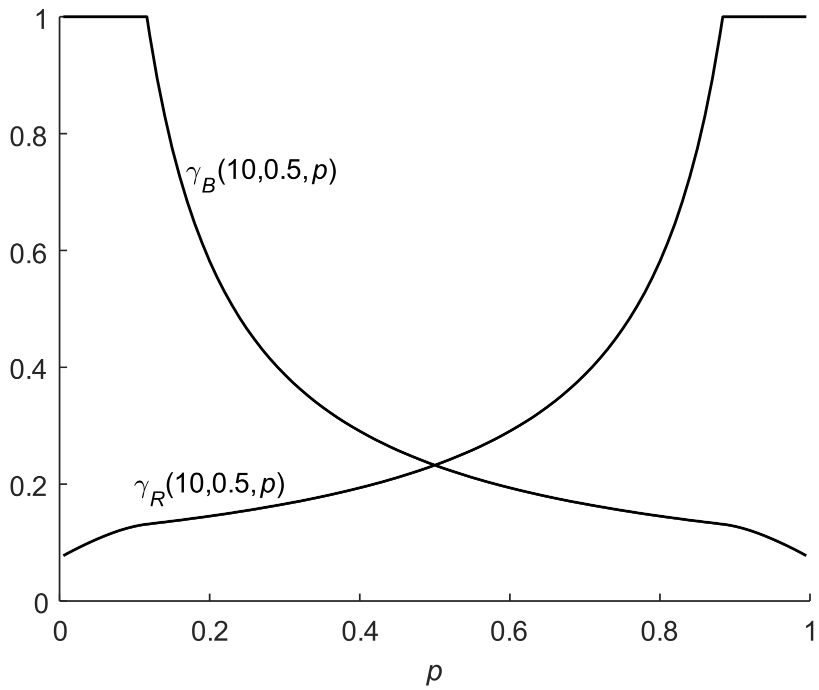

Figure 3

Figure 3 showcases five properties of type-symmetric electoral equilibria stated in the propositions above (and proved in appendix A). First, that $\gamma_R(n,c,p) = \gamma_B(n,c,1-p), \forall p \in (0,1)$, which is due to the symmetry of the electoral game: (γ,δ) is a type-symmetric equilibrium of the electoral game with parameter values n, c and p if and only if (δ,γ) is a type-symmetric equilibrium of the electoral game with parameters values n, c and $1-p$. Second, $\gamma_B(n,c,\cdot)$ is decreasing and $\gamma_R(n,c,\cdot)$ is increasing, as stated in Proposition 2. Third, restricted to all domain points p such that $(\gamma_B(n,c,p), \gamma_R(n,c,p)) \in (0,1)^2$, the graph of $\gamma_B(n,c,\cdot)$ is an equilateral hyperbola. In fact, Proposition 1 implies that the equilibrium is interior if, and only if, $c \in (\bar{c}_n(p), 1)$ – in which case Lemma 1 gives $p\gamma_B(n,c,p) = 0.5\gamma^*(n,c)$, constant in p. Fourth, for low enough values of p, B-citizens feel so outnumbered that they all decide to go cast their votes (by Proposition 1, this happens precisely for those $p \in (0,0.5)$ such that $c \leq \bar{c}_n(p)$). Finally, although for high enough values of p (in more precise terms, for those $p \in (0.5,1)$ such that $c \leq \bar{c}_n(p)$, which in this figure corresponds to $p \approx 0.88$) the graph of $\gamma_B(n,c,\cdot)$ is no longer a hyperbola, there is no break there, neither in terms of continuity nor in terms of differentiability, as shown in the proof of Proposition 2.

Out of these five properties, only the last one is new in the literature. It is the one that allows for the above statement of Proposition 2. The following lemma collects a few more basic results of the pivotal voting model. The convergence of $\gamma_B(n,c,p)$ and $\gamma_R(n,c,p)$ to 0 comes originally from Palfrey and Rosenthal (1985), while the convergence of $n\gamma_B(n,c,p)$ and $n\gamma_R(n,c,p)$ comes from Taylor and Yildirim (2010a).

Lemma 2 *Given* $p \in (0,1)$, $\bar{c}_n(p)$ *is strictly decreasing in* n *and* $\lim_{n\to\infty} \bar{c}_n(p) = 0$. *Moreover,* $\lim_{n\to\infty} \gamma_B(n,c,p) = \lim_{n\to\infty} \gamma_R(n,c,p) = 0$, $\lim_{n\to\infty} n\gamma_B(n,c,p) > 0$ *and* $\lim_{n\to\infty} n\gamma_R(n,c,p) > 0$.

It may be noted that Lemma 2 implies that, given p and $c \in (0,1)$, there exists a critical population size $n_0(c,p)$ such that, if $n \geq n_0(c,p)$, then $\bar{c}_n(p) < c$ and the electoral equilibrium will be interior, as proved in Proposition 1. The following lemma will be essential for the analysis in the remainder of this work, in that it guarantees the existence of a critical population size that does not depend on p, a choice variable of the electoral pollster.[12]

Lemma 3 *Given* $c \in (0,1)$, *there exists* $n_0(c) \in \mathbb{N}$ *such that, for all* $n \geq n_0(c)$,

$$c > \bar{c}_n(p) \text{ and } (\gamma_B(n,c,p), \gamma_R(n,c,p)) \in (0,1)^2, \forall p \in [\bar{q}, 1-\bar{q}].$$

As an illustration of this lemma, given $\bar{q} = 0.05$, for $c = 0.3$, $c = 0.5$ and $c = 0.7$, it would suffice to take $n \geq 69$, $n \geq 23$ and $n \geq 10$, respectively, in order to ensure the interiority of the electoral equilibrium for any $p \in [\bar{q}, 1-\bar{q}]$. Regarding the existence of two conclusions in Lemma 3, the interiority conclusion $(\gamma_B(n,c,p), \gamma_R(n,c,p)) \in (0,1)^2$ follows immediately from $\bar{c}_n(p) < c < 1$ and Proposition 1.

[12] This is where the assumption $q \in [\bar{q}, 1-\bar{q}]$ comes into play. With this assumption in hand, the pollster's choice of p is naturally constrained to the $[\bar{q}, 1-\bar{q}]$ interval as well. If we had let $q \in (0,1)$, there would not exist a uniform critical population size capable of guaranteeing that, regardless of the $p \in (0,1)$ chosen by the pollster, the electoral equilibrium would be interior.

All the above results hold regardless of the announced probability p of any given citizen being a B supporter being equal to the true probability q or not. In the next two sections, p will be endogenized, and the two previous lemmas will be used both in the analysis and the intuition regarding this endogenization. There, we will also see that it makes sense, both in analyzing the pollster's behavior and its welfare implications, to consider the distribution of the difference in the number of votes cast for the two candidates. Since such analysis is nonstandard in this literature, it is presented below.

The election result is a random vector (b, r, a), where b, r and a denote the number of votes for B, the number of votes for R and the number of abstentions, respectively. The true probability distribution (known only to the pollster) of (b, r, a) is

$$\text{Multinomial}\left(n, q\gamma_B\left(n, c, p\right), \left(1-q\right)\gamma_R\left(n, c, p\right), 1 - q\gamma_B\left(n, c, p\right) - \left(1-q\right)\gamma_R\left(n, c, p\right)\right). \quad (5)$$

To simplify the notation momentarily, write the distribution of (b, r, a) in (5) as $\text{Multinomial}\left(n, \beta, \rho, 1-\beta-\rho\right)$. Let us call the distribution of $b - r$ the multinomial difference distribution, $\text{MultiDiff}\left(n, \beta, \rho\right)$. Its characteristic function $\varphi_{\text{MultiDiff}(n,\beta,\rho)}$ can be obtained as follows.

First, note that $b - r = d$ if and only if $d + n = b + (n - r) = b + (b + a) = 2b + a$. Now, on the one hand, we have, for any $t \in \mathbb{R}$,

$$\varphi_{\text{MultiDiff}(n,\beta,\rho)}\left(t\right) = \text{E}\left(e^{itd}\right) = \text{E}\left(e^{it(d+n-n)}\right) = e^{-itn}\,\text{E}\left(e^{it(d+n)}\right) = e^{-itn}\,\text{E}\left(e^{i(2tb+ta)}\right).$$

On the other hand, $\text{E}\left(e^{i(2tb+ta)}\right)$ could be thought of as

$$\varphi_{\text{Multinomial}(n,\beta,\rho,1-\beta-\rho)}\left(2t, 0, t\right),$$

which equals $\left(\beta e^{i2t}+\rho e^{i0}+(1-\beta-\rho)\,e^{it}\right)^n$.[13] Therefore,

$$\varphi_{\text{MultiDiff}(n,\beta,\rho)}\left(t\right)=e^{-itn}\left(\beta e^{i2t}+\rho e^{i0}+(1-\beta-\rho)\,e^{it}\right)^n=\left(1+\beta\left(e^{it}-1\right)+\rho\left(e^{-it}-1\right)\right)^n. \tag{6}$$

Once we have studied the determination of p by the electoral pollster, the characteristic function given in (6), in conjunction with the two previous lemmas, will also allow us to find (through Levy's Continuity Theorem) the asymptotic distribution of the difference of votes. This will be key in establishing our welfare results, which will hold not only in the limit, but also for any sufficiently large n.

4 Electoral pollster's behavior

The main concern of the pollster is to be considered of good quality, by "getting the election right." The election result, a realization of the random vector (b,r,a), is known to all citizens right after the election. Although (b,r,a) is distributed according to (5), the citizens, based on the information reported by the pollster, p, believe that it is distributed as

$$\text{Multinomial}\left(n,p\gamma_B\left(n,c,p\right),\left(1-p\right)\gamma_R\left(n,c,p\right),1-p\gamma_B\left(n,c,p\right)-\left(1-p\right)\gamma_R\left(n,c,p\right)\right).$$

In particular, the citizens believe that, on average, candidate B will receive $np\gamma_B\left(n,c,p\right)$ votes and candidate R will receive $n\left(1-p\right)\gamma_R\left(n,c,p\right)$ votes. By Lemmas 3 and 1, these expected turnouts, in the eyes of citizens, are equal, as long as $n\geq n_0\left(c\right)$.

The pollster's rating depends on how the actual election result compares to the result implied by the report. We assume that, given an election result (b,r,a), the pollster's rating is given by

$$-\left[\left(b-np\gamma_B\left(n,c,p\right)\right)^2+\left(r-n\left(1-p\right)\gamma_R\left(n,c,p\right)\right)^2\right],$$

[13] The characteristic function of the Multinomial distribution can be checked in Johnson et al. (1997, p. 37).

which is a measure of how far the election result was from the result expected by the citizens.[14]

Defining $e_B(b,n,c,p) := b - np\gamma_B(n,c,p)$ and $e_R(r,n,c,p) := r - n(1-p)\gamma_R(n,c,p)$ as the pollster's prediction error, from the perspective of the citizens, regarding B-votes and R-votes, respectively, the rating is

$$- \left[(e_B(b,n,c,p))^2 + (e_R(r,n,c,p))^2 \right].$$

As the actual election result is random, the utility function of the pollster is the expected value of its rating. That is,

$$U(n,c,p,q) = -\,\mathrm{E}\left[(e_B(b,n,c,p))^2 + (e_R(r,n,c,p))^2 \right], \tag{7}$$

where E is being taken with respect to the true probability distribution (5), which depends on q. The pollster's optimization problem is:

$$\max_{p\in[\bar{q},1-\bar{q}]} U(n,c,p,q). \tag{8}$$

We denote a solution to this problem by $p_n^*(c,q)$. Because it is commonly known that $q \in [\bar{q}, 1-\bar{q}]$ (you may think that a society with a degree of ideological homogeneity so high that $q < \bar{q}$ or $q > 1-\bar{q}$ would hardly find it necessary to have elections), all citizens would know for sure that the pollster was not being truthful if it released $q < \bar{q}$ or $q > 1-\bar{q}$. That is why the constraint in (8) is set as $p \in [\bar{q}, 1-\bar{q}]$, and that (together with continuity of U in p) is why we are assured of the existence of a solution to that problem.

[14]The website FiveThirtyEight.com, for instance, ranks electoral pollsters in a similar way. The difference is that usually pollsters report a direct prediction of the election result (e.g., "candidate B will win by 10 points"), whereas here the pollster reports the probability that someone favors candidate B.

Remark 1 *It may be noted that our utility function resembles, in absolute value,*

$$\frac{\mathrm{E}\left[\left(b-np\gamma_B\left(n,c,p\right)\right)^2\right]}{np\gamma_B\left(n,c,p\right)}+\frac{\mathrm{E}\left[\left(r-n\left(1-p\right)\gamma_R\left(n,c,p\right)\right)^2\right]}{n\left(1-p\right)\gamma_R\left(n,c,p\right)}$$
$$+\frac{\mathrm{E}\left[\left(\left(n-b-r\right)-n\left(1-p\gamma_B\left(n,c,p\right)-\left(1-p\right)\gamma_R\left(n,c,p\right)\right)\right)^2\right]}{n\left(1-p\gamma_B\left(n,c,p\right)-\left(1-p\right)\gamma_R\left(n,c,p\right)\right)},$$

which is the expected value of the statistic of the Pearson Chi-Squared Test of goodness of fit for a multinomial distribution, with a null hypothesis that the true probability distribution of (b,r,a) *has parameters* $\left(p\gamma_B\left(n,c,p\right),\left(1-p\right)\gamma_R\left(n,c,p\right),1-p\gamma_B\left(n,c,p\right)-\left(1-p\right)\gamma_R\left(n,c,p\right)\right)$, *that is, that the pollster is being truthful.*

In fact, by Proposition 1, in an interior equilibrium, the second denominator in this formula equals the first one, so that, by Lemma 2, they both converge, as $n\to\infty$, *to the same positive number, whereas the third denominator, equal to* $n-2np\gamma_B\left(n,c,p\right)$, *diverges. Therefore, it is only logical for a zero weight to be used for the third term in Pearson's formula, and the same positive weights to be given to the first two terms, as in* U. *Thus, for analytical tractability and without loss of insight, we assume that the reputational concerns of the pollster can be reasonably represented by* U.

Also, even with a truthful report ($p=q$), *in an interior equilibrium, the expected value of the above test statistic can be seen to equal* 2, *which corresponds to a larger than 10% p-value.*[15] *Therefore, the null hypothesis would not be rejected for any significance level up to 10%. In other words, our model is not forcing, in any way, misreporting by the pollster, since the latter already achieves a good result by releasing a truthful report – although, as will be shown in Proposition 4, it can do even better by misreporting. In this way, although our model concerns a single election, it could also be used to emulate a dynamic environment with multiple elections and a stable presence in the market of the polling firm, with no reason for a lack of trust on the*

[15] In an interior equilibrium, by Lemma 1, the expression becomes $\frac{E\left[\left(i-\frac{n\gamma^*}{2}\right)^2\right]}{\frac{n\gamma^*}{2}}+\frac{E\left[\left(j-\frac{n\gamma^*}{2}\right)^2\right]}{\frac{n\gamma^*}{2}}+\frac{E\left[\left(k-n(1-\gamma^*)\right)^2\right]}{n(1-\gamma^*)}=\frac{\mathrm{Var}(i)+\left(\frac{n\gamma^*}{2}-\frac{n\gamma^*}{2}\right)^2}{\frac{n\gamma^*}{2}}+\frac{\mathrm{Var}(j)+\left(\frac{n\gamma^*}{2}-\frac{n\gamma^*}{2}\right)^2}{\frac{n\gamma^*}{2}}+\frac{\mathrm{Var}(k)+\left(n\left(1-\frac{\gamma^*}{2}-\frac{\gamma^*}{2}\right)-n(1-\gamma^*)\right)^2}{n(1-\gamma^*)}=\frac{\mathrm{Var}(i)}{\frac{n\gamma^*}{2}}+\frac{\mathrm{Var}(j)}{\frac{n\gamma^*}{2}}+\frac{\mathrm{Var}(k)}{n(1-\gamma^*)}=\frac{\frac{n\gamma^*}{2}\left(1-\frac{\gamma^*}{2}\right)}{\frac{n\gamma^*}{2}}+\frac{\frac{n\gamma^*}{2}\left(1-\frac{\gamma^*}{2}\right)}{\frac{n\gamma^*}{2}}+\frac{n\gamma^*(1-\gamma^*)}{n(1-\gamma^*)}=1-\frac{\gamma^*}{2}+1-\frac{\gamma^*}{2}+\gamma^*=2$, where $\gamma^*=\gamma^*\left(n,c\right)$.

part of the citizens.

To better grasp the pollster's incentives, we can rewrite this function as

$$\begin{aligned} U(n,c,p,q) &= -\left[\mathrm{E}\left(e_B^2(b,n,c,p)\right) + \mathrm{E}\left(e_R^2(r,n,c,p)\right)\right] \\ &= -\left[\begin{array}{c} \left[\mathrm{E}\left(e_B(b,n,c,p)\right)\right]^2 + \mathrm{Var}\left(e_B(b,n,c,p)\right) \\ +\left[\mathrm{E}\left(e_R(r,n,c,p)\right)\right]^2 + \mathrm{Var}\left(e_R(r,n,c,p)\right) \end{array}\right], \end{aligned} \tag{9}$$

where expected values and variances are computed using the real distribution (5). Thus, the pollster faces a tradeoff between minimizing the expected value and the variance of its prediction errors.

On the one hand, since

$$\begin{aligned} \mathrm{E}\left(e_B(b,n,c,p)\right) &= \mathrm{E}\left(b - np\gamma_B(n,c,p)\right) = nq\gamma_B(n,c,p) - np\gamma_B(n,c,p) \\ &= n(q-p)\gamma_B(n,c,p) \end{aligned} \tag{10}$$

and

$$\begin{aligned} \mathrm{E}\left(e_R(r,n,c,p)\right) &= \mathrm{E}\left(r - n(1-p)\gamma_R(n,c,p)\right) = n(1-q)\gamma_R(n,c,p) - n(1-p)\gamma_R(n,c,p) \\ &= n(p-q)\gamma_R(n,c,p), \end{aligned} \tag{11}$$

in order to minimize

$$\left[\mathrm{E}\left(e_B(b,n,c,p)\right)\right]^2 + \left[\mathrm{E}\left(e_R(r,n,c,p)\right)\right]^2 = n^2(p-q)^2\left(\gamma_B^2(n,c,p) + \gamma_R^2(n,c,p)\right), \tag{12}$$

the pollster should release a truthful report ($p = q$).

On the other hand, we have

$$\begin{aligned}\operatorname{Var}\left(e_B\left(b,n,c,p\right)\right) &= \operatorname{Var}\left(b - np\gamma_B\left(n,c,p\right)\right) = \operatorname{Var}\left(b\right) \\ &= n\left(q\gamma_B\left(n,c,p\right)\right)\left(1 - q\gamma_B\left(n,c,p\right)\right) \qquad (13)\end{aligned}$$

and

$$\begin{aligned}\operatorname{Var}\left(e_R\left(r,n,c,p\right)\right) &= \operatorname{Var}\left(r - n\left(1-p\right)\gamma_R\left(n,c,p\right)\right) = \operatorname{Var}\left(r\right) \\ &= n\left(\left(1-q\right)\gamma_R\left(n,c,p\right)\right)\left(1 - \left(1-q\right)\gamma_R\left(n,c,p\right)\right). \qquad (14)\end{aligned}$$

If $c \in (0,1)$ and $n \geq n_0(c)$, by Lemma 3, the electoral equilibrium will be interior, so that Lemma 1 applies and we obtain

$$\begin{aligned}&\operatorname{Var}\left(e_B\left(b,n,c,p\right)\right) + \operatorname{Var}\left(e_R\left(r,n,c,p\right)\right) = \operatorname{Var}\left(b\right) + \operatorname{Var}\left(r\right) \\ &= n\left(q\gamma_B\left(n,c,p\right)\left(1 - q\gamma_B\left(n,c,p\right)\right) + \left(1-q\right)\gamma_R\left(n,c,p\right)\left(1 - \left(1-q\right)\gamma_R\left(n,c,p\right)\right)\right) \\ &= n\left(\frac{q}{p}\frac{\gamma^*}{2}\left(1 - \frac{q}{p}\frac{\gamma^*}{2}\right) + \frac{1-q}{1-p}\frac{\gamma^*}{2}\left(1 - \frac{1-q}{1-p}\frac{\gamma^*}{2}\right)\right), \qquad (15)\end{aligned}$$

where γ^* is short for $\gamma^*(n,c)$. This is not minimized by taking $p = q$. In fact, the derivative of this expression with respect to p is

$$\frac{n\gamma^*}{2p^3\left(1-p\right)^3}\left(\begin{array}{c}\left(2q-1\right)p^4 - \left(2\gamma^* q^2 + 2\left(2-\gamma^*\right)q + \gamma^* - 1\right)p^3 \\ +3q\left(\gamma^* q + 1\right)p^2 - q\left(3\gamma^* q + 1\right)p + \gamma^* q^2\end{array}\right),$$

which, at $p = q$, equals $n\gamma^*\left(1-\gamma^*\right)\left(2q-1\right)/\left(2q\left(1-q\right)\right)$, which is only 0 if $q = 0.5$ (in which case the pollster will indeed choose not to misreport, as shown in Proposition 4 ahead).

Given this tradeoff, in order to avoid putting its reputation at risk, the pollster is typically willing to accept nonzero expected errors, as long as their variance is sufficiently small. For instance, if $q > 0.5$, since this expression for the partial derivative of $\operatorname{Var}\left(e_B\left(b,n,c,p\right)\right) + \operatorname{Var}\left(e_R\left(r,n,c,p\right)\right)$ with respect to p at $p = q$ becomes positive, reporting $p < q$ may be beneficial to the pollster – even in the absence of any partisan interest whatsoever. That this is in fact the case will be proven in Proposition 4, the central result of this section.

An alternative but related way of understanding this sum of variances is to note, from the covariance formula of a multinomial distribution, that

$$\mathrm{Cov}\,(b,r) = -n\,(q\gamma_B\,(n,c,p))\,((1-q)\,\gamma_R\,(n,c,p)) = -n\left(\frac{\gamma^*}{2}\right)^2 \frac{q}{p}\frac{1-q}{1-p},$$

so that (15) gives

$$\begin{aligned}\mathrm{Var}\,(b-r) &= \mathrm{Var}\,(b) + \mathrm{Var}\,(r) - 2\,\mathrm{Cov}\,(b,r)\\ &= \frac{n\gamma^*}{2}\left(\frac{q}{p}\left(1-\frac{q}{p}\frac{\gamma^*}{2}\right)+\frac{1-q}{1-p}\left(1-\frac{1-q}{1-p}\frac{\gamma^*}{2}\right)\right)+\frac{n\gamma^*}{2}\frac{q}{p}\frac{1-q}{1-p}\gamma^*.\end{aligned}$$

Lemma 2 allows us to define $m\,(c) := \lim_{n\to\infty} n\gamma^*\,(n,c)$ $(= \lim_{n\to\infty} n\gamma_B\,(n,c,0.5))$ and thus yields that, for a large n and fixed p, $\mathrm{Var}\,(b-r)$ should approach

$$\frac{m\,(c)}{2}\left(\frac{q}{p}+\frac{1-q}{1-p}\right), \tag{16}$$

just as $\mathrm{Var}\,(b) + \mathrm{Var}\,(r)$ does (from (15)). Expression (16) can also be interpreted as the asymptotic expected turnout of the election,

$$\lim_{n\to\infty}\left(nq\gamma_B\,(n,c,p) + n\,(1-q)\,\gamma_R\,(n,c,p)\right) = \lim_{n\to\infty}\frac{n\gamma^*\,(n,c)}{2}\left(\frac{q}{p}+\frac{1-q}{1-p}\right).$$

In this sense, for large n, in order to minimize $\mathrm{Var}\,(b)+\mathrm{Var}\,(r)$, the pollster may want to choose a p close to the one that minimizes (16) instead – that is, $p = \phi\,(q) := \left(1+\sqrt{1/q-1}\right)^{-1}$.[16] In fact, we will establish, as a step in showing the comparative statics results in Proposition 5, that the p reported by the pollster will necessarily lie between q and $\phi\,(q)$ (see Lemma 4).

This makes sense since, although the only way of matching the B- and the R-turnouts on av-

[16] This also happens to be the value that the pollster would choose if it wished its report to reflect the expected turnout for B over the expected turnout for R in the election (that is, so that $p/\,(1-p) = (q\gamma_B\,(n,c,p))\,/\,((1-q)\,\gamma_R\,(n,c,p)) = (q/p)\,/\,((1-q)\,/\,(1-p))$).

erage (in an interior equilibrium setting, $nq\gamma_B(n,c,p) = n(q/p)\gamma^*/2$ and $n(1-q)\gamma_R(n,c,p) = n((1-q)/(1-p))\gamma^*/2)$ is by reporting $p = q$, the pollster's rating is based not on averages of elections, but on this particular election. In this way, it is willing to report a p that does not imply an expected election tie, as long as the difference between the number of votes cast for B and R is still small and has less dispersion. The simplest way to achieve this effect is by making less people show up on election day.

The reader may want to note that our assumption on the pollster's behavior in no way forces it to misreport its poll results. In fact, if the model environment was such that voting was not considered to be costly, then it would never be in the pollster's best interest to misreport, as shows the following proposition (the proof of which can be found in appendix B, alongside the proofs of all the remaining results of the paper).

Proposition 3 *Given $n \geq 2$ and $q \in [\overline{q}, 1-\overline{q}]$, if $c = 0$, then $p_n^*(c,q) = q$ is the unique solution to the pollster's problem.*

Thus, it is most natural to analyze the issue of misreporting of pre-election poll results within a costly voting framework, in which the pollster's choice of p affects turnouts, as presented in Proposition 2 (and, as a consequence, the probability that either party will win the election).

We are now ready to state and prove the main result of this section, which confirms the misreporting behavior suggested in the discussion above. According to it, misreporting is the norm, rather than the exception.

Proposition 4 *Given $q \in [\bar{q}, 1-\bar{q}]$, $c \in (0,1)$ and $n \geq n_0(c)$, the solution of the pollster's optimization problem, $p_n^*(c,q)$, is unique and such that,*

i. if $q = 0.5$, then $p_n^*(c,q) = q$;

ii. if $q > 0.5$, then $p_n^*(c,q) \in (0.5, q)$;

iii. if $q < 0.5$, then $p_n^*(c,q) \in (q, 0.5)$.

Thus, if there exists an expected majority in the society ($q \neq 0.5$) and n is sufficiently large so that the electoral equilibrium is interior, then a rational pollster driven purely by reputational motives will always misreport information. Moreover, it will do so in such a way that the citizens will believe that the majority is not as large as it actually is. As we will see in the next section, contrary to common belief, misreporting is not only in the best interest of the pollster, but also of society.

This proposition also establishes that, fixed $n \geq 2$, $c \in (0,1)$ and $n \geq n_0(c)$, $p_n^*(c, \cdot) : [\bar{q}, 1-\bar{q}] \to [\bar{q}, 1-\bar{q}]$ is a well-defined, single-valued, function. It is plotted below for $n = 100$ and $c \in \{0.3, 0.5, 0.7\}$, together with the 45° line (which corresponds to p_n^* when $c = 0$, by Proposition 3).

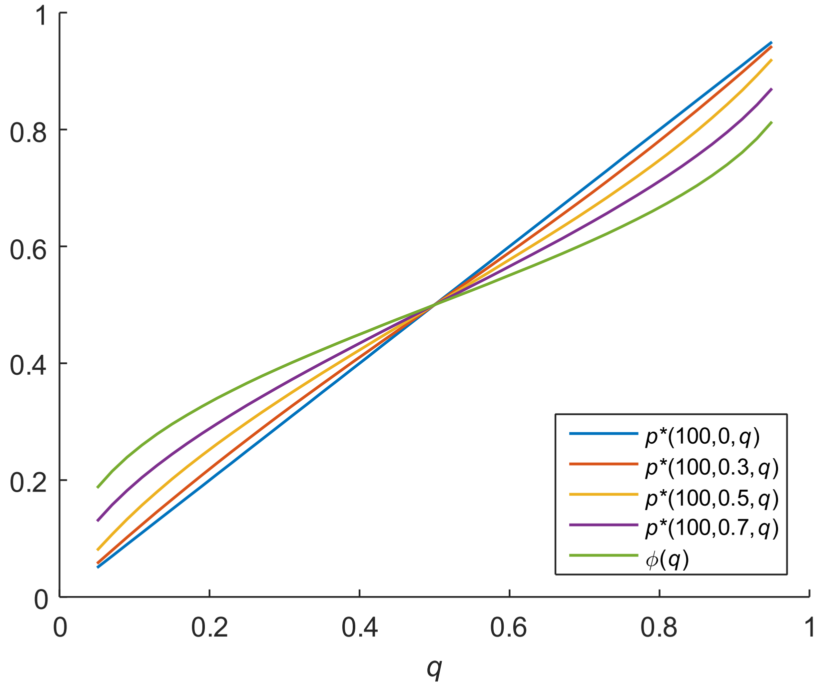

Figure 4

Also present in Figure 4 is the aforementioned bound $\phi(q)$, which lies between 0.5 and q. This hints to the following strengthening of Proposition 4.

Lemma 4 *Given $c \in (0,1)$ and $n \geq n_0(c)$, if $q \in (0.5, 1-\bar{q}]$, then $p_n^*(c,q) \in (\phi(q), q)$, and if $q \in [\bar{q}, 0.5)$, then $p_n^*(c,q) \in (q, \phi(q))$, where $\phi(q) := \left(1 + \sqrt{1/q - 1}\right)^{-1}$.*

This more precise bound helps in the addressing of the comparative statics issues regarding p_n^*.

Proposition 5 *Given $q \in (\bar{q}, 1-\bar{q})$, $c \in (0,1)$ and $n \geq n_0(c)$, the function p_n^* is continuously differentiable. Furthermore,*

i. $\partial p_n^*(c,q)/\partial q > 0$*;*

ii. $\partial p_n^*(c,q)/\partial c \lesseqqgtr 0$ *if* $q \gtreqqless 0.5$.

Part (i) should not come as a surprise, since both channels that go into the pollster's decision process seem to point in the same direction: if q is nudged from 0.6 to 0.7 for instance, then the value of p that minimizes $[\mathrm{E}(e_B(b,n,c,p))]^2 + [\mathrm{E}(e_R(r,n,c,p))]^2$ also moves from 0.6 to 0.7, while the value of p that minimizes $\mathrm{Var}(e_B(b,n,c,p)) + \mathrm{Var}(e_R(r,n,c,p))$, at least for a large n, should also move upward, from something around $\phi(0.6) = \left(1+\sqrt{1/0.6-1}\right)^{-1}$ to something around $\phi(0.7) = \left(1+\sqrt{1/0.7-1}\right)^{-1} > \phi(0.6)$.

As for part (ii), note, by Lemmas 3 and 1, that $[\mathrm{E}(e_B(b,n,c,p))]^2 + [\mathrm{E}(e_R(r,n,c,p))]^2$ converges to

$$\left(\frac{m(c)}{2}\right)^2 (p-q)^2 \left(\frac{1}{p^2} + \frac{1}{(1-p)^2}\right)$$

(see (12)). Thus, at least for large n, maximizing U with respect to p entails minimizing

$$\left(\frac{m(c)}{2}\right)^2 (p-q)^2 \left(\frac{1}{p^2} + \frac{1}{(1-p)^2}\right) + \frac{m(c)}{2}\left(\frac{q}{p} + \frac{1-q}{1-p}\right)$$

(see (16)) or, equivalently,

$$\frac{m(c)}{2}(p-q)^2 \left(\frac{1}{p^2} + \frac{1}{(1-p)^2}\right) + 1\left(\frac{q}{p} + \frac{1-q}{1-p}\right).$$

The first of these summands pushes p to q, while the second one brings p to $\phi(q)$, which is lower than q if we use the $q > 0.5$ case to fix ideas. The weights attributed to these two forces are $m(c)/2$ and 1. If c is nudged upward, then $m(c)$ should fall, by Proposition 2 (recall that $m(c) = \lim_{n\to\infty} n\gamma_B(n, c, 0.5)$ and $\gamma_B(n, c, 0.5)$ decreases with c). Thus, the first of these channels loses importance vis-à-vis the second one, which results in a lower, closer to $\phi(q)$, value of p.

A few comments about the implications of our model should be made.

Firstly, given the possibility of misreporting, the election will not be a tie in expected terms, as formalized in the lemma below. Rather, if $q > 0.5$, then the expected result of the election is a win for candidate B, as could be anticipated by the fact that the expected number of votes for B $(= nq\gamma_B(n, c, p) = nq\gamma^*/(2p))$ is greater than the expected number of votes for R $(= n(1-q)\gamma_R(n, c, p) = n(1-q)\gamma^*/(2(1-p)))$. Thus, our model suggests a mechanism to explain why, given $q \neq 0.5$, the minority candidate is not expected to win every other election, as would be implied by the canonical pivotal voting model with fixed and equal voting costs and truthful pre-election poll reports.[17]

Lemma 5 *Given* $c \in (0, 1)$, $n \geq n_0(c)$ *and* $q \in [\overline{q}, 1-\overline{q}]$, *then* $\Pr(B\ \textit{wins} \mid n, c, q, q) = 0.5$, *and:*[18]

i. if $q = 0.5$, *then* $\Pr(B\ \textit{wins} \mid n, c, p_n^*(c, q), q) = 0.5$;

ii. if $q > 0.5$, *then* $\Pr(B\ \textit{wins} \mid n, c, p_n^*(c, q), q) > 0.5$;

iii. if $q < 0.5$, *then* $\Pr(B\ \textit{wins} \mid n, c, p_n^*(c, q), q) < 0.5$.

Secondly, we provide a new interpretation for the emergence of the so-called "bandwagon effect" – the phenomenon according to which supporters of a specific candidate are more likely to cast their votes if their candidate ranks first in pre-election polls. Our model implies that part of this effect could actually be an illusion. In an interior equilibrium, citizens expect that,

[17] Alternatively, Taylor and Yildirim (2010b) show that, for a finite population size, the neutrality result may also break in favor of the majority, as long as voting costs are drawn from a common nondegenerate distribution.

[18] There is a slight abuse of notation here, since n, c, p and q are actually parameters, not events.

on average, the election will be a tie, or, in another way, that the ratio of the expected number of votes for each candidate should be

$$\frac{np_n^*(c,q)\,\gamma_B(n,c,p_n^*(c,q))}{n(1-p_n^*(c,q))\,\gamma_R(n,c,p_n^*(c,q))} = 1,$$

by Proposition 1. However, in actuality, this ratio would be

$$\frac{nq\gamma_B(n,c,p_n^*(c,q))}{n(1-q)\,\gamma_R(n,c,p_n^*(c,q))} = \frac{nq\gamma_B(n,c,p_n^*(c,q))}{n\frac{1-q}{1-p_n^*(c,q)}(1-p_n^*(c,q))\,\gamma_R(n,c,p_n^*(c,q))} = \frac{\frac{q}{p_n^*(c,q)}}{\frac{1-q}{1-p_n^*(c,q)}},$$

which is greater than 1, by Proposition 4.

For example, if $q = 0.80$, $c = 0.7$ and $n = 100$, the pollster would report $p^* = 0.71$, and the above ratio would be approximately 1.62. Thus, candidate B will receive, on average, 62% more votes than candidate R. An observer who believes that poll results are truthful, by observing such a discrepancy in a given election, could erroneously be led into thinking that, instead of the underdog effect predicted by Proposition 2, the bandwagon effect was in place. It could occur to him/her that the electoral equilibrium (γ_B, γ_R) as generated by the pivotal voting model was incorrect – namely, that γ_B (γ_R) was larger (lower) than predicted –, without realizing that the p and the $1-p$ terms in the ratio of the expected number of votes were wrong in the first place. In this way, our model implies that part of what is usually attributed to the bandwagon effect can actually be a direct consequence of misreporting.

Having studied the pollster's rational reporting behavior for a fixed (and sufficiently large) population size, one could ask whether such behavior is qualitatively different as the population size grows without bound. The following proposition shows that this is not the case.

Proposition 6 *Given $c \in (0,1)$ and $q \in [\overline{q}, 1-\overline{q}]$, $\lim_{n\to\infty} p_n^*(c,q)$ exists, and, if denoted by $p_\infty^*(c,q)$, is such that:*

i. if $q = 0.5$, then $p_\infty^*(c,q) = q$;

ii. if $q > 0.5$, then $p_\infty^*(c,q) \in (0.5, q)$;

iii. if $q < 0.5$*, then* $p^*_\infty(c, q) \in (q, 0.5)$*.*

This result will prove important also in the welfare discussion in the next section, in that, as anticipated in the brief discussion following the characteristic function of the multinomial difference distribution in (6), it makes possible the derivation of the asymptotic distribution of the difference in the number of votes cast for the two candidates, as stated in the lemma ahead.

Lemma 6 *Given* $c \in (0, 1)$ *and* $q \in [\bar{q}, 1 - \bar{q}]$*, the number of* B*-votes minus the number of* R*-votes converges in distribution to a random variable* Z *distributed as* $\text{Skellam}\left((q/p^*_\infty(c, q))\, m(c)/2, ((1-q)/(1-p^*_\infty(c, q)))\, m(c)/2\right)$.[19]

Given the possibility of misreporting (for $q \neq 0.5$), $p^*_\infty(c, q)$ exists and is not equal to q, by Proposition 6. Lemma 6 says that even the asymptotic distribution of the difference of votes will be skewed. In fact, if we use $q > 0.5$ to fix ideas, it will be skewed in favor of candidate B, since, by Proposition 6, we will have $(q/p^*_\infty(c, q))\, m(c)/2 > m(c)/2 > ((1-q)/(1-p^*_\infty(c, q)))\, m(c)/2$.

5 Welfare analysis

Now that we have characterized the electoral equilibrium and the solution to the pollster's problem, we are able to analyze the welfare implications of misreporting.

Let $\mathcal{I}(n, c, p, q)$ correspond to the expected (in the eyes of an outside observer, who knows not only p but also q) aggregate ideological component of utility within society (that is, citizens favoring the winning candidate earn $+1$, all others earn -1).

In order to proceed with this computation, the random vector (b, r, a) is insufficient, since we do not know how many of those a abstentions correspond to B supporters and how many correspond to R supporters. If we denote by n_B the number of B-citizens, the expected aggregate

[19] Given $m_B, m_R > 0$, $\text{Skellam}(m_B, m_R)$ is the distribution of $X - Y$, where $X \sim \text{Poisson}(m_B)$ and $Y \sim \text{Poisson}(m_R)$ are independent. It is also called the Poisson difference distribution.

ideological component of utility within society, given n, c, p, q, is:[20]

$$\begin{aligned}
\mathcal{I}(n,c,p,q) &= \mathrm{E}_{n_B}\begin{bmatrix} \Pr(B \text{ wins} \mid n,c,p,q,n_B)(n_B-(n-n_B)) \\ +\Pr(R \text{ wins} \mid n,c,p,q,n_B)((n-n_B)-n_B) \end{bmatrix} \\
&= \mathrm{E}_{n_B}\begin{bmatrix} \Pr(B \text{ wins} \mid n,c,p,q,n_B)(2n_B-n) \\ +(1-\Pr(B \text{ wins} \mid n,c,p,q,n_B))(n-2n_B) \end{bmatrix} \\
&= \mathrm{E}_{n_B}\left[(2\Pr(B \text{ wins} \mid n,c,p,q,n_B)-1)(2n_B-n)\right].
\end{aligned}$$

An exact expression for this ideological component of utility would be rather difficult to work with.[21] However, a tremendous shortcut would be made possible if we were to disregard the dependence between n_B and the probability that candidate B wins the election. In that case, we could use the following approximation for $\mathcal{I}(n,c,p,q)$:

$$\begin{aligned}
I(n,c,p,q) &= (2\Pr(B \text{ wins} \mid n,c,p,q)-1)\,\mathrm{E}_{n_B}(2n_B-n) \\
&= (2\Pr(B \text{ wins} \mid n,c,p,q)-1)(2nq-n), \qquad (17)
\end{aligned}$$

where we have used that n_B is binomially distributed with parameters (n,q), so that $\mathrm{E}_{n_B}(n_B) = nq$.

Let us argue that this is a reasonable approximation when n is large, in the sense that $\lim_{n\to\infty}((\mathcal{I}(n,c,p,q)-I(n,c,p,q))/n) = 0$ (that is, the difference in the per capita ideological component of utility when measured through $\mathcal{I}$ and I is negligible). Call $x_{n,n_B} := \Pr(B \text{ wins} \mid n,c,p,q,n_B)$ and $y_n := \Pr(B \text{ wins} \mid n,c,p,q)$, just to save a bit on notation. Note

[20] We are only interested in the utility of the citizens, that is, the utility function of the pollster does not enter the welfare function.

[21] If abstentions a are decomposed as $\tilde{b}+\tilde{r}$, where $\tilde{b}$ and $\tilde{r}$ represent B and R supporters who choose not to vote, then we could write $\mathcal{I}(n,c,p,q) = \sum_{\substack{b,r,\tilde{b},\tilde{r}=0 \\ b+r+\tilde{b}+\tilde{r}=n \\ b>r}} \binom{n}{b,r,\tilde{b},\tilde{r}} (q\gamma_B)^b ((1-q)\gamma_R)^r (q(1-\gamma_B))^{\tilde{b}} ((1-q)(1-\gamma_R))^{\tilde{r}}$ $-\sum_{\substack{b,r,\tilde{b},\tilde{r}=0 \\ b+r+\tilde{b}+\tilde{r}=n \\ b<r}} \binom{n}{b,r,\tilde{b},\tilde{r}} (q\gamma_B)^b ((1-q)\gamma_R)^r (q(1-\gamma_B))^{\tilde{b}} ((1-q)(1-\gamma_R))^{\tilde{r}}$.

that

$$\begin{aligned}
\frac{\mathcal{I}(n,c,p,q)-I(n,c,p,q)}{n} &= \mathrm{E}_{n_B}\left[(2x_{n,n_B}-1)\left(2\frac{n_B}{n}-1\right)-(2y_n-1)(2q-1)\right] \\
&= \mathrm{E}_{n_B}\left[(2x_{n,n_B}-1)\left(2\frac{n_B}{n}-2q\right)+(2x_{n,n_B}-2y_n)(2q-1)\right] \\
&= 2\,\mathrm{E}_{n_B}\left[(2x_{n,n_B}-1)\left(\frac{n_B}{n}-q\right)\right],
\end{aligned}$$

where we have used the Law of Iterated Expectations to get $\mathrm{E}_{n_B}[(2x_{n,n_B}-2y_n)(2q-1)] = 2(2q-1)\,\mathrm{E}_{n_B}[x_{n,n_B}-y_n]=0$.

Note that both $2x_{n,n_B}-1$ and $n_B/n-q$ have absolute values bounded by 1, and that, by the Strong Law of Large Numbers, $n_B/n-q \overset{a.s.}{\to} 0$. Therefore, the problem is reduced to one of showing that, if X_n and Y_n are random variables such that $|X_n|,|Y_n| \le 1, \forall n \in \mathbb{N}$, and $Y_n \overset{a.s.}{\to} 0$, then $\lim_{n\to\infty} \mathrm{E}(X_nY_n)=0$. That this is true can be shown as follows: since E is an increasing operator and $-|Y_n| \le X_nY_n \le |Y_n|, \forall n \in \mathbb{N}$, necessarily $\mathrm{E}(X_nY_n)$ is always between $-\mathrm{E}|Y_n|$ and $\mathrm{E}|Y_n|$, so that the Squeeze Theorem would yield $\lim_{n\to\infty}\mathrm{E}(X_nY_n)=0$ if we were to show that $\lim_{n\to\infty}\mathrm{E}|Y_n|=0$. And indeed, this is the case, as can be seen by noting that $|Y_n| \overset{a.s.}{\to} |0|=0$, $|Y_n|\le 1$, $\mathrm{E}\,1=1<\infty$, and applying the Dominated Convergence Theorem to ensure that $\lim_{n\to\infty}\mathrm{E}|Y_n|=\mathrm{E}\,0=0$.

The above approximation plays a major role in the proof of the main result of this section, Proposition 7, in appendix B. Using the limiting distribution given in Lemma 6, we are able to establish an asymptotic version of Lemma 5, so that, even as $n\to\infty$, one should expect that the candidate supported by the majority of the citizens will win the election more often than not.

Lemma 7 *Given* $c\in(0,1)$ *and* $q\in[\overline{q},1-\overline{q}]$*,*

i. if $q=0.5$*, then* $\lim_{n\to\infty}\Pr(B \text{ } wins \mid n,c,p_n^*(c,q),q)=0.5$*;*

ii. if $q>0.5$*, then* $\lim_{n\to\infty}\Pr(B \text{ } wins \mid n,c,p_n^*(c,q),q)>0.5$*;*

iii. if $q<0.5$*, then* $\lim_{n\to\infty}\Pr(B \text{ } wins \mid n,c,p_n^*(c,q),q)<0.5$*.*

Thus, the probability of a victory of the "right" candidate, namely the candidate who most likely has the support of the majority of the population, is greater than 50%. By allowing the possibility of misreporting, the neutrality prediction of the canonical pivotal voting model (that election results, on average, should be ties) disappears in favor of the majority candidate, even as n tends to infinity. This contrasts with the conclusions of Campbell (1999) and Taylor and Yildirim (2010b), where, given a truthful report of poll results, departure from the neutrality result in the limit can only occur if the voting cost (or, equivalently, the ideological component of utility) of B supporters is different from that of R supporters.[22] Once we relax the assumption of truthful poll results, even in a model with fixed and homogenous voting costs, we avoid the unrealistic prediction that both candidates have equal chances of winning the election. In fact, we should expect the majority candidate to win more often than not – not only for a fixed electorate size, but also as the electorate size grows without bound.

Back to our finite n environment, the expected cost $\mathcal{C}$ of the election is given by

$$\mathcal{C}\left(n,c,p,q\right)=nq\gamma_{B}\left(n,c,p\right)c+n\left(1-q\right)\gamma_{R}\left(n,c,p\right)c. \tag{18}$$

Thus, the welfare function $\mathcal{W}$ is given by

$$\mathcal{W}\left(n,c,p,q\right)=\mathcal{I}\left(n,c,p,q\right)-\mathcal{C}\left(n,c,p,q\right).$$

This approach, akin to Goeree and Großer (2007), differs from other welfare analyses present in the literature, in the sense that it considers an approximate welfare function ($I-\mathcal{C}$, as in Taylor and Yildirim, 2010a) not as a means of suggesting a result about the exact welfare function ($\mathcal{I}-\mathcal{C}$), but to actually prove such a result (with the aid of properties of the Skellam distribution).

[22] More precisely, Taylor and Yildirim (2010b) show that, as opposed to their finite population case, even if voting costs are stochastic, if they are drawn from a common distribution, then, no matter how large the majority of a candidate may be, asymptotically, we should expect an election tie. Both Campbell (1999) and Taylor and Yildirim (2010b) show a converse result: if the voting costs for R-supporters are first-order stochastically dominated by the voting costs for B-supporters, then, no matter how large the majority of B, one should expect a victory of R.

The next proposition shows that, for sufficiently large elections, misreporting of poll results actually generates a welfare improvement. Properties of the Skellam distribution are key in arriving at this central result of this work (see appendix B).

Proposition 7 *Given $c \in (0,1)$ and $q \in [\overline{q}, 1-\overline{q}] \setminus \{0.5\}$, for sufficiently large n,*

i. $\mathcal{C}\left(n, c, p_n^*\left(c, q\right), q\right) < \mathcal{C}\left(n, c, q, q\right)$;

ii. $\mathcal{I}\left(n, c, p_n^*\left(c, q\right), q\right) > \mathcal{I}\left(n, c, q, q\right)$;

iii. $\mathcal{W}\left(n, c, p_n^*\left(c, q\right), q\right) > \mathcal{W}\left(n, c, q, q\right)$.

The intuition behind this proposition is as follows. Suppose that $q > 0.5$. In this case, it would be best for society if candidate B was the winner of the election, as, more often than not, the number of B-citizens will be larger than the number of R-citizens. Recall that, with a truthful report ($p = q$), the expected result of the election is a tie, so that the probability that candidate B wins is 1/2. However, due to the misreporting behavior explained in Proposition 4, $p_n^*\left(c, q\right) \in (0.5, q)$, and by Proposition 2, the B-citizens (R-citizens) are going to vote with a higher (lower) probability relative to the truthful poll, so that the probability that B wins is greater than 1/2. Besides that, in absolute value, the probability of a B-vote ends up changing less than the probability of an R-vote, in such a way that the expected voting cost decreases. Thus, as the expected benefit increases with misreporting and the expected voting cost decreases, we conclude that a misreported poll is unambiguously welfare-improving relative to a truthful poll.

Finally, with Lemma 7 at our disposal, we can also check the robustness of the conclusion that the misreporting of polls is welfare-improving, by showing that it holds even as the size of the electorate grows without bound. This is facilitated by our proof of Proposition 7 which, given the asymptotic nature of the approximation of $\mathcal{I}$ via I, naturally calls for the taking of limits as $n \to \infty$ (in opposition to the fixed n approach in Proposition 4 of Goeree and Großer, 2007).

Proposition 8 *Given $c \in (0,1)$ and $q \in [\overline{q}, 1-\overline{q}] \setminus \{0.5\}$,*

i. $\lim_{n\to\infty} (\mathcal{C}(n, c, p_n^*(c,q), q)/n) = \lim_{n\to\infty} (\mathcal{C}(n, c, q, q)/n) = 0$*;*

ii. $\lim_{n\to\infty} (\mathcal{I}(n, c, p_n^*(c,q), q)/n) > \lim_{n\to\infty} (\mathcal{I}(n, c, q, q)/n)$*;*

iii. $\lim_{n\to\infty} (\mathcal{W}(n, c, p_n^*(c,q), q)/n) > \lim_{n\to\infty} (\mathcal{W}(n, c, q, q)/n)$*.*

Thus, the result that misreporting of pre-elections poll results enhances the welfare of society relative to a truthful report is true not only for a finite population size, but also asymptotically.

6 Concluding remarks

In this paper, we formally introduced an electoral pollster in an two-candidate costly voting model, and have concluded that a pollster driven only by reputational reasons will underreport the expected number of supporters of the most preferred candidate. This implies that, relative a truthful report, citizens in the expected majority group will vote with more intensity; the contrary being true for citizens in the expected minority group. With misreporting, not only the chances of a victory by the most preferred candidate in society are raised, but also total election costs are reduced, thus yielding a welfare gain.

By acknowledging the possibility of nontruthful pre-election poll results, we show that, even in a model with fixed and homogenous voting costs, the candidate who is supported by most citizens wins the election with probability greater than 50%. This result also holds asymptotically, that is, for an electorate size growing without bound.

Finally, our work suggests that part of what is usually attributed to the bandwagon effect could actually be an illusion due to misreporting, in the sense that, even if the citizens' behavior generates the underdog effect instead, an observer who disregards the possibility of nontruthful polls could incorrectly conclude that the bandwagon effect was taking place.

A possible extension of this work would be to analyze the effects of different incentive schemes regarding the release of poll results, especially when there is competition among different pollsters. Also of interest would be an analysis of the behavioral and welfare effects in the case of multiple simultaneous state elections (for governors or legislative representatives), with or without state-dependent q variables.

The present analysis can also be linked to the issue of voluntary vs. compulsory voting, studied in Börgers (2004) and Krasa and Polborn (2009). This could be achieved by letting a social planner aware of the pollster's misreporting behavior choose c as to maximize welfare.

Finally, Campbell (1999) provides a model which favors the electability of the minority group, due to an asymmetry in the ideological component of both groups' payoffs. Whether the pollster's report bias under such circumstances, as well as the resulting welfare implications, would be strengthened or weakened, is left for future research.

References

BISCHOPING, K. and SCHUMAN, H. (1992), "Pens and Polls in Nicaragua: An Analysis of the 1990 Preelection Surveys", *American Journal of Political Science*, **36**, 331–350.

BÖRGERS, T. (2004), "Costly Voting", *American Economic Review*, **94**, 57–66.

BURSZTYN, L., CANTONI, D., FUNK, P. and YUCHTMAN, N. (2018), "Polls, the Press, and Political Participation: The Effects of Anticipated Election Closeness on Voter Turnout" (Working Paper 23490, National Bureau of Economic Research).

CAMPBELL, C. (1999), "Large Electorates and Decisive Minorities", *Journal of Political Economy*, **107**, 1199–1217.

GOEREE, J. K. and GROßER, J. (2007), "Welfare Reducing Polls", *Economic Theory*, **31**, 51–68.

GRILLO, A. (2017), "Risk Aversion and Bandwagon Effect in the Pivotal Voter Model", *Public Choice*, **172**, 465–482.

JOHNSON, N. L., KOTZ, S. and BALAKRISHNAN, L. (1997), *Discrete Multivariate Distributions* (New York: John Wiley & Sons).

KRASA, S. and POLBORN, M. K. (2009), "Is Mandatory Voting Better than Voluntary Voting?", *Games and Economic Behavior*, **66**, 275–291.

LEDYARD, J. O. (1984), "The Pure Theory of Large Two-candidate Elections", *Public Choice*, **44**, 7–41.

MORTON, R. B. and OU, K. (2015), "What Motivates Bandwagon Voting Behavior: Altruism or a Desire to Win?", *European Journal of Political Economy*, **40B**, 224–241.

PALFREY, T. R. and ROSENTHAL, H. (1983), "A Strategic Calculus of Voting", *Public Choice*, **41**, 7–53.

PALFREY, T. R. and ROSENTHAL, H. (1985), "Voter Participation and Strategic Uncertainty", *American Political Science Review*, **79**, 62–78.

RUDIN, W. (1976), *Principles of Mathematical Analysis*, 3rd ed. (New York: McGraw-Hill).

SCHOCHETMAN, I. E. (1990), "Pointwise Versions of the Maximum Theorem with Applications in Optimization", *Applied Mathematics Letters*, **3**, 89–92.

SKELLAM, J. G. (1946), "The Frequency Distribution of the Difference between Two Poisson Variates Belonging to Different Populations", *Journal of the Royal Statistical Society*, **109**, 296.

TAYLOR, C. R. and YILDIRIM, H. (2010a), "Public Information and Electoral Bias", *Games and Economic Behavior*, **68**, 353–375.

TAYLOR, C. R. and YILDIRIM, H. (2010b), "A Unified Analysis of Rational Voting with Private Values and Group-specific Costs", *Games and Economic Behavior*, **70**, 457–471.

WADSWORTH, G. P. and BRYAN, J. G. (1974), *Applications of Probability and Random Variables*, 2nd ed. (New York: McGraw-Hill).

Appendix A

This appendix proves only the basic results regarding the pivotal voting model as presented in sections 2 and 3 (that is, prior to our discussion of the pollster's optimizing behavior). Unfortunately, it is not always true that the more basic result has the shortest proof. In this way, let the reader interested primarily in understanding the pollster's misreporting behavior and its welfare implications – to be tackled in appendix B –, be advised that the technical results stated in the present appendix as "claims" will not be invoked in any of the proofs of appendix B. Only "lemmas" and "propositions" herein will be called upon in that appendix.

In order to obtain the characterization of the electoral equilibrium, take the point of view of a citizen that prefers candidate B, and suppose that this citizen votes with probability $\tilde{\gamma} \in [0,1]$, while the other $n-1$ citizens are voting with probabilities γ (B-citizens) and δ (R-citizens). Firstly, note that, as this citizen votes with probability $\tilde{\gamma}$, the expected voting cost in which he/she incurs is given by $\tilde{\gamma}c$. Moreover, note that the electoral result depends on the votes for candidates B and R that come from the other $n-1$ citizens. Denote by Δ the difference between the number of votes that candidate B and R received from the other citizens. There are four relevant events:

i) $\Delta \geq 1$, in which case the B candidate wins and the ideological component of our citizen's utility is equal to 1;

ii) $\Delta \leq -2$, in which case the R candidate wins and the ideological component of our citizen's utility is equal to -1;

iii) $\Delta = 0$, in which case, with probability $\tilde{\gamma}$ our citizen will break the tie and thus the ideological component of his/her utility will be 1. With probability $1-\tilde{\gamma}$, as the tie will be broken by the toss of a fair coin, each candidate will be elected with probability $1/2$, which implies that the expected ideological component of his/her utility will be 0;

iv) $\Delta = -1$, in which case, with probability $\tilde{\gamma}$ our citizen creates a tie, so that the ideological component of his/her utility will be 0, and with probability $1-\tilde{\gamma}$ our candidate does not vote, implying that candidate R wins the election and the ideological component of our citizen's utility will be -1.

If $p \in (0,1)$ is the believed probability that a citizen favors candidate B, then the multinomial probabilities associated to each of these four events are, in order:

$$S_1 \;:=\; \sum_{\substack{b,r=0\\ b+r\leq n-1\\ b\geq r+1}} \binom{n-1}{b,r} (p\gamma)^b \left((1-p)\,\delta\right)^r \left(1-p\gamma-(1-p)\,\delta\right)^{n-1-b-r};$$

$$S_2 \;:=\; \sum_{\substack{b,r=0\\ b+r\leq n-1\\ b\leq r-2}} \binom{n-1}{b,r} (p\gamma)^b \left((1-p)\,\delta\right)^r \left(1-p\gamma-(1-p)\,\delta\right)^{n-1-b-r};$$

$$S_3 \;:=\; \sum_{\substack{b,r=0\\ b+r\leq n-1\\ b=r}} \binom{n-1}{b,r} (p\gamma)^b \left((1-p)\,\delta\right)^r \left(1-p\gamma-(1-p)\,\delta\right)^{n-1-b-r};$$

$$S_4 \;:=\; \sum_{\substack{b,r=0\\ b+r\leq n-1\\ b=r-1}} \binom{n-1}{b,r} (p\gamma)^b \left((1-p)\,\delta\right)^r \left(1-p\gamma-(1-p)\,\delta\right)^{n-1-b-r}.$$

Taking these probabilities into account, the expected utility of this citizen will be given by multiplying them, in turn, by 1, -1, $\tilde{\gamma}$ and $-(1-\tilde{\gamma})$, adding them all up and then subtracting $\tilde{\gamma}c$. This generates

$$\begin{aligned} V_B\left(n,c,p,\tilde{\gamma},\gamma,\delta\right) &= S_1 - S_2 + \tilde{\gamma}S_3 - (1-\tilde{\gamma})\,S_4 - \tilde{\gamma}c \\ &= S_1 - S_2 - S_4 + \tilde{\gamma}\left(-c + S_3 + S_4\right) \\ &= \xi_1 + \tilde{\gamma}\left(-c + \Pi_B\left(n,p,\gamma,\delta\right)\right), \end{aligned}$$

where $\xi_1 := S_1 - S_2 - S_4$ is constant in $\tilde{\gamma}$, and $S_3 + S_4$ is simply another way of writing the probability that this B-citizen's vote is pivotal in the election, i.e., $\Pi_B\left(n,p,\gamma,\delta\right)$.

Likewise, the expected utility of a citizen that prefers candidate R and votes with probability $\tilde{\delta}$ is given by

$$V_R\left(n,c,p,\tilde{\delta},\gamma,\delta\right) = \xi_2 + \tilde{\delta}\left(-c + \Pi_R\left(n,p,\gamma,\delta\right)\right).$$

By definition, (γ,δ) is a type-symmetric Nash equilibrium if and only if each B-citizen (R-citizen) chooses to vote with probability γ (δ) given that all other B-citizens are voting with probability γ and all other R-citizens are voting with probability δ – i.e., if and only if $\gamma \in \arg\max_{\tilde{\gamma}\in[0,1]} V_B\left(n,c,p,\tilde{\gamma},\gamma,\delta\right)$ and $\delta \in \arg\max_{\tilde{\delta}\in[0,1]} V_R\left(n,c,p,\tilde{\delta},\gamma,\delta\right)$. Since $\tilde{\delta}$ ($\tilde{\gamma}$) does not enter the first (second) of these maximization problems, we can rewrite these conditions in a simpler way by defining $\Psi\left(n,c,p,\tilde{\gamma},\tilde{\delta},\gamma,\delta\right) := V_B\left(n,c,p,\tilde{\gamma},\gamma,\delta\right) + V_R\left(n,c,p,\tilde{\delta},\gamma,\delta\right)$ and stating that (γ,δ) solves the following problem:

$$\max_{(\tilde{\gamma},\tilde{\delta})\in[0,1]\times[0,1]} \Psi\left(n,c,p,\tilde{\gamma},\tilde{\delta},\gamma,\delta\right).$$

Thus, (γ, δ) is a type-symmetric Nash equilibrium if and only if (γ, δ) belongs to the arg max of the problem above.

The solution of this problem yields nine regions of the $[0,1] \times [0,1]$ square where a type-symmetric electoral equilibrium (γ, δ) might, in principle, be located: the interior, the sides (vertices excluded) and the vertices. Since Ψ is affine in $\tilde{\gamma}$ and $\tilde{\delta}$, the conditions that characterize each of these type-symmetric equilibrium possibilities are as follows:

$\mathbf{E}_{\gamma\delta}$. $(\gamma, \delta) \in (0,1) \times (0,1)$: $\Pi_B(n, p, \gamma, \delta) = \Pi_R(n, p, \gamma, \delta) = c$;

$\mathbf{E}_{\gamma 0}$. $(\gamma, \delta) \in (0,1) \times \{0\}$: $\Pi_B(n, p, \gamma, \delta) = c$ and $\Pi_R(n, p, \gamma, \delta) \leq c$;

$\mathbf{E}_{\gamma 1}$. $(\gamma, \delta) \in (0,1) \times \{1\}$: $\Pi_B(n, p, \gamma, \delta) = c$ and $\Pi_R(n, p, \gamma, \delta) \geq c$;

$\mathbf{E}_{0\delta}$. $(\gamma, \delta) \in \{0\} \times (0,1)$: $\Pi_B(n, p, \gamma, \delta) \leq c$ and $\Pi_R(n, p, \gamma, \delta) = c$;

$\mathbf{E}_{00}$. $(\gamma, \delta) \in \{(0,0)\}$: $\Pi_B(n, p, \gamma, \delta) \leq c$ and $\Pi_R(n, p, \gamma, \delta) \leq c$;

$\mathbf{E}_{01}$. $(\gamma, \delta) \in \{(0,1)\}$: $\Pi_B(n, p, \gamma, \delta) \leq c$ and $\Pi_R(n, p, \gamma, \delta) \geq c$;

$\mathbf{E}_{1\delta}$. $(\gamma, \delta) \in \{1\} \times (0,1)$: $\Pi_B(n, p, \gamma, \delta) \geq c$ and $\Pi_R(n, p, \gamma, \delta) = c$;

$\mathbf{E}_{10}$. $(\gamma, \delta) \in \{(1,0)\}$: $\Pi_B(n, p, \gamma, \delta) \geq c$ and $\Pi_R(n, p, \gamma, \delta) \leq c$;

$\mathbf{E}_{11}$. $(\gamma, \delta) \in \{(1,1)\}$: $\Pi_B(n, p, \gamma, \delta) \geq c$ and $\Pi_R(n, p, \gamma, \delta) \geq c$.

The following claims on properties of the functions Π_B and Π_R and on the conditions for emergence of different equilibrium types will be key in the proof of Proposition 1. Whenever not mentioned explicitly, we are considering the electoral game in which $n \geq 2$ is the number of citizens and $c \in \mathbb{R}_+$ is the voting cost.

Claim 1 *Given $n \geq 2$, $p \in (0,1)$ and $(\gamma, \delta) \in [0,1]^2$, we have $\Pi_R(n, p, \gamma, \delta) = \Pi_B(n, 1-p, \delta, \gamma)$.*

Proof. This can be checked immediately from (1) and (2). ∎

Claim 2 *Given $n \geq 2$, $c \in \mathbb{R}_+$ and $p \in (0,1)$, (γ, δ) is a type-symmetric equilibrium of the electoral game with these parameter values if and only if (δ, γ) is a type-symmetric equilibrium of the electoral game with parameters values n, c and $1-p$.*

Proof. This follows immediately from the nine cases considered above and Claim 1. ∎

Claim 3 *$E_{\gamma 0}$-, $E_{0\delta}$-, E_{01}- and E_{10}-type electoral equilibria do not occur.*

Proof. Let $\delta \in (0,1]$. Note from (1) and Claim 1 that

$$\Pi_B(n,p,0,\delta) = (1-(1-p)\delta)^{n-1} + (n-1)((1-p)\delta)(1-(1-p)\delta)^{n-2}$$

and

$$\Pi_R(n,p,0,\delta) = \Pi_B(n,1-p,\delta,0) = (1-(1-p)\delta)^{n-1}.$$

Since $(1-p)\delta < 1$, we have $\Pi_B(n,p,0,\delta) - \Pi_R(n,p,0,\delta) = (n-1)((1-p)\delta)(1-(1-p)\delta)^{n-2} > 0$. This rules out both $E_{0\delta}$- and E_{01}-type equilibria.

Having established the impossibility of occurrence of these equilibrium types, $E_{\gamma 0}$- and E_{10}-type equilibria are also impossible, by Claim 2. ■

Claim 4 *An E_{00}-type electoral equilibrium occurs if and only if $c \geq 1$.*

Proof. By (1) and (2), we have

$$\Pi_B(n,p,0,0) = \binom{n-1}{0,0,n-1} 0^0 0^0 1^{n-1} = 1 \times 1 \times 1 \times 1 = 1 \tag{19}$$

and

$$\Pi_R(n,p,0,0) = \binom{n-1}{0,0,n-1} 0^0 0^0 1^{n-1} = 1 \times 1 \times 1 \times 1 = 1.$$

Thus, the necessary and sufficient condition for an E_{00}-type equilibrium, $\Pi_B(n,p,\gamma,\delta) \leq c$ and $\Pi_R(n,p,\gamma,\delta) \leq c$, becomes simply $c \geq 1$. ■

Claim 5 *If (γ,δ) is an $E_{\gamma\delta}$-type electoral equilibrium, then $\delta = (p/(1-p))\gamma$.*

Proof. In such an electoral equilibrium, we must have, by (1),

$$\begin{aligned}
c &= \Pi_B(n,p,\gamma,\delta) = \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} \binom{n-1}{k,k,n-1-2k} (p\gamma)^k ((1-p)\delta)^k (1-p\gamma-(1-p)\delta)^{n-1-2k} \\
&\quad + \sum_{k=0}^{\lfloor \frac{n-2}{2} \rfloor} \binom{n-1}{k,k+1,n-2-2k} (p\gamma)^k ((1-p)\delta)^{k+1} (1-p\gamma-(1-p)\delta)^{n-2-2k} \\
&= \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} \binom{n-1}{k,k,n-1-2k} (p\gamma)^k ((1-p)\delta)^k (1-p\gamma-(1-p)\delta)^{n-1-2k} \\
&\quad + (1-p)\delta \sum_{k=0}^{\lfloor \frac{n-2}{2} \rfloor} \binom{n-1}{k,k+1,n-2-2k} (p\gamma)^k ((1-p)\delta)^k (1-p\gamma-(1-p)\delta)^{n-2-2k}
\end{aligned}$$

and

$$
\begin{aligned}
c &= \Pi_R(n,p,\gamma,\delta) = \Pi_B(n,1-p,\delta,\gamma) \\
&= \sum_{k=0}^{\left\lfloor \frac{n-1}{2} \right\rfloor} \binom{n-1}{k,k,n-1-2k} ((1-p)\,\delta)^k (p\gamma)^k (1-(1-p)\,\delta - p\gamma)^{n-1-2k} \\
&\quad + p\gamma \sum_{k=0}^{\left\lfloor \frac{n-2}{2} \right\rfloor} \binom{n-1}{k,k+1,n-2-2k} ((1-p)\,\delta)^k (p\gamma)^k (1-(1-p)\,\delta - p\gamma)^{n-2-2k},
\end{aligned}
$$

where we have applied Claim 1.

In more concise notation, we can write $c = \kappa + (1-p)\,\delta\lambda$ and $c = \kappa + p\gamma\lambda$, where

$$
\begin{aligned}
\lambda &:= \sum_{k=0}^{\left\lfloor \frac{n-2}{2} \right\rfloor} \binom{n-1}{k,k+1,n-2-2k} (p\gamma)^k ((1-p)\,\delta)^k (1-p\gamma-(1-p)\,\delta)^{n-2-2k} \\
&\geq \binom{n-1}{0,1,n-2} (p\gamma)^0 ((1-p)\,\delta)^0 (1-p\gamma-(1-p)\,\delta)^{n-2} = (n-1)(1-p\gamma-(1-p)\,\delta)^{n-2},
\end{aligned}
$$

which is positive since $p\gamma + (1-p)\,\delta$ is an average between γ and δ, both of which are lower than 1. Therefore, $(1-p)\,\delta = (c-\kappa)/\lambda = p\gamma$, and the thesis follows. ■

Claim 6 *If $(\gamma, 1)$ is an $E_{\gamma 1}$-type electoral equilibrium, then $p > 0.5$ and $\gamma \geq (1-p)/p$. If $(1, \delta)$ is an $E_{1\delta}$-type electoral equilibrium, then $p < 0.5$ and $\delta \geq p/(1-p)$.*

Proof. First note that, by (1) and (2), we have

$$
\begin{aligned}
\Pi_B(n,p,\gamma,1) &= \sum_{k=0}^{\left\lfloor \frac{n-1}{2} \right\rfloor} \binom{n-1}{k,k,n-1-2k} (p\gamma)^k (1-p)^k (p(1-\gamma))^{n-1-2k} \\
&\quad + \sum_{k=0}^{\left\lfloor \frac{n-2}{2} \right\rfloor} \binom{n-1}{k,k+1,n-2-2k} (p\gamma)^k (1-p)^{k+1} (p(1-\gamma))^{n-2-2k}
\end{aligned}
$$

and

$$
\begin{aligned}
\Pi_R(n,p,\gamma,1) &= \sum_{k=0}^{\left\lfloor \frac{n-1}{2} \right\rfloor} \binom{n-1}{k,k,n-1-2k} (1-p)^k (p\gamma)^k (p(1-\gamma))^{n-1-2k} \\
&\quad + \sum_{k=0}^{\left\lfloor \frac{n-2}{2} \right\rfloor} \binom{n-1}{k,k+1,n-2-2k} (1-p)^k (p\gamma)^{k+1} (p(1-\gamma))^{n-2-2k}.
\end{aligned}
$$

Now assume $(\gamma, 1)$ is an $\mathrm{E}_{\gamma 1}$-type electoral equilibrium. Thus $\Pi_B(n,p,\gamma,1) = c \leq \Pi_R(n,p,\gamma,1)$, which implies

$$\begin{aligned} 0 &\leq \Pi_R(n,p,\gamma,1) - \Pi_B(n,p,\gamma,1) \\ &= \sum_{k=0}^{\lfloor \frac{n-2}{2} \rfloor} \binom{n-1}{k, k+1, n-2-2k} (p\gamma)^k (1-p)^k (p\gamma - (1-p)) (p(1-\gamma))^{n-2-2k}. \end{aligned}$$

Since $\gamma \in (0,1)$, the terms $(p\gamma)^k$ and $(p(1-\gamma))^{n-2-2k}$ are positive, so that the above inequality can only be true if $p\gamma - (1-p) \geq 0$, that is, $\gamma \geq (1-p)/p$. Since $\gamma < 1$, this also implies $(1-p)/p < 1$, that is, $p > 0.5$.

Finally, if $(1,\delta)$ is an $\mathrm{E}_{1\delta}$-type electoral equilibrium of the game with parameter p, then, by Claim 2, $(\delta, 1)$ is a type-symmetric equilibrium of the game with parameter $1-p$. As seen above, this implies $1-p > 0.5$ (that is, $p < 0.5$) and $\delta \geq (1-(1-p))/(1-p) = p/(1-p)$. ■

Claim 7 *Given $n \geq 2$ and $p \in [0.5, 1)$, we have $\Pi_R(n,p,\gamma,1) = \Pi_B(n,p,\gamma,1)$ if $\gamma = (1-p)/p$, $\Pi_R(n,p,\gamma,1) > \Pi_B(n,p,\gamma,1)$ if $\gamma \in ((1-p)/p, 1)$, and $\Pi_R(n,p,\gamma,1) \geq \Pi_B(n,p,\gamma,1)$ if $\gamma = 1$.*

Proof. Expressions (1) and (2) yield

$$\begin{aligned} & \Pi_R(n,p,\gamma,1) - \Pi_B(n,p,\gamma,1) \\ = & \sum_{k=0}^{\lfloor \frac{n-2}{2} \rfloor} \binom{n-1}{k, k+1, n-2-2k} (1-p)^k (p\gamma)^{k+1} (1-p\gamma-(1-p))^{n-2-2k} \\ & - \sum_{k=0}^{\lfloor \frac{n-2}{2} \rfloor} \binom{n-1}{k, k+1, n-2-2k} (p\gamma)^k (1-p)^{k+1} (1-p\gamma-(1-p))^{n-2-k} \\ = & \sum_{k=0}^{\lfloor \frac{n-2}{2} \rfloor} \binom{n-1}{k, k+1, n-2-2k} (p\gamma)^k (1-p)^k (1-p\gamma-(1-p))^{n-2-2k} (p\gamma-(1-p)) \\ = & (p\gamma-(1-p)) \sum_{k=0}^{\lfloor \frac{n-2}{2} \rfloor} \binom{n-1}{k, k+1, n-2-2k} (p\gamma)^k (1-p)^k (p(1-\gamma))^{n-2-2k}. \end{aligned}$$

The result follows immediately from this computation. One may note that, in case $\gamma = 1$, the above expression is zero if $p = 0.5$ or n is odd, and positive otherwise. ■

Given $n \geq 2$, let $\Upsilon_n : [0.5, 1) \times [0,1] \to \mathbb{R}$ be given by

$$\Upsilon_n(p,\gamma) = \begin{cases} \Pi_B\left(n,p,\gamma,\frac{p}{1-p}\gamma\right) & \text{if } \gamma \in \left[0, \frac{1-p}{p}\right) \\ \Pi_B(n,p,\gamma,1) & \text{if } \gamma \in \left[\frac{1-p}{p}, 1\right] \end{cases}.$$

As a consequence of the previous claims, we have

Claim 8 *Given $n \geq 2$, $c \in \mathbb{R}_+$ and $p \in [0.5, 1)$, the conditions that characterize the four kinds of type-symmetric electoral equilibria that may emerge can be rewritten more simply as:*

$\mathbf{E}_{11}$**.** $(\gamma, \delta) \in \{(1, 1)\}$ and $\Upsilon_n(p, \gamma) \geq c$;

$\mathbf{E}_{\gamma 1}$**.** $(\gamma, \delta) \in [(1-p)/p, 1) \times \{1\}$ and $\Upsilon_n(p, \gamma) = c$;

$\mathbf{E}_{\gamma\delta}$**.** $(\gamma, \delta) \in (0, (1-p)/p) \times (0, 1)$, $\delta = (p/(1-p))\gamma$ and $\Upsilon_n(p, \gamma) = c$;

$\mathbf{E}_{00}$**.** $(\gamma, \delta) \in \{(0, 0)\}$ and $\Upsilon_n(p, \gamma) \leq c$.

Proof. That these are the only four possibilities for a type-symmetric electoral equilibria for a $p \in [0.5, 1)$ follows from Claims 3 and 6.

Note that $(1, 1)$ is an E_{11}-type electoral equilibrium if and only if $\Pi_B(n, p, 1, 1) \geq c$ and $\Pi_R(n, p, 1, 1) \geq c$, which occurs if and only if $\Pi_B(n, p, 1, 1) \geq c$ (Claim 7 guarantees $\Pi_R(n, p, 1, 1) \geq \Pi_B(n, p, 1, 1)$), that is, if and only if $\Upsilon_n(p, 1) \geq c$.

In the same way, $(\gamma, 1)$ is an $\mathrm{E}_{\gamma 1}$-type electoral equilibrium if and only if $\gamma \in (0, 1)$, $\Pi_B(n, p, \gamma, 1) = c$ and $\Pi_R(n, p, \gamma, 1) \geq c$, which occurs if and only if $\gamma \in [(1-p)/p, 1)$ (by Claim 6) and $\Pi_B(n, p, \gamma, 1) = c$ (Claim 7 guarantees $\Pi_R(n, p, \gamma, 1) \geq \Pi_B(n, p, \gamma, 1)$ for $\gamma \geq (1-p)/p$), that is, if and only if $\gamma \in [(1-p)/p, 1)$ and $\Upsilon_n(p, \gamma) = c$.

As for the interior equilibrium case, (γ, δ) is an $\mathrm{E}_{\gamma\delta}$-type equilibrium if and only if $(\gamma, \delta) \in (0, 1) \times (0, 1)$, $\Pi_B(n, p, \gamma, \delta) = c$ and $\Pi_R(n, p, \gamma, \delta) = c$, which occurs if and only if $\gamma \in (0, (1-p)/p)$, $\delta = (p/(1-p))\gamma$ (by Claim 5) and $\Pi_B(n, p, \gamma, \delta) = c$, since, by (1) and (2),

$$\Pi_R\left(n, p, \gamma, \frac{p}{1-p}\gamma\right) = \sum_{k=0}^{\left\lfloor \frac{n-1}{2} \right\rfloor} \binom{n-1}{k, k, n-1-2k} (p\gamma)^k (p\gamma)^k (1-2p\gamma)^{n-1-2k}$$

$$+ \sum_{k=0}^{\left\lfloor \frac{n-2}{2} \right\rfloor} \binom{n-1}{k, k+1, n-2-2k} (p\gamma)^k (p\gamma)^{k+1} (1-2p\gamma)^{n-2-2k} = \Pi_B\left(n, p, \gamma, \frac{p}{1-p}\gamma\right).$$

In other words, if and only if $\gamma \in (0, (1-p)/p)$, $\delta = (p/(1-p))\gamma$ and $\Upsilon_n(p, \gamma) = c$.

Finally, $(0, 0)$ is an E_{00}-type electoral equilibrium if and only if $\Pi_B(n, p, 0, 0) \leq c$ and $\Pi_R(n, p, 0, 0) \leq c$, which occurs if and only if $\Pi_B(n, p, 0, 0) \leq c$ (the proof of Claim 4 shows that $\Pi_R(n, p, 0, 0) = 1 = \Pi_B(n, p, 0, 0)$), that is, if and only if $\Upsilon_n(p, 0) \leq c$. ■

A function that appears in the above proof will appear many other times in this appendix, making it useful to dedicate a separate lemma to its behavior.

Lemma 8 *The function* $P_n : [0, 0.5] \to \mathbb{R}$ *given by* $P_n(\alpha) = \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} \binom{n-1}{k,k,n-1-2k} \alpha^{2k} (1-2\alpha)^{n-1-2k} + \sum_{k=0}^{\lfloor \frac{n-2}{2} \rfloor} \binom{n-1}{k,k+1,n-2-2k} \alpha^{2k+1} (1-2\alpha)^{n-2-2k}$ *is continuous on* $[0, 0.5]$ *and continuously differentiable on* $(0, 0.5)$, *where* $P_n' < 0$.

Proof. Continuity and continuous differentiability follow from the simple fact that P_n is a polynomial function. That its derivative is negative is proved in Lemma 2 of Taylor and Yildirim (2010a). ■

The conditions stated in Claim 8 are more convenient to work with than those involving Π_B and Π_R derived at the beginning of this appendix, since they depend on one function (and with one less argument) only. It may also be noted, by (3) and (4), that, for any $p \in [0.5, 1)$,

$$\underline{c}_n(p) = \Upsilon_n(p, 1) \tag{20}$$

and

$$\bar{c}_n(p) = \Upsilon_n(p, (1-p)/p) = P_n(1-p). \tag{21}$$

Continuity of Υ_n is clear: for all $(p, \gamma) \in [0.5, 1) \times [0, 1]$, we can write $\Upsilon_n(p, \gamma) = \Pi_B(n, p, \gamma, \min(1, (p/(1-p))\gamma))$, composed of continuous functions only (recall $\Pi_B(n, p, \cdot, \cdot)$ is polynomial). Differentiability of Υ_n is much less clear – nonetheless, it will be proven in Claims 10 and 11 ahead, with aid from the following basic fact from real analysis.

Claim 9 *Given an open interval* I, $\bar{x} \in I$, *and a continuous* $f : I \to \mathbb{R}$ *that is differentiable over* $I \setminus \{\bar{x}\}$, *if* $\lim_{x\to\bar{x}} f'(x) = L \in \mathbb{R}$, *then* f *is differentiable at* $\bar{x}$, *and* $f'(\bar{x}) = L$.

Proof. Given $\varepsilon > 0$, since I is open and $\lim_{x\to\bar{x}} f'(x) = L$, we know that there exists $\delta > 0$ such that $0 < |x - \bar{x}| < \delta$ implies $x \in I$ and $|f'(x) - L| < \varepsilon$. Now, for any x such that $0 < |x - \bar{x}| < \delta$, since f is continuous over $[\min(\bar{x}, x), \max(\bar{x}, x)]$ and differentiable over $(\min(\bar{x}, x), \max(\bar{x}, x))$, the Mean Value Theorem yields existence of a $x_m \in (\min(\bar{x}, x), \max(\bar{x}, x))$ such that $f'(x_m) = (f(x) - f(\bar{x}))/(x - \bar{x})$. Finally, because we will have $0 < |x_m - \bar{x}| < |x - \bar{x}| < \delta$, $|(f(x) - f(\bar{x}))/(x - \bar{x}) - L| = |f'(x_m) - L| < \varepsilon$. ■

Claim 10 *Given* $n \geq 2$, *we have* $\partial\Upsilon_n(p, \gamma)/\partial\gamma < 0$, *for all* $(p, \gamma) \in [0.5, 1) \times (0, 1)$.

Proof. Let $(p,\gamma) \in [0.5,1) \times (0,1)$. In case $\gamma \in (0,(1-p)/p)$, then, as noted in the proof of Claim 8,

$$\begin{aligned}\Upsilon_n(p,\gamma) &= \Pi_B\left(n,p,\gamma,\frac{p}{1-p}\gamma\right) = \sum_{k=0}^{\lfloor\frac{n-1}{2}\rfloor}\binom{n-1}{k,k,n-1-2k}(p\gamma)^{2k}(1-2p\gamma)^{n-1-2k} \qquad (22)\\ &+ \sum_{k=0}^{\lfloor\frac{n-2}{2}\rfloor}\binom{n-1}{k,k+1,n-2-2k}(p\gamma)^{2k+1}(1-2p\gamma)^{n-2-2k},\end{aligned}$$

a polynomial in (p,γ), hence continuously differentiable (this observation will be used later, in the proof of Claim 12). The fact that the partial derivative $\partial\Upsilon_n(p,\gamma)/\partial\gamma$ is negative follows from Lemma 8: since $p\gamma \in (0,1-p) \subseteq (0,0.5)$, we can write $\Upsilon_n(p,\gamma) = P_n(p\gamma)$, so that, by the Chain Rule, $\partial\Upsilon_n(p,\gamma)/\partial\gamma = pP'(p\gamma) < 0$.

If $p = 0.5$, then $(1-p)/p = 1$ and the proof is finished. If $p > 0.5$, then $(1-p)/p = 1/p - 1 < 1/0.5 - 1 = 1$, so that we still have to consider the possibility that $\gamma \in [(1-p)/p, 1)$.

If $\gamma \in ((1-p)/p,1)$, then, by (1),

$$\begin{aligned}\Upsilon_n(p,\gamma) &= \Pi_B(n,p,\gamma,1) = \sum_{k=0}^{\lfloor\frac{n-1}{2}\rfloor}\binom{n-1}{k,k,n-1-2k}(p\gamma)^{k}(1-p)^{k}(p(1-\gamma))^{n-1-2k} \qquad (23)\\ &+ \sum_{k=0}^{\lfloor\frac{n-2}{2}\rfloor}\binom{n-1}{k,k+1,n-2-2k}(p\gamma)^{k}(1-p)^{k+1}(p(1-\gamma))^{n-2-2k},\end{aligned}$$

again a polynomial, hence continuously differentiable. In this case, the fact that $\partial\Upsilon_n(p,\gamma)/\partial\gamma < 0$ will follow from Lemma 1, part (iii) of Taylor and Yildirim (2010b). There, it is shown that $\partial P(\alpha_B,\alpha_R,n)/\partial\alpha_B < 0$ if $\alpha_B \in (0,p)$, $\alpha_R \in (0,1-p]$ and $\alpha_B \geq \left(1-\lfloor n/2\rfloor^{-1}\right)\alpha_R$, where their α_B is our $p\gamma$, their α_R is our $(1-p)\delta$, and their $P(\alpha_B,\alpha_R,n)$ is our $\Pi_B(n,p,\gamma,\delta)$.[23] By putting $\delta = 1$, we have $\alpha_B = p\gamma \in (1-p,p) \subseteq (0,p)$, $\alpha_R = 1-p \in (0,1-p]$ and $\alpha_B = p\gamma > p(1-p)/p = 1-p > \left(1-\lfloor n/2\rfloor^{-1}\right)(1-p)$, whence the result from their paper follows and, by the Chain Rule, $\partial\Upsilon_n(p,\gamma)/\partial\gamma = \partial\Pi_B(n,p,\gamma,1)/\partial\gamma = p\partial P(\alpha_B,\alpha_R,n)/\partial\alpha_B < 0$.

Finally, let us consider the $\gamma = \bar{\gamma} := (1-p)/p$ case. If we show that $\lim_{\gamma\to\bar{\gamma}}\partial\Upsilon_n(p,\gamma)/\partial\gamma$ exists and equals, say, L (which will be shown to be negative), then $\partial\Upsilon_n(p,\bar{\gamma})/\partial\gamma$ will also exist and equal L, by Claim 9 applied to function $\Upsilon_n(p,\cdot)$ (which is continuous over $(0,1)$ and, as seen above, differentiable at least over $(0,1)\setminus\{\bar{\gamma}\}$). The reason why we choose to show this stronger property is that it will turn out useful later on, in the proof of Claim 12.

For any $\gamma \in (0,\bar{\gamma})$, $\Upsilon_n(p,\gamma) = \Pi_B(n,p,\gamma,(p/(1-p))\gamma)$. Therefore, the left-hand side limit

[23] Although their lemma is stated for $\alpha_R \in (0,1-p)$, their proof still holds perfectly if $\alpha_R = 1-p$, that is, $\delta = 1$ (even their last step, which requires noting that $1-\alpha_B-\alpha_R \neq 0$), as long as $\alpha_B \neq p$.

$\lim_{\gamma \to \bar{\gamma}-} \partial \Upsilon_n (p, \gamma) / \partial \gamma$ can be obtained by first differentiating (22) with respect to γ, which yields

$$\begin{aligned}
& p \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} \frac{(n-1)!}{k!k!(n-1-2k)!} 2k (p\gamma)^{2k-1} (1-2p\gamma)^{n-1-2k} \\
& -p \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} \frac{(n-1)!}{k!k!(n-1-2k)!} 2(n-1-2k)(p\gamma)^{2k} (1-2p\gamma)^{n-2-2k} \\
& +p \sum_{k=0}^{\lfloor \frac{n-2}{2} \rfloor} \frac{(n-1)!}{k!(k+1)!(n-2-2k)!} (2k+1)(p\gamma)^{2k} (1-2p\gamma)^{n-2-2k} \\
& -p \sum_{k=0}^{\lfloor \frac{n-2}{2} \rfloor} \frac{(n-1)!}{k!(k+1)!(n-2-2k)!} 2(n-2-2k)(p\gamma)^{2k+1} (1-2p\gamma)^{n-3-2k},
\end{aligned} \tag{24}$$

and then taking the limit, which results in

$$\begin{aligned}
& p \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} \frac{(n-1)!}{k!k!(n-1-2k)!} 2k (1-p)^{2k-1} (2p-1)^{n-1-2k} \\
& -p \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} \frac{(n-1)!}{k!k!(n-1-2k)!} 2(n-1-2k)(1-p)^{2k} (2p-1)^{n-2-2k} \\
& +p \sum_{k=0}^{\lfloor \frac{n-2}{2} \rfloor} \frac{(n-1)!}{k!(k+1)!(n-2-2k)!} (2k+1)(1-p)^{2k} (2p-1)^{n-2-2k} \\
& -p \sum_{k=0}^{\lfloor \frac{n-2}{2} \rfloor} \frac{(n-1)!}{k!(k+1)!(n-2-2k)!} 2(n-2-2k)(1-p)^{2k+1} (2p-1)^{n-3-2k}.
\end{aligned} \tag{25}$$

Similarly, for any $\gamma \in (\bar{\gamma}, 1)$, $\Upsilon_n (p, \gamma) = \Pi_B (n, p, \gamma, 1)$. Therefore, the right-hand side limit $\lim_{\gamma \to \bar{\gamma}+} \partial \Upsilon_n (p, \gamma) / \partial \gamma$ can be obtained by first differentiating (23) with respect to γ, and then taking the limit. This results in

$$\begin{aligned}
& p \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} \frac{(n-1)!}{k!k!(n-1-2k)!} k (1-p)^{k-1} (1-p)^{k} (2p-1)^{n-1-2k} \\
& -p \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} \frac{(n-1)!}{k!k!(n-1-2k)!} (n-1-2k)(1-p)^{k} (1-p)^{k} (2p-1)^{n-2-2k} \\
& +p \sum_{k=0}^{\lfloor \frac{n-2}{2} \rfloor} \frac{(n-1)!}{k!(k+1)!(n-2-2k)!} k (1-p)^{k-1} (1-p)^{k+1} (2p-1)^{n-2-2k} \\
& -p \sum_{k=0}^{\lfloor \frac{n-2}{2} \rfloor} \frac{(n-1)!}{k!(k+1)!(n-2-2k)!} (n-2-2k)(1-p)^{k} (1-p)^{k+1} (2p-1)^{n-3-2k}.
\end{aligned} \tag{26}$$

That this is negative follows from Lemma 1, part (iii) of Taylor and Yildirim (2010b), just like in the $\gamma \in ((1-p)/p, 1)$ case (note that, since $p > 0.5$, $\alpha_B = p\gamma = 1 - p < p$).

That (25) and (26) are in fact equal can be seen by subtracting the former from the latter expression, to obtain

$$\begin{aligned}
& -p \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} \frac{(n-1)!}{k!k!(n-1-2k)!} k (1-p)^{2k-1} (2p-1)^{n-1-2k} \\
& +p \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} \frac{(n-1)!}{k!k!(n-1-2k)!} (n-1-2k) (1-p)^{2k} (2p-1)^{n-2-2k} \\
& -p \sum_{k=0}^{\lfloor \frac{n-2}{2} \rfloor} \frac{(n-1)!}{k!(k+1)!(n-2-2k)!} (k+1) (1-p)^{2k} (2p-1)^{n-2-2k} \\
& +p \sum_{k=0}^{\lfloor \frac{n-2}{2} \rfloor} \frac{(n-1)!}{k!(k+1)!(n-2-2k)!} (n-2-2k) (1-p)^{2k+1} (2p-1)^{n-3-2k} \\
= & -p \sum_{k=1}^{\lfloor \frac{n-1}{2} \rfloor} \frac{(n-1)!}{k!k!(n-1-2k)!} k (1-p)^{2k-1} (2p-1)^{n-1-2k} \\
& +p \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} \frac{(n-1)!}{k!k!(n-1-2k)!} (n-1-2k) (1-p)^{2k} (2p-1)^{n-2-2k} \\
& -p \sum_{k=0}^{\lfloor \frac{n-2}{2} \rfloor} \frac{(n-1)!}{k!k!(n-1-2k)!} (n-1-2k) (1-p)^{2k} (2p-1)^{n-2-2k} \\
& +p \sum_{k'=1}^{\lfloor \frac{n-2}{2} \rfloor + 1} \frac{(n-1)!}{(k'-1)!k'!(n-2k')!} \left(n-2k'\right) (1-p)^{2k'-1} (2p-1)^{n-1-2k'} \\
= & -p \sum_{k=1}^{\lfloor \frac{n-1}{2} \rfloor} \frac{(n-1)!}{(k-1)!k!(n-2k)!} (n-2k) (1-p)^{2k-1} (2p-1)^{n-1-2k} \\
& +p \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} \frac{(n-1)!}{k!k!(n-1-2k)!} (n-1-2k) (1-p)^{2k} (2p-1)^{n-2-2k} \\
& -p \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor - 1} \frac{(n-1)!}{k!k!(n-1-2k)!} (n-1-2k) (1-p)^{2k} (2p-1)^{n-2-2k} \\
& +p \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor} \frac{(n-1)!}{(k-1)!k!(n-2k)!} (n-2k) (1-p)^{2k-1} (2p-1)^{n-1-2k} .
\end{aligned}$$

This is zero because the first of these sums equals the fourth one (even when n is even and the fourth sum has one extra term, since this term will be zero) and the second one equals the third one (even when n is odd and the second sum has one extra term, since this term will be zero).

Thus, $\partial\Upsilon_n(p,\bar{\gamma})/\partial\gamma$ also exists, and is negative. ■

Claim 11 *Given $n \geq 2$, we have $\partial\Upsilon_n(p,\gamma)/\partial p < 0$, for all $(p,\gamma) \in (0.5,1) \times (0,1)$.*

Proof. Let $(p,\gamma) \in (0.5,1) \times (0,1)$. In case $\gamma \in (0,(1-p)/p)$, just as we noted in the proof of Claim 10, we have $p\gamma \in (0,1-p) \subseteq (0,0.5)$ and $\Upsilon_n(p,\gamma) = P_n(p\gamma)$, so that we can use Lemma 8 to conclude, using the Chain Rule, that $\partial\Upsilon_n(p,\gamma)/\partial p = \gamma P_n'(p\gamma) < 0$.

In the $\gamma \in ((1-p)/p,1)$ case, as in the proof of Claim 10, we can use Taylor and Yildirim's (2010b) Lemma 1. In their notation, $\Upsilon_n(p,\gamma) = P(p\gamma, 1-p, n)$. By the Chain Rule,

$$\frac{\partial}{\partial p}\Upsilon_n(p,\gamma) = \frac{\partial}{\partial\alpha_B}P(p\gamma,1-p,n)\gamma - \frac{\partial}{\partial\alpha_R}P(p\gamma,1-p,n),$$

where $\partial P/\partial\alpha_B$ and $\partial P/\partial\alpha_R$ are partial derivatives of P with respect to its first and second argument, respectively.

That the first of these derivatives is negative has been covered in the proof of Claim 10, where it was noted that the proof in Taylor and Yildirim (2010b) remains valid when $\delta = 1$ (that is, when $\alpha_R = 1-p$). Similarly, the expression for $\partial P(p\gamma,(1-p)\delta,n)/\partial\alpha_R$ obtained in their proof also remains valid when $\delta = 1$ and, as they show, it equals 0 if $n = 2$ and is positive (since $p\gamma - (1-p) > 0$) if $n > 2$. Thus, $\partial\Upsilon_n(p,\gamma)/\partial p < 0$ also in this case.

Finally, for the $p = \bar{p} := 1/(1+\gamma)$ (i.e., $\bar{p}$ is such that $\gamma = (1-\bar{p})/\bar{p}$) case, we mimic the discussion in the proof of Claim 10. If we show that $\lim_{p\to\bar{p}}\partial\Upsilon_n(p,\gamma)/\partial p$ exists, then $\partial\Upsilon_n(\bar{p},\gamma)/\partial p$ exists, and equals this limit (which will be shown to be negative), by Claim 9 applied to function $\Upsilon_n(\cdot,\gamma)$ (which is continuous over $(0.5,1)$ and, as seen above, differentiable at least over $(0.5,1)\setminus\{\bar{p}\}$). As noted in the proof of Claim 10, we show this stronger property because it will be useful in proving Claim 12.

For any $p \in (0.5,\bar{p})$ (so that p can be made to approach $\bar{p}$ from the left), $\Upsilon_n(p,\gamma) = \Pi_B(n,p,\gamma,(p/(1-p))\gamma)$, since $\gamma = 1/\bar{p} - 1 < 1/p - 1 = (1-p)/p$. Therefore, the left-hand side limit $\lim_{p\to\bar{p}-}\partial\Upsilon_n(p,\gamma)/\partial p$ can be obtained by first differentiating (22) with respect to p, and then taking the limit. By the symmetry of (22) with respect to p and γ, the derivative at hand will be similar to (24), the difference being that p and γ would be switching places. A quick look at (24) shows that this means that γ, instead of p, will be multiplying each of the four sums there. After plugging in $\gamma = (1-p)/p$, as seen in the proof of Claim 10, (24) becomes (25), which turns out to be equal to (26). Therefore, $\lim_{p\to\bar{p}-}\partial\Upsilon_n(p,\gamma)/\partial p$ amounts to (26) evaluated at $(\bar{p},\gamma)$, divided by $\bar{p}$

and then multiplied by $\gamma = (1-\bar{p})/\bar{p}$:

$$\begin{aligned}
&\frac{1-\bar{p}}{\bar{p}}\frac{1}{\bar{p}}\frac{\partial}{\partial\gamma}\Upsilon_n\left(\bar{p},\frac{1-\bar{p}}{\bar{p}}\right) \\
=&\sum_{k=0}^{\lfloor\frac{n-1}{2}\rfloor}\frac{(n-1)!}{k!k!(n-1-2k)!}k\bar{p}^{-1}(1-\bar{p})^{2k}(2\bar{p}-1)^{n-1-2k} \\
&-\sum_{k=0}^{\lfloor\frac{n-1}{2}\rfloor}\frac{(n-1)!}{k!k!(n-1-2k)!}(n-1-2k)\bar{p}^{-1}(1-\bar{p})^{2k+1}(2\bar{p}-1)^{n-2-2k} \\
&+\sum_{k=0}^{\lfloor\frac{n-2}{2}\rfloor}\frac{(n-1)!}{k!(k+1)!(n-2-2k)!}k\bar{p}^{-1}(1-\bar{p})^{2k+1}(2\bar{p}-1)^{n-2-2k} \\
&-\sum_{k=0}^{\lfloor\frac{n-2}{2}\rfloor}\frac{(n-1)!}{k!(k+1)!(n-2-2k)!}(n-2-2k)\bar{p}^{-1}(1-\bar{p})^{2k+2}(2\bar{p}-1)^{n-3-2k},
\end{aligned} \tag{27}$$

the negativity of which follows simply from the negativity of $\partial\Upsilon_n(p,(1-p)/p)/\partial\gamma, \forall p \in (0.5,1)$ proven in Claim 10.

For any $p \in (\bar{p},1)$ (so that p can be made to approach $\bar{p}$ from the right), $\Upsilon_n(p,\gamma) = \Pi_B(n,p,\gamma,1)$, since $\gamma = 1/\bar{p}-1 > 1/p-1 = (1-p)/p$. Therefore, the right-hand side limit $\lim_{p\to\bar{p}+}\partial\Upsilon_n(p,\gamma)/\partial p$ can be obtained by first rewriting (23) in a slightly more convenient form,

$$\begin{aligned}
\Upsilon_n(p,\gamma) \;=\; &\sum_{k=0}^{\lfloor\frac{n-1}{2}\rfloor}\binom{n-1}{k,k,n-1-2k}\gamma^k(1-\gamma)^{n-1-2k}p^{n-1-k}(1-p)^k \\
&+\sum_{k=0}^{\lfloor\frac{n-2}{2}\rfloor}\binom{n-1}{k,k+1,n-2-2k}\gamma^k(1-\gamma)^{n-2-2k}p^{n-2-k}(1-p)^{k+1},
\end{aligned}$$

then differentiating it with respect to p, to obtain

$$\begin{aligned}
&\sum_{k=0}^{\lfloor\frac{n-1}{2}\rfloor}\binom{n-1}{k,k,n-1-2k}\gamma^k(1-\gamma)^{n-1-2k}(n-1-k)p^{n-2-k}(1-p)^k \\
&-\sum_{k=0}^{\lfloor\frac{n-1}{2}\rfloor}\binom{n-1}{k,k,n-1-2k}\gamma^k(1-\gamma)^{n-1-2k}kp^{n-1-k}(1-p)^{k-1} \\
&+\sum_{k=0}^{\lfloor\frac{n-2}{2}\rfloor}\binom{n-1}{k,k+1,n-2-2k}\gamma^k(1-\gamma)^{n-2-2k}(n-2-k)p^{n-3-k}(1-p)^{k+1} \\
&-\sum_{k=0}^{\lfloor\frac{n-2}{2}\rfloor}\binom{n-1}{k,k+1,n-2-2k}\gamma^k(1-\gamma)^{n-2-2k}(k+1)p^{n-2-k}(1-p)^k,
\end{aligned}$$

and finally applying the $\lim_{p\to\bar{p}+}$ operator and substituting $\gamma = (1-\bar{p})/\bar{p}$, which yields:

$$\begin{aligned}
&\sum_{k=0}^{\left\lfloor\frac{n-1}{2}\right\rfloor} \frac{(n-1)!}{k!k!\,(n-1-2k)!}\,(n-1-k)\,\bar{p}^{-1}\,(1-\bar{p})^{2k}\,(2\bar{p}-1)^{n-1-2k} \\
&-\sum_{k=0}^{\left\lfloor\frac{n-1}{2}\right\rfloor} \frac{(n-1)!}{k!k!\,(n-1-2k)!}k\,(1-\bar{p})^{2k-1}\,(2\bar{p}-1)^{n-1-2k} \\
&+\sum_{k=0}^{\left\lfloor\frac{n-2}{2}\right\rfloor} \frac{(n-1)!}{k!\,(k+1)!\,(n-2-2k)!}\,(n-2-k)\,\bar{p}^{-1}\,(1-\bar{p})^{2k+1}\,(2\bar{p}-1)^{n-2-2k} \\
&-\sum_{k=0}^{\left\lfloor\frac{n-2}{2}\right\rfloor} \frac{(n-1)!}{k!\,(k+1)!\,(n-2-2k)!}\,(k+1)\,(1-\bar{p})^{2k}\,(2\bar{p}-1)^{n-2-2k}\,.
\end{aligned} \tag{28}$$

That (28) coincides with (27) can be seen by subtracting the latter from the former. This time, a useful start will be to subtract the two first sums in (27) from the first sum in (28), and the two last sums in (27) from the third sum in (28), while only marginally rewriting the second and fourth sums in (28). This yields

$$\begin{aligned}
& \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} \frac{(n-1)!}{k!k!\,(n-1-2k)!} (n-1-2k)\,\bar{p}^{-1} \begin{pmatrix} (1-\bar{p})^{2k} (2\bar{p}-1)^{n-1-2k} \\ +(1-\bar{p})^{2k+1} (2\bar{p}-1)^{n-2-2k} \end{pmatrix} \\
& - \sum_{k=1}^{\lfloor \frac{n-1}{2} \rfloor} \frac{(n-1)!}{k!k!\,(n-1-2k)!} k\,(1-\bar{p})^{2k-1} (2\bar{p}-1)^{n-1-2k} \\
& + \sum_{k=0}^{\lfloor \frac{n-2}{2} \rfloor} \frac{(n-1)!}{k!\,(k+1)!\,(n-2-2k)!} (n-2-2k)\,\bar{p}^{-1} \begin{pmatrix} (1-\bar{p})^{2k+1} (2\bar{p}-1)^{n-2-2k} \\ +(1-\bar{p})^{2k+2} (2\bar{p}-1)^{n-3-2k} \end{pmatrix} \\
& - \sum_{k=0}^{\lfloor \frac{n-2}{2} \rfloor} \frac{(n-1)!}{k!k!\,(n-2-2k)!} (1-\bar{p})^{2k} (2\bar{p}-1)^{n-2-2k} \\
= & \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} \frac{(n-1)!}{k!k!\,(n-1-2k)!} (n-1-2k)\,(1-\bar{p})^{2k} (2\bar{p}-1)^{n-2-2k} \\
& - \sum_{k=1}^{\lfloor \frac{n-1}{2} \rfloor} \frac{(n-1)!}{(k-1)!k!\,(n-2k)!} (n-2k)\,(1-\bar{p})^{2k-1} (2\bar{p}-1)^{n-1-2k} \\
& + \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor - 1} \frac{(n-1)!}{k!\,(k+1)!\,(n-2-2k)!} (n-2-2k)\,(1-\bar{p})^{2k+1} (2\bar{p}-1)^{n-3-2k} \\
& - \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor - 1} \frac{(n-1)!}{k!k!\,(n-1-2k)!} (n-1-2k)\,(1-\bar{p})^{2k} (2\bar{p}-1)^{n-2-2k} \\
= & \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} \frac{(n-1)!}{k!k!\,(n-1-2k)!} (n-1-2k)\,(1-\bar{p})^{2k} (2\bar{p}-1)^{n-2-2k} \\
& - \sum_{k=1}^{\lfloor \frac{n-1}{2} \rfloor} \frac{(n-1)!}{(k-1)!k!\,(n-2k)!} (n-2k)\,(1-\bar{p})^{2k-1} (2\bar{p}-1)^{n-1-2k} \\
& + \sum_{k'=1}^{\lfloor \frac{n}{2} \rfloor} \frac{(n-1)!}{(k'-1)!k'!\,(n-2k')!} \left(n-2k'\right) (1-\bar{p})^{2k'-1} (2\bar{p}-1)^{n-1-2k'} \\
& - \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor - 1} \frac{(n-1)!}{k!k!\,(n-1-2k)!} (n-1-2k)\,(1-\bar{p})^{2k} (2\bar{p}-1)^{n-2-2k}.
\end{aligned}$$

This is zero because the first of these sums equals the fourth one (even when n is odd and the first sum has one extra term, since this term will be zero) and the second one equals the third one (even when n is even and the third sum has one extra term, since this term will be zero).

Thus, $\partial\Upsilon_n(\bar{p},\gamma)/\partial p$ also exists, and is negative. ■

Claim 12 *The function* Υ_n *is continuously differentiable over* $(0.5,1)\times(0,1)$.

Proof. As noted in the proof of Claim 10, the partial derivatives of Υ_n at any point $(p,\gamma)\in(0.5,1)\times(0,1)$ such that $\gamma\in(0,(1-p)/p)$ or $\gamma\in((1-p)/p,1)$ are continuous, since (22) and (23) are polynomials in (p,γ). Furthermore, the proof of Claims 10 and 11 show that both partial derivatives of Υ_n at any $(p,\gamma)\in(0.5,1)\times(0,1)$ such that $\gamma=(1-p)/p$ are also continuous. Therefore, Υ_n is a C^1 function. ■

Claim 13 *Given* $n\geq 2$ *and* $p\in(0,1)$, $\underline{c}_n(p)=\underline{c}_n(1-p)$ *and* $\overline{c}_n(p)=\overline{c}_n(1-p)$.

Proof. It suffices to note that, for $p<0.5$, (3) and Claim 1 give

$$\underline{c}_n(p)=\Pi_R(n,p,1,1)=\Pi_B(n,1-p,1,1)=\underline{c}_n(1-p),$$

whereas (4) and Claim 1 give

$$\overline{c}_n(p)=\Pi_R\left(n,p,1,\frac{p}{1-p}\right)=\Pi_B\left(n,1-p,\frac{p}{1-p},1\right)=\overline{c}_n(1-p).$$

■

Claim 14 *Given* $n\geq 2$ *and* $p\in(0,0.5)\cup(0.5,1)$, $\underline{c}_n(p)<\overline{c}_n(p)$. *If* $p=0.5$, *then* $\underline{c}_n(p)=\overline{c}_n(p)$.

Proof. If $p\in[0.5,1)$, then (20), (21) give

$$\overline{c}_n(p)-\underline{c}_n(p)=\Upsilon_n\left(p,\frac{1-p}{p}\right)-\Upsilon_n(p,1),$$

which is zero if $p=0.5$ and, by Claim 10, is positive if $p>0.5$.

If $p\in(0,0.5)$, then Claim 13 yields $\overline{c}_n(p)-\underline{c}_n(p)=\overline{c}_n(1-p)-\underline{c}_n(1-p)$, which, as shown above (since $1-p>0.5$), is positive. ■

Claim 15 *The function* $\underline{c}_n|_{[0.5,1)}$ *is continuous, strictly decreasing, and* $\lim_{p\to 1-}\underline{c}_n(p)=0$. *The function* $\overline{c}_n|_{[0.5,1)}$ *is continuous, strictly increasing, and* $\lim_{p\to 1-}\overline{c}_n(p)=1$.

Proof. Let $p \in [0.5, 1)$. Expressions (3) and (1) give

$$\begin{aligned}
\underline{c}_n(p) &= \sum_{k=0}^{\left\lfloor \frac{n-1}{2} \right\rfloor} \binom{n-1}{k, k, n-1-2k} p^k (1-p)^k \, 0^{n-1-2k} \qquad (29)\\
&\quad + \sum_{k=0}^{\left\lfloor \frac{n-2}{2} \right\rfloor} \binom{n-1}{k, k+1, n-2-2k} p^k (1-p)^{k+1} \, 0^{n-2-2k} \\
&= \begin{cases} \frac{(n-1)!}{\frac{n}{2}!\left(\frac{n}{2}-1\right)!} p^{\frac{n}{2}-1} (1-p)^{\frac{n}{2}} & \text{if } n \text{ is even} \\ \frac{(n-1)!}{\frac{n-1}{2}!\left(\frac{n-1}{2}\right)!} p^{\frac{n-1}{2}} (1-p)^{\frac{n-1}{2}} & \text{if } n \text{ is odd} \end{cases},
\end{aligned}$$

so that the conclusion $\lim_{p \to 1-} \underline{c}_n(p) = 0$ is valid. Both of these are polynomials in p, whence continuous. They can also be seen to be strictly decreasing in p for $p \in [0.5, 1)$. In fact, using the symbol "$\sim$" to mean "shares its sign with", we have, for all $p \in (0.5, 1)$,

$$\begin{aligned}
\frac{\partial}{\partial p}\left(p^{\frac{n}{2}-1}(1-p)^{\frac{n}{2}}\right) &= \left(\frac{n}{2}-1\right) p^{\frac{n}{2}-2}(1-p)^{\frac{n}{2}} - \frac{n}{2} p^{\frac{n}{2}-1}(1-p)^{\frac{n}{2}-1} \\
&= p^{\frac{n}{2}-2}(1-p)^{\frac{n}{2}-1}\left(\left(\frac{n}{2}-1\right)(1-p) - \frac{n}{2}p\right) \\
&\sim \left(\frac{n}{2}-1\right)(1-p) - \frac{n}{2}p = -(1-p) + \frac{1-2p}{2} n,
\end{aligned}$$

which is negative, since it is decreasing in n, and at $n = 2$ it is already negative ($= -p$). In the same way,

$$\begin{aligned}
\frac{\partial}{\partial p}\left(p^{\frac{n-1}{2}}(1-p)^{\frac{n-1}{2}}\right) &= \frac{n-1}{2} p^{\frac{n-1}{2}-1}(1-p)^{\frac{n-1}{2}} - \frac{n-1}{2} p^{\frac{n-1}{2}}(1-p)^{\frac{n-1}{2}-1} \\
&= \frac{n-1}{2} p^{\frac{n-1}{2}-1}(1-p)^{\frac{n-1}{2}-1}(1-p-p) < 0.
\end{aligned}$$

Therefore, $\underline{c}_n(p)$ is strictly decreasing in p for $p \in [0.5, 1)$.

At the same time, (4) and (1) give, for $p \in [0.5, 1)$,

$$\begin{aligned}
\overline{c}_n(p) &= \Pi_B\left(n, p, \frac{1-p}{p}, 1\right) = \sum_{k=0}^{\left\lfloor \frac{n-1}{2} \right\rfloor} \binom{n-1}{k, k, n-1-2k} (1-p)^{2k} (2p-1)^{n-1-2k} \qquad (30)\\
&\quad + \sum_{k=0}^{\left\lfloor \frac{n-2}{2} \right\rfloor} \binom{n-1}{k, k+1, n-2-2k} (1-p)^{2k+1} (2p-1)^{n-2-2k},
\end{aligned}$$

that is, $P_n(1-p)$ (note that $1 - p \leq 0.5$). Thus, Lemma 8 yields continuity and strict increasingness of $\overline{c}_n|_{[0.5,1)}$, since, for all $p \in (0.5, 1)$, the Chain Rule gives $\overline{c}_n'(p) = -P_n'(1-p) > 0$.

Finally,

$$\lim_{p\to 1-}\bar{c}_n(p)=\sum_{k=0}^{\left\lfloor\frac{n-1}{2}\right\rfloor}\binom{n-1}{k,k,n-1-2k}0^{2k}+\sum_{k=0}^{\left\lfloor\frac{n-2}{2}\right\rfloor}\binom{n-1}{k,k+1,n-2-2k}0^{2k+1}=\binom{n-1}{0,0,n-1}=1.$$

■

Claim 16 *Given $n \geq 2$, $c \in \mathbb{R}_+$ and $p \in (0,1)$, there cannot be more than one type-symmetric equilibrium for the electoral game with these parameter values.*

Proof. Assume $p \geq 0.5$. If (γ, δ) is an electoral equilibrium, according to Claims 3 and 6, it can only be of types E_{11}, $E_{\gamma 1}$, $E_{\gamma\delta}$ or E_{00}. Moreover, Claim 6 assures us that if it is an $E_{\gamma 1}$-type equilibrium, then $\gamma \geq (1-p)/p$. By invoking Claim 5, we will then have (γ, δ) either of the form $(\gamma, (p/(1-p))\gamma)$, for $\gamma \in [0, (1-p)/p)$, or $(\gamma, 1)$, for $\gamma \in [(1-p)/p, 1]$. This is true of any electoral equilibrium, so that if there were another one, (γ', δ'), it would also take one of these forms.

Without loss of generality, let $\gamma < \gamma'$ ($\gamma = \gamma' \in [0, (1-p)/p)$ would imply $\delta = (p/(1-p))\gamma = (p/(1-p))\gamma' = \delta'$, while $\gamma = \gamma' \in [(1-p)/p, 1]$ would imply $\delta = 1 = \delta'$). Claim 10 then yields $\Upsilon_n(p,\gamma) > \Upsilon_n(p,\gamma')$.

We are now left with three possibilities for the equilibrium type of (γ, δ) (E_{11} is excluded because we cannot have $1 < \gamma'$).

Possibility 1: (γ, δ) if of the E_{00} type. Then (γ', δ') is either of the E_{11}, the $E_{\gamma 1}$ or the $E_{\gamma\delta}$ types. In any case, we would obtain $\Upsilon_n(p,\gamma) \leq c \leq \Upsilon_n(p,\gamma')$, a contradiction.

Possibility 2: (γ, δ) if of the $E_{\gamma\delta}$ type. Then (γ', δ') is either of the E_{11}, the $E_{\gamma 1}$ or the $E_{\gamma\delta}$ types. In any case, we would obtain $\Upsilon_n(p,\gamma) = c \leq \Upsilon_n(p,\gamma')$, a contradiction.

Possibility 3: (γ, δ) if of the $E_{\gamma 1}$ type. Then (γ', δ') is either of the E_{11} or the $E_{\gamma 1}$ types. In any case, we would obtain $\Upsilon_n(p,\gamma) = c \leq \Upsilon_n(p,\gamma')$, a contradiction.

If $p < 0.5$ and (γ, δ) and (γ', δ') are two different type-symmetric electoral equilibria, then, by Claim 1, (δ, γ) and (δ', γ') would be two different type-symmetric electoral equilibria for the game with parameter value $1-p$ (≥ 0.5) instead of p, contradicting the argument above. ■

We are now ready to prove Proposition 1.

Proof of Proposition 1. First, let us assume $p \in [0.5, 1)$. By Claims 14 and 15, we know that $0 < \underline{c}_n(p) \leq \bar{c}_n(p) < 1$, with $\underline{c}_n(p) = \bar{c}_n(p)$ only in the $p = 0.5$ case.

If $c \leq \underline{c}_n(p)$, since $\underline{c}_n(p) = \Upsilon_n(p, 1)$ by (20), then the first condition in Claim 8 shows that $(\gamma, \delta) = (1, 1)$ is an electoral equilibrium. Thus Claim 16 guarantees its uniqueness among type-symmetric equilibria. The same comment regarding uniqueness applies to each of the situations considered below, and will be omitted for brevity.

If $c \in (\underline{c}_n(p), \overline{c}_n(p))$ (which is a possibility as long as $p \neq 0.5$), then (20) and (21) give $\Upsilon_n(p, 1) = \underline{c}_n(p) < c < \overline{c}_n(p) = \Upsilon_n(p, (1-p)/p)$. By the continuity of $\Upsilon_n(p, \cdot)$ and the Intermediate Value Theorem, there exists $\gamma \in ((1-p)/p, 1)$ such that $c = \Upsilon_n(p, \gamma)$ $(= \Pi_B(n, p, \gamma, 1))$. As can be seen from the second condition in Claim 8, this means that $(\gamma, 1)$ is an equilibrium.

If $c = \overline{c}_n(p)$, then (21) gives $\Upsilon_n(p, (1-p)/p) = c$. Thus, again from the second condition in Claim 8, $((1-p)/p, 1)$ is an equilibrium.

If $c \in (\overline{c}_n(p), 1)$, then (21) and (19) give $\Upsilon_n(p, (1-p)/p) = \overline{c}_n(p) < c < 1 = \Pi_B(n, p, 0, 0) = \Upsilon_n(p, 0)$. Again by the Intermediate Value Theorem, there exists $\gamma \in (0, (1-p)/p) = (0, \min(1, (1-p)/p))$ such that $c = \Upsilon_n(p, \gamma)$ $(= \Pi_B(n, p, \gamma, p/(1-p)\gamma))$. As can be seen from the third condition in Claim 8, this means that $(\gamma, p/(1-p)\gamma)$ is an equilibrium.

Finally, if $c \geq 1$, then by Claim 4 we know that $(\gamma, \delta) = (0, 0)$ is an equilibrium.

Now for the $p < 0.5$ case, in which, by Claim 13, we have $\underline{c}_n(p) = \underline{c}_n(1-p)$ and $\overline{c}_n(p) = \overline{c}_n(1-p)$.

If $c \leq \underline{c}_n(p)$ $(= \underline{c}_n(1-p))$, then, as proved above, $(1, 1)$ will be an equilibrium of the electoral game with parameter value $1-p$. Thus, by Claim 2, $(1, 1)$ will be an equilibrium of the game with parameter value p.

If $c \in (\underline{c}_n(p), \overline{c}_n(p))$ $(= (\underline{c}_n(1-p), \overline{c}_n(1-p))$ by Claim 13), then, as proved above, there will exist $\gamma' \in (p/(1-p), 1)$ such that $\Pi_B(n, 1-p, \gamma', 1) = c$ and $(\gamma', 1)$ is an equilibrium of the electoral game with parameter value $1-p$. Let $\delta := \gamma'$. Then, by Claim 2, $(1, \delta)$ will be an equilibrium of the game with parameter value p, and it follows from Claim 1 that $\Pi_R(n, p, 1, \delta) = \Pi_B(n, 1-p, \gamma', 1) = c$.

If $c = \overline{c}_n(p)$ $(= \overline{c}_n(1-p))$, then, as proved above, $\Pi_B(n, 1-p, p/(1-p), 1) = c$ and $(p/(1-p), 1)$ will be an equilibrium of the electoral game with parameter value $1-p$. Thus, by Claim 2, $(1, p/(1-p))$ will be an equilibrium of the game with parameter value p, and it follows from Claim 1 that $\Pi_R(n, p, 1, p/(1-p)) = \Pi_B(n, 1-p, p/(1-p), 1) = c$.

If $c \in (\overline{c}_n(p), 1)$ $(= (\overline{c}_n(1-p), 1))$, then, as proved above, there will exist $\gamma' \in (0, p/(1-p))$ such that $\Pi_B(n, 1-p, \gamma', ((1-p)/p)\gamma') = \Pi_R(n, 1-p, \gamma', ((1-p)/p)\gamma') = c$ and $(\gamma', ((1-p)/p)\gamma')$ is an equilibrium of the electoral game with parameter value $1-p$. Let $\gamma := ((1-p)/p)\gamma' \in (0, 1) = (0, \min(1, (1-p)/p))$. Then, by Claim 2, $(\gamma, (p/(1-p))\gamma)$ will be an equilibrium of the game with parameter value p, and it follows from Claim 1 that $\Pi_B(n, p, \gamma, (p/(1-p))\gamma) = \Pi_R(n, 1-p, \gamma', ((1-p)/p)\gamma') = c$.

Finally, the argument already given for the $c \geq 1$ case remains valid. ■

Given the proof of Proposition 1 and the resulting definition of functions γ_B and γ_R, Claim 2 could be restated in a much simpler form:

Claim 17 *Given* $n \geq 2$*,* $c \in \mathbb{R}_+$ *and* $p \in (0,1)$*,* $\gamma_B(n,c,p) = \gamma_R(n,c,1-p)$*.*

Next we pave the way for the proof of Proposition 2.

Claim 18 *Given* $n \geq 2$*, the type-symmetric Nash equilibrium* $(\gamma_B(n,c,p), \gamma_R(n,c,p))$ *varies continuously with* c *and* p*.*

Proof. Consider the problem used to find the type-symmetric Nash equilibria, stated at the beginning of this appendix:

$$\max_{(\tilde{\gamma},\tilde{\delta})\in[0,1]\times[0,1]} \Psi\left(n,c,p,\tilde{\gamma},\tilde{\delta},\gamma,\delta\right).$$

Let n be fixed. The solution to this problem, the existence of which is assured by the Extreme Value Theorem, is the correspondence Γ that maps each tuple (c,p,γ,δ) to the set $\left\{\left(\tilde{\gamma},\tilde{\delta}\right) : \left(\tilde{\gamma},\tilde{\delta}\right) \in \arg\max_{(\tilde{\gamma},\tilde{\delta})\in[0,1]\times[0,1]} \Psi\left(n,c,p,\tilde{\gamma},\tilde{\delta},\gamma,\delta\right)\right\}$. For notational purposes, denote $\theta := (c,p)$ and $x := (\gamma,\delta)$, so that the solution of the problem can be denoted $\Gamma(\theta,x)$. This problem fulfills all the conditions of Berge's Maximum Theorem (fixed n, $\Psi(n,\cdot,\cdot,\cdot,\cdot,\cdot,\cdot)$ is continuous, and the constraint correspondence is continuous for being constant), which implies that Γ is upper hemicontinuous in (θ,x).

For a given θ, note that a fixed point of $\Gamma(\theta,\cdot)$ (i.e., x such that $x \in \Gamma(\theta,x)$) is precisely what we have called a type-symmetric Nash equilibrium of the electoral game, and by Proposition 1, such a fixed point exists and is unique. The function that maps each θ to the fixed point shall be denoted x^* (that is, $x^*(\cdot) = (\gamma_B(n,\cdot), \gamma_R(n,\cdot))$).

Consider now the problem of choosing x so as to minimize the Euclidean distance d between x and $\Gamma(\theta,x)$, $d(x,\Gamma(\theta,x)) := \inf\{\|x-y\| : y \in \Gamma(\theta,x)\}$:

$$\min_{x\in[0,1]\times[0,1]} d(x,\Gamma(\theta,x)).$$

Since $\Gamma(\theta,\cdot)$ is compact-valued (because Γ is upper hemicontinuous), we have $d(x,\Gamma(\theta,x)) = 0$ if and only if $x \in \Gamma(\theta,x)$. But note that this zero distance is actually attainable, if we take $x = x^*(\theta)$ (and only if, because of the uniqueness of this fixed point).

Thus, $x^*(\theta)$ is the unique solution to the above minimization problem. A second application of Berge's Maximum Theorem now implies the continuity of x^*, that is, that γ and δ vary continuously with p and c. ■

Proof of Lemma 1. Let us first assume $p \in [0.5, 1)$. Since $c \in (\overline{c}_n(p), 1)$, Proposition 1 guarantees that $(\gamma_B(n,c,p), \gamma_R(n,c,p))$ will be an $E_{\gamma\delta}$-type equilibrium, that is, $\gamma_B(n,c,p) \in (0, (1-p)/p)$, $\gamma_R(n,c,p) = (p/(1-p))\gamma_B(n,c,p)$ and $\Upsilon_n(p, \gamma_B(n,c,p)) = c$.[24]

Claim 15 gives $\overline{c}_n(p) \geq \overline{c}_n(0.5)$, so that also $c \in (\overline{c}_n(0.5), 1)$, and Proposition 1 guarantees that $(\gamma^*(n,c), (0.5/(1-0.5))\gamma^*(n,c)) = (\gamma^*(n,c), \gamma^*(n,c))$ will be an $E_{\gamma\delta}$-type equilibrium of the electoral game with the probability parameter 0.5. That is, $\gamma^*(n,c) \in (0,1)$ and $\Upsilon_n(0.5, \gamma^*(n,c)) = c$.

Now, by (22), when $\gamma < (1-p)/p$ as in both cases above, $\Upsilon_n(p,\gamma) = P_n(p\gamma)$ (note that $p\gamma \in (0, 1-p) \subseteq [0, 0.5]$). By Lemma 8, P_n is strictly decreasing, whence injective. Therefore, $\Upsilon_n(p, \gamma_B(n,c,p)) = c = \Upsilon_n(0.5, \gamma^*(n,c))$ implies $P_n(p\gamma_B(n,c,p)) = P_n(0.5\gamma^*(n,c))$, which, in its turn, yields $p\gamma_B(n,c,p) = 0.5\gamma^*(n,c)$, and $\gamma_R(n,c,p) = (p/(1-p))\gamma_B(n,c,p) = 0.5\gamma^*(n,c)/(1-p)$.

If $p \in (0, 0.5)$ (so that $1-p \in (0.5, 1)$) and $c \in (\overline{c}_n(p), 1)$, then, since $\overline{c}_n(p) = \overline{c}_n(1-p)$ (Claim 13), we have $c \in (\overline{c}_n(1-p), 1)$ and, from the argument above and Claim 17, we get $\gamma^*(n,c) \in (0,1)$, $\gamma^*(n,c)/(2(1-p)) = \gamma_B(n,c,1-p) = \gamma_R(n,c,p)$ and $\gamma^*(n,c)/(2p) = \gamma_R(n,c,1-p) = \gamma_B(n,c,p)$. ■

Proof of Proposition 2. Continuity was proved in Claim 18. Therefore, in order to check the claimed monotonicity properties of γ_B and γ_R with respect to p and to c, it suffices to check these properties over the following four open regions of the plane: $\{(p,c) \in (0,1) \times \mathbb{R}_+ : c < \underline{c}_n(p)\}$, $A_1 := \{(p,c) \in (0,1) \times \mathbb{R}_+ : \underline{c}_n(p) < c < \overline{c}_n(p)\}$, $A_2 := \{(p,c) \in (0,1) \times \mathbb{R}_+ : \overline{c}_n(p) < c < 1\}$ and $\{(p,c) \in (0,1) \times \mathbb{R}_+ : c > 1\}$. Over the first and fourth of these regions, besides $A_1 \cap (0, 0.5) \times \mathbb{R}_+$, Proposition 1 shows that γ_B will be a constant (1, 0 and 1, respectively), so that it is decreasing (although not strictly) in c and in p. The same applies to γ_R over the first and fourth regions, besides $A_1 \cap [0.5, 1) \times \mathbb{R}_+$, which will thus be decreasing in c and increasing (not strictly) in p.

We may then argue for the claimed monotonicity properties by simply computing the requested partial derivatives in the remaining regions. It will also be convenient to define $A_3 := \{(p,c) \in (0,1) \times \mathbb{R}_+ : \underline{c}_n(p) < c < 1\}$, which includes both A_1 and A_2. Note that these three sets are open, because of the continuity of $\underline{c}_n$ and $\overline{c}_n$ (by Claims 15 and 13). This makes the computation of partial derivatives of γ_B and γ_R within these regions fairly straightforward.

It will be helpful in the computation of these derivatives to note two properties beforehand. The first one is that, if $p \in (0.5, 1)$ and $(p,c) \in A_3$, then Proposition 1 implies $\gamma_B(n,c,p) \in (0,1)$ and $\Upsilon_n(p, \gamma_B(n,c,p)) = c$. The second one is that, if $(p,c) \in A_2$, then Proposition 1 implies $\gamma_B(n,c,p) \in (0, \min(1, (1-p)/p))$ and $c = \Pi_B(n, p, \gamma_B(n,c,p), (p/(1-p))\gamma_B(n,c,p)) = P_n(p\gamma_B(n,c,p))$ (this can be computed since $p\gamma_B(n,c,p) \in (0, p\min(1, (1-p)/p)) = (0, \min(p, (1-p))) \subseteq (0, 0.5)$).

[24] After this initial observation, the proof could also be followed as in Proposition 3 of Goeree and Großer (2007), where an $E_{\gamma\delta}$-type equilibrium is assumed from the start.

i)

a) Assume $c \in (\bar{c}_n(p), 1)$, so that $(p, c) \in A_2$. In this case, as noted above, $P_n(p\gamma_B(n, c, p)) = c$. Lemma 8 guarantees that P_n is continuously differentiable at $p\gamma_B(n, c, p)$ and $P_n'(p\gamma_B(n, c, p)) < 0$, so that the Implicit Function Theorem gives

$$\frac{\partial}{\partial c}\gamma_B(n, c, p) = \frac{1}{pP_n'(p\gamma_B(n, c, p))} < 0.$$

Also, Proposition 1 yields $\gamma_R(n, c, p) = (p/(1-p))\gamma_B(n, c, p)$, so that $\partial\gamma_R(n, c, p)/\partial c = (p/(1-p))\partial\gamma_B(n, c, p)/\partial c < 0$.

Now assume $c \in (\underline{c}_n(p), 1)$, so that $(p, c) \in A_3$. If $p = 0.5$, then $\underline{c}_n(p) = \bar{c}_n(p)$ by Claim 14, so that $(p, c) \in A_2$ and the above proof is applicable. If $p \in (0.5, 1)$, then $\gamma_B(n, c, p) \in (0, 1)$ and $\Upsilon_n(p, \gamma_B(n, c, p)) = c$, as noted above. By Claim 12, Υ_n is continuously differentiable at $(p, \gamma_B(n, c, p)) \in (0.5, 1) \times (0, 1)$ and, by Claim 10, $\partial\Upsilon_n(p, \gamma_B(n, c, p))/\partial\gamma < 0$. Thus, we can apply the Implicit Function Theorem to get

$$\frac{\partial}{\partial c}\gamma_B(n, c, p) = \frac{1}{\frac{\partial}{\partial\gamma}\Upsilon_n(p, \gamma_B(n, c, p))} < 0.$$

b) Let us now assume $p \in (0, 0.5)$. For $c \in (\bar{c}_n(p), 1)$, it was already shown in part (a) that $\partial\gamma_B(n, c, p)/\partial c < 0$ (that part of the argument holds for all $p \in (0, 1)$). For $c \in (\underline{c}_n(p), 1)$ $(= (\underline{c}_n(1-p), 1)$, by Claim 13), since $1 - p \in (0.5, 1)$, we already know from part (a) that $\partial\gamma_B(n, c, 1-p)/\partial c < 0$. Since $\gamma_R(n, \cdot, p) = \gamma_B(n, \cdot, 1-p)$ (Claim 17), we then have $\partial\gamma_R(n, c, p)/\partial c = \partial\gamma_B(n, c, 1-p)/\partial c < 0$.

ii)

a) The structure of the proof of part (a) can be applied here as well. If $c \in (\bar{c}_n(p), 1)$, then $(p, c) \in A_2$ and, as explained above, $c = P_n(p\gamma_B(n, c, p)) = P_n((1-p)\gamma_R(n, c, p))$, where the last equality follows from Lemma 1. Therefore, the Implicit Function Theorem yields

$$\frac{\partial}{\partial p}\gamma_R(n, c, p) = \frac{\gamma_R(n, c, p)}{1-p} > 0.$$

If $c \in (\underline{c}_n(p), 1)$, then $(p, c) \in A_3$, and there are two cases to consider. If $p = 0.5$, then $\underline{c}_n(p) = \bar{c}_n(p)$ by Claim 14, so that $(p, c) \in A_2$ and $\gamma_B(n, c, p)$ is determined through $P_n(p\gamma_B(n, c, p)) = c$. In this case, an application of the Implicit Function Theorem gives

$$\frac{\partial}{\partial p}\gamma_B(n, c, p) = -\frac{\gamma_B(n, c, p)}{p} < 0.$$

If $p \in (0.5, 1)$, then $\Upsilon_n(p, \gamma_B(n, c, p)) = c$ and, as argued in part (a), $\partial\gamma_B(n, c, p)/\partial p$ can be obtained via Implicit Function Theorem:

$$\frac{\partial}{\partial p}\gamma_B(n,c,p) = -\frac{\frac{\partial}{\partial p}\Upsilon_n(p,\gamma_B(n,c,p))}{\frac{\partial}{\partial \gamma}\Upsilon_n(p,\gamma_B(n,c,p))}$$

Since both numerator and denominator are negative, by Claims 11 and 10, we have $\partial\gamma_B(n,c,p)/\partial p < 0$.

b) Let $p \in (0, 0.5)$. For $c \in (\overline{c}_n(p), 1)$, as argued in part (a), we have $P_n(p\gamma_B(n,c,p)) = c$, so that the Implicit Function Theorem gives

$$\frac{\partial}{\partial p}\gamma_B(n,c,p) = -\frac{\gamma_B(n,c,p)}{p} < 0.$$

Finally, for $c \in (\underline{c}_n(p), 1) = (\underline{c}_n(1-p), 1)$, as shown in part (a), $\partial\gamma_B(n,c,1-p)/\partial p < 0$. Therefore, since $\gamma_R(n,\cdot,p) = \gamma_B(n,\cdot,1-p)$, the Chain Rule yields $\partial\gamma_R(n,c,p)/\partial p = -\partial\gamma_B(n,c,1-p)/\partial p > 0$. ■

Proof of Lemma 2. If $p \in [0.5, 1)$, then, as observed in the proof of Claim 15, $\overline{c}_n(p) = P_n(1-p)$, which is strictly decreasing in n, as shown in the beginning of the proof of Lemma 3 in Taylor and Yildirim (2010a, p. 367).

If $p \in (0, 0.5)$, then, by Claim 13, $\overline{c}_n(p) = \overline{c}_n(1-p)$, which, as argued above (note that $1-p \in (0.5, 1)$), is strictly decreasing in n.

In what concerns the limit of $\overline{c}_n(p)$, it follows from Proposition 2 and Lemma 3, parts (i) and (ii) in Taylor and Yildirim (2010a) that, given $p \in (0,1)$ and $c \in (0,1)$, there exists $n_0(c,p) \in \mathbb{N}$ such that, for any $n \geq n_0(c,p)$, the equilibrium of the electoral game with parameter values n, c and p will be of the interior type. Therefore, given $\varepsilon > 0$, without loss of generality $\varepsilon < 1$, our Proposition 1 ensures that, for all $n \geq n_0(\varepsilon, p)$, $\varepsilon > \overline{c}_n(p) = |\overline{c}_n(p) - 0|$, i.e., $\lim_{n\to\infty}\overline{c}_n(p) = 0$.

The fact that $\lim_{n\to\infty}\gamma_B(n,c,p) = \lim_{n\to\infty}\gamma_R(n,c,p) = 0$ is also proven in Lemma 3, part (i) of Taylor and Yildirim (2010a), whereas the existence and positivity of $\lim_{n\to\infty} n\gamma_B(n,c,p)$ and $\lim_{n\to\infty} n\gamma_R(n,c,p)$ are proven in Lemma 4 of that work. ■

Proof of Lemma 3. Let $c \in (0,1)$ be given. For each $p \in [\overline{q}, 1-\overline{q}]$, Lemma 2 and the Well Ordering Principle guarantee the existence of the smallest natural number $k \geq 2$ such that $\overline{c}_k(p) < c$. Call this number $\overline{n}(c,p)$.

Claim 13 readily yields $\overline{n}(c, 1-p) = \overline{n}(c,p)$. We now argue that $\overline{n}(c,p)$ is decreasing in p for $p \in [\overline{q}, 0.5]$ (and, therefore, increasing in p for $p \in [0.5, 1-\overline{q}]$). In fact, let $p_1, p_2 \in [\overline{q}, 0.5]$ be such that $p_1 < p_2$. By Claim 15, we have, for any natural $k \geq 2$, $\overline{c}_k(p_1) > \overline{c}_k(p_2)$. In particular,

$\overline{c}_{\bar{n}(c,p_1)}(p_2) < \overline{c}_{\bar{n}(c,p_1)}(p_1) < c$, so that $\bar{n}(c,p_2) \leq \bar{n}(c,p_1)$.

Thus, we have shown that, for a given $c \in (0,1)$, $\bar{n}(c,p)$ is bounded for $p \in [\bar{q}, 1-\bar{q}]$, by $\bar{n}(c, 1-\bar{q}) = \bar{n}(c,\bar{q})$. Define $n_0(c) = \bar{n}(c,\bar{q})$.

Now, given $n \geq n_0(c)$ and $p \in [\bar{q}, 1-\bar{q}]$, since $n_0(c) \geq \bar{n}(c,p)$, we have $n \geq \bar{n}(c,p)$, so that Lemma 2 yields $\overline{c}_n(p) \leq \overline{c}_{\bar{n}(c,p)}(p) < c < 1$. Therefore, by Proposition 1, $(\gamma_B(n,c,p), \gamma_R(n,c,p)) \in (0,1)^2$. ■

Appendix B

This appendix presents proofs for all of the results stated in sections 4 and 5.

Before getting started, we show a computation that will be used in several proofs. Given $n \geq 2$, $c \in \mathbb{R}_+$ and $p \in [\bar{q}, 1-\bar{q}]$, by plugging (10), (11), (13) and (14) into (9), we get

$$U(n,c,p,q) = -\left[\begin{array}{c} n^2 (p-q)^2 \left(\gamma_B^2 (n,c,p) + \gamma_R^2 (n,c,p)\right) + n\left(q\gamma_B(n,c,p)\right)\left(1 - q\gamma_B(n,c,p)\right) \\ +n\left((1-q)\gamma_R(n,c,p)\right)\left(1-(1-q)\gamma_R(n,c,p)\right) \end{array}\right]. \tag{31}$$

If $c \in (0,1)$ and $n \geq n_0(c)$, Lemmas 3 and 1 yield $\gamma_B(n,c,p) = \gamma^*/(2p)$ and $\gamma_R(n,c,p) = \gamma^*/(2(1-p))$, where $\gamma^* = \gamma^*(n,c) \in (0,1)$. In this case, (31) can be rewritten as

$$\begin{aligned} U(n,c,p,q) &= -n^2(p-q)^2\left(\frac{\gamma^{*2}}{4p^2} + \frac{\gamma^{*2}}{4(1-p)^2}\right) - n\frac{q}{p}\frac{\gamma^*}{2}\left(1 - \frac{q}{p}\frac{\gamma^*}{2}\right) - n\frac{1-q}{1-p}\frac{\gamma^*}{2}\left(1-\frac{1-q}{1-p}\frac{\gamma^*}{2}\right) \\ &= \left[-\frac{n\gamma^*}{2}\frac{q}{p} + \frac{n\gamma^{*2}}{4}\frac{q^2}{p^2} - \frac{n^2\gamma^{*2}}{4}\frac{(p-q)^2}{p^2}\right] \\ &\quad + \left[-\frac{n\gamma^*}{2}\frac{1-q}{1-p} + \frac{n\gamma^{*2}}{4}\frac{(1-q)^2}{(1-p)^2} - \frac{n^2\gamma^{*2}}{4}\frac{(p-q)^2}{(1-p)^2}\right]. \end{aligned} \tag{32}$$

Several proofs below will involve differentiation of U with respect to p, the pollster's choice variable (viewed as a function of p, the above expression obviously belongs to $C^\infty([\bar{q}, 1-\bar{q}])$). As seen in (32), in doing so, we need not worry about p entering $U(n,c,p,q)$ implicitly, perhaps inside a $\gamma_B(n,c,p)$ or $\gamma_R(n,c,p)$ expression as in (31). This shows how Lemma 1 (and Lemma 3, which allows us to use Lemma 1 for a sufficiently large n) is key in this study.

Proof of Proposition 3. For any $p \in [\bar{q}, 1-\bar{q}]$, since $c = 0$, Proposition 1 yields the electoral equilibrium is $(\gamma_B(n,c,p), \gamma_R(n,c,p)) = (1,1)$. Therefore, by (31),

$$U(n,c,p,q) = -\left[n^2(p-q)^2 2 + nq(1-q) + n(1-q)(1-(1-q))\right] = -2n^2(p-q)^2 - 2nq(1-q).$$

It is clear that the maximum value of $U(n,c,p,q)$ occurs only at $p = q$. ■

Claim 19 *Given $q \in [\bar{q}, 1-\bar{q}]$, $c \in (0,1)$ and $n \geq n_0(c)$, if $p_n^*(c,q)$ is a solution to $\max_{p\in[\bar{q},1-\bar{q}]} U(n,c,p,q)$, then $1 - p_n^*(c,q)$ is a solution to $\max_{p\in[\bar{q},1-\bar{q}]} U(n,c,p,1-q)$.*

Proof. Since $c \in (0,1)$ and $n \geq n_0(c)$, (32) holds. Note from that expression that $U(n,c,p,1-q) = U(n,c,1-p,q)$.

Now, if $r \in [\bar{q}, 1-\bar{q}]$ were such that $U(n,c,r,1-q) > U(n,c,1-p_n^*(c,q), 1-q)$, then we would have $U(n,c,1-r,q) > U(n,c,p_n^*(c,q), q)$. Since $1-r \in [\bar{q}, 1-\bar{q}]$, this would contradict the fact that $p_n^*(c,q)$ is a solution of $\max_{p\in[\bar{q},1-\bar{q}]} U(n,c,p,q)$. ■

Before proving Proposition 4, we state a couple of lemmas about the pollster's utility function. Since we assume that the reader is more interested in following the development of proof strategies, rather than big calculations, in the proofs below we unashamedly make use of mathematical software to help with the computation of derivatives and algebraic manipulations; this is indicated by the "$\doteq$" symbol. We will continue to employ the symbol "$\sim$" to mean "shares its sign with."

Lemma 9 *Given $q \in [\bar{q}, 1-\bar{q}]$, $c \in (0,1)$ and $n \geq n_0(c)$, if $\gamma^*(n,c) \leq 1/n$, then $\partial^2 U(n,c,p,q)/\partial p^2 < 0$ for all $p \in (\bar{q}, 1-\bar{q})$.*

Proof. Since $c \in (0,1)$ and $n \geq n_0(c)$, (32) holds, regardless of the value of $p \in (\bar{q}, 1-\bar{q})$. We now apply the $\partial^2/\partial p^2$ operator to each term in square brackets in (32). First, note that

$$\begin{aligned}
&\frac{\partial^2}{\partial p^2}\left(-\frac{n\gamma^*}{2}\frac{q}{p} + \frac{n\gamma^{*2}}{4}\frac{q^2}{p^2} - \frac{n^2\gamma^{*2}}{4}\frac{(p-q)^2}{p^2}\right) \\
\doteq\ & -\frac{1}{2}\frac{n\gamma^*}{p^4} q\left(2p - 3q\gamma^* - 2np\gamma^* + 3nq\gamma^*\right) \sim (3q + 2np - 3nq)\gamma^* - 2p.
\end{aligned}$$

This is an affine function of γ^* which, when evaluated at $\gamma^* = 0$, equals $-2p < 0$, and when evaluated at $\gamma^* = 1/n$, equals $3q/n + 2p - 3q - 2p = 3q(1-n)/n < 0$. Therefore, it is negative for all $\gamma^* \in (0, 1/n]$.

Similarly, we have

$$\begin{aligned}
&\frac{\partial^2}{\partial p^2}\left(-\frac{n\gamma^*}{2}\frac{1-q}{1-p} + \frac{n\gamma^{*2}}{4}\frac{(1-q)^2}{(1-p)^2} - \frac{n^2\gamma^{*2}}{4}\frac{(p-q)^2}{(1-p)^2}\right) \\
\doteq\ & -\frac{1}{2}\frac{n\gamma^*}{(1-p)^4}(1-q)\left(n\gamma^* - 3\gamma^* - 2p + 3q\gamma^* + 2np\gamma^* - 3nq\gamma^* + 2\right) \\
\sim\ & (3nq - 3q - 2np - n + 3)\gamma^* + 2p - 2,
\end{aligned}$$

and the same reasoning applies: it is an affine function of γ^* which, when evaluated at $\gamma^* = 0$, equals $2p - 2 < 0$, and when evaluated at $\gamma^* = 1/n$, equals $3q - 3q/n - 2p - 1 + 3/n + 2p - 2 = 3q - 3q/n + 3/n - 3 = -3(1-q)(n-1)/n < 0$. Therefore, it is negative for all $\gamma^* \in (0, 1/n]$.

Thus, $\partial^2 U(n,c,p,q)/\partial p^2 < 0$. ■

In order to be able to analyze the $\gamma^*(n,c) > 1/n$ case as well (for which, unfortunately, U will not necessarily be concave in p), it will be useful to establish the notation

$$f_n(p,q,\gamma) := \left[\begin{array}{c} (2nq\gamma - n\gamma - 2q + 1)\,p^4 + \left(4q + \gamma - 2q\gamma + 2q^2\gamma - 2nq\gamma - 2nq^2\gamma - 1\right)p^3 \\ +\left(3nq\gamma - 3q^2\gamma - 3q + 3nq^2\gamma\right)p^2 + \left(q + 3q^2\gamma - nq\gamma - 3nq^2\gamma\right)p + nq^2\gamma - q^2\gamma \end{array}\right], \tag{33}$$

since differentiation of (32) yields

$$\frac{\partial}{\partial p} U(n,c,p,q) \mathrel{\mathring{=}} \frac{1}{2}\frac{n\gamma^*}{p^3(1-p)^3} f_n(p,q,\gamma^*), \tag{34}$$

where, as usual, γ^* stands for $\gamma^*(n,c)$.

Lemma 10 *Given $q \in [\bar{q}, 1-\bar{q}]$, $c \in (0,1)$ and $n \geq n_0(c)$, if $\gamma^*(n,c) > 1/n$, then $f_n(p,q,\gamma^*(n,c))$ is strictly decreasing in p for $p \in [\bar{q}, 1-\bar{q}]$.*

Proof. Since $c \in (0,1)$ and $n \geq n_0(c)$, (32) holds, regardless of the value of $p \in [\bar{q}, 1-\bar{q}]$. From the hypothesis on γ^* $(= \gamma^*(n,c))$ and Lemma 1, we have $\gamma^* \in (1/n, 1)$.

Since $f_n(p,q,\gamma^*)$ is an affine function of γ^*, so will be its derivative with respect to p, for any $p \in (\bar{q}, 1-\bar{q})$:

$$\begin{aligned} \frac{\partial}{\partial p} f_n(p,q,\gamma^*) \mathrel{\mathring{=}} & \left(\begin{array}{c} 6p^2q^2 - 3nq^2 - 6pq^2 - 6p^2q - 4np^3 - nq + 3p^2 \\ +3q^2 + 6npq^2 - 6np^2q + 8np^3q - 6np^2q^2 + 6npq \end{array}\right)\gamma^* \\ & +q + 12p^2q - 8p^3q - 6pq - 3p^2 + 4p^3. \end{aligned}$$

At $\gamma^* = 1/n$, this expression amounts to

$$\begin{aligned} \frac{\partial}{\partial p} f_n(p,q,1/n) &\mathrel{\mathring{=}} -3\left(-2pq^2 - 2p^2q + 2p^2q^2 + p^2 + q^2\right)\frac{n-1}{n} \\ &= -3\left(p^2(1-q)^2 + q^2(1-p)^2\right)\frac{n-1}{n} < 0. \end{aligned}$$

At $\gamma^* = 1$, it is equal to

$$\begin{aligned} \frac{\partial}{\partial p} f_n(p,q,1) &\mathrel{\mathring{=}} \left(\left(6p - 6p^2 - 3\right)q^2 + \left(8p^3 - 6p^2 + 6p - 1\right)q - 4p^3\right)(n-1) \\ &\sim (8q-4)\,p^3 + \left(-6q^2 - 6q\right)p^2 + \left(6q^2 + 6q\right)p + \left(-3q^2 - q\right) =: g(p,q). \end{aligned}$$

If $q = 0.5$, then this becomes the quadratic $\left(-18p^2 + 18p - 5\right)/4$, the maximum value of which, obtained at the vertex $p = 0.5$, is $-1/8$. If $q \neq 0.5$, then $g(p, q)$ is a cubic in p, with discriminant

$$\begin{aligned} & \left(-6q^2 - 6q\right)^2 \left(6q^2 + 6q\right)^2 - 4(8q-4)\left(6q^2 + 6q\right)^3 - 4\left(-6q^2 - 6q\right)^3 \left(-3q^2 - q\right) \\ & -27(8q-4)^2 \left(-3q^2 - q\right)^2 + 18(8q-4)\left(-6q^2 - 6q\right)\left(6q^2 + 6q\right)\left(-3q^2 - q\right) \\ \mathrel{\mathring{=}} \; & -432q^2(1-q)^2\left(3q^4 - 6q^3 + q^2 + 2q + 1\right) \sim 16\left(-3q^4 + 6q^3 - q^2 - 2q - 1\right). \end{aligned}$$

By letting $s := (2q-1)^2$, it may be noted that $-3s^2 + 14s - 27$ is identical to the above expression. Since the discriminant of this concave quadratic in s is negative, it is negative itself, and so is the discriminant above. Thus, $g(\cdot, q)$ has exactly one real root p'. This root cannot lie in the $(0, 1)$ interval. In fact, in case $q > 0.5$, $\lim_{p \to +\infty} g(p, q) = +\infty$ and $g(1, q) = -3q^2 + 7q - 4 = (1-q)(3q-4) < 0$ yield $p' > 1$, whereas in case $q < 0.5$, $\lim_{p \to -\infty} g(p, q) = +\infty$ and $g(0, q) = -3q^2 - q < 0$ yield $p' < 0$. Therefore, in either case, $g(p, q) < 0$ for $p \in (\bar{q}, 1 - \bar{q})$.

Since $\partial f_n(p, q, \gamma^*)/\partial p$ is affine in γ^* and both $\partial f_n(p, q, 1/n)/\partial p$ and $\partial f_n(p, q, 1)/\partial p$ are negative, we shall have, for all $\gamma^* \in (1/n, 1)$, $\partial f_n(p, q, \gamma^*)/\partial p < 0$, whence $f_n(p, q, \gamma^*)$ is strictly decreasing in p for $p \in [\bar{q}, 1 - \bar{q}]$. ■

The same way the thesis of Lemma 9 may not hold if $\gamma^*(n, c) > 1/n$, the thesis of Lemma 10 may also not hold if $\gamma^*(n, c) \leq 1/n$ (these claims can be checked numerically). This is why we will need both lemmas in the following proof.

Proof of Proposition 4. Since $c \in (0, 1)$ and $n \geq n_0(c)$, (32) holds, regardless of the value of $p \in [\bar{q}, 1 - \bar{q}]$, and $\partial U(n, c, p, q)/\partial p$ shares its sign with $f_n(p, q, \gamma^*)$, where $\gamma^* = \gamma^*(n, c)$.

i) Note that

$$f_n(p, 0.5, \gamma^*) \mathrel{\mathring{=}} \frac{1}{4}(2p-1)\left(\gamma^* - 2p - n\gamma^* - p\gamma^* + p^2\gamma^* + 2p^2 + 3np\gamma^* - 3np^2\gamma^*\right),$$

so that $f_n(0.5, 0.5, \gamma^*) = 0$, and 0.5 is a critical point of $U(n, c, \cdot, 0.5)$. If $\gamma^* \leq 1/n$, then Lemma 9 ensures 0.5 will be the unique solution to (8). If $\gamma^* > 1/n$, then Lemma 10 implies that, for every $p \in [\bar{q}, 0.5)$, $f_n(p, 0.5, \gamma^*) > f_n(0.5, 0.5, \gamma^*) = 0$, and, for every $p \in (0.5, 1 - \bar{q}]$, $f_n(p, 0.5, \gamma^*) < f_n(0.5, 0.5, \gamma^*) = 0$. Since $\partial U(n, c, p, q)/\partial p$ shares its sign with $f_n(p, q, \gamma^*)$, also in this case $U(n, c, \cdot, 0.5)|_{[\bar{q}, 0.5]}$ will be strictly increasing and $U(n, c, \cdot, 0.5)|_{[0.5, 1-\bar{q}]}$, strictly decreasing, so that 0.5 will again be the unique solution to (8).

ii) The proof follows the exact same lines as the previous case, once we are able to pin down a zero of $f_n(\cdot, q, \gamma^*)$ between 0.5 and q. It should be noted that

$$f_n(0.5, q, \gamma^*) \mathrel{\mathring{=}} \frac{1}{16}\left((n-2)\gamma^* + 1\right)(2q-1) > 0$$

and

$$f_n(q,q,\gamma^*) \mathrel{\mathring{=}} -(1-\gamma^*)\,q^2(2q-1)(1-q)^2 < 0.$$

Because $f_n(p,q,\gamma^*)$ is a polynomial in p (hence, continuous), by the Intermediate Value Theorem, there exists $p^* \in (0.5,q)$ such that $f_n(p^*,q,\gamma^*) = 0$. That is, p^* is a critical point of $U(n,c,\cdot,q)$ in the interval $[\bar{q},1-\bar{q}]$. Now both cases $\gamma^* \leq 1/n$ and $\gamma^* > 1/n$ must be considered, in the exact same fashion as in case (i).

iii) This follows directly from the previous case and Claim 19. ■

The following claim follows directly from the proof of Proposition 4.

Claim 20 *Given $q \in [\bar{q},1-\bar{q}]$, $c \in (0,1)$ and $n \geq n_0(c)$, $p_n^*(c,q)$ is the only $p \in [\bar{q},1-\bar{q}]$ such that $f_n(p,q,\gamma^*(n,c)) = 0$.*

Lemma 11 *Given $c \in (0,1)$ and $n \geq n_0(c)$, the function $p_n^*(c,\cdot)$ is one to one.*

Proof. In fact, let $q_1,q_2 \in [\bar{q},1-\bar{q}]$ be such that $p_n^*(c,q_1) = p_n^*(c,q_2)$ and $q_1 \neq q_2$. Without loss of generality, assume $q_1 < q_2$. By Proposition 4, it must then be the case that either $q_1,q_2 \in [\bar{q},0.5)$ or $q_1,q_2 \in (0.5,1-\bar{q}]$. To fix ideas, we assume the latter for the time being.

Claim 20 guarantees that $f_n(p,q,\gamma^*(n,c)) = 0$, together with the constraint $p \in (0,1)$, defines $p_n^*(c,q)$. In order to conclude that this function is smooth, we must first check for the smoothness of γ^* with respect to its second argument.

Proposition 1 guarantees the existence of the electoral equilibrium $(\gamma_B(n,c,0.5),\gamma_R(n,c,0.5))$. Since $n \geq n_0(c)$, Lemma 3 gives the interiority of this electoral equilibrium, which, by Lemma 1, then equals $(\gamma^*(n,c),\gamma^*(n,c)) \in (0,1)^2$. Proposition 1 implies $P_n(0.5\gamma^*(n,c)) = c$. Since P_n is a C^1 function and $P_n'(0.5\gamma^*(n,c)) < 0$ (Lemma 8), the Implicit Function Theorem guarantees that $\gamma^*(n,c)$ will also be continuously differentiable in c for $c \in (0,1)$.

It must also be noted that $\partial f_n(p^*,q,\gamma^*)/\partial p$, where p^* is short for $p_n^*(c,q)$, cannot be zero. In fact, the $\gamma^* > 1/n$ case is covered in Lemma 10, while if $\gamma^* \leq 1/n$, then Lemma 9 and (34) give

$$\begin{aligned} 0 &> \frac{\partial^2}{\partial p^2}U(n,c,p^*,q) = \frac{n\gamma^*}{2}\left(\frac{1}{p^{*3}(1-p^*)^3}\frac{\partial}{\partial p}f_n(p^*,q,\gamma^*) + f_n(p^*,q,\gamma^*)\frac{\partial}{\partial p}\left(\frac{1}{p^{*3}(1-p^*)^3}\right)\right) \\ &= \frac{n\gamma^*}{2}\left(\frac{1}{p^{*3}(1-p^*)^3}\frac{\partial}{\partial p}f_n(p^*,q,\gamma^*) + 0\right) \sim \frac{\partial}{\partial p}f_n(p^*,q,\gamma^*). \end{aligned}$$

Now, since f_n is C^1 as well (it is polynomial), a new application of the Implicit Function Theorem, but now to the identity provided in Claim 20, guarantees that p_n^* itself is C^1 and

$$\frac{\partial}{\partial q}p_n^*\left(c,q\right)=-\frac{\frac{\partial}{\partial q}f_n\left(p_n^*\left(c,q\right),q,\gamma^*\left(n,c\right)\right)}{\frac{\partial}{\partial p}f_n\left(p_n^*\left(c,q\right),q,\gamma^*\left(n,c\right)\right)} \tag{35}$$

at any q. The most important feature of this identity is that the derivative in the numerator is only partial (with respect to the second argument).

From Rolle's Theorem applied to $p_n^*\left(c,\cdot\right)$, there exists $\tilde{q}\in\left(q_1,q_2\right)$ such that $\partial p_n^*\left(c,\tilde{q}\right)/\partial q=0$. Expression (35) then yields $\partial f_n\left(p_n^*\left(c,\tilde{q}\right),\tilde{q},\gamma^*\right)/\partial q=0$, where $\gamma^*=\gamma^*\left(n,c\right)$. In other words, $f_n\left(p_n^*\left(c,\tilde{q}\right),\cdot,\gamma^*\right)$ is a quadratic in its second argument, as can be seen in (33), with $\tilde{q}$ as a root, and at which its slope is zero. This implies that this quadratic has a zero discriminant.

However, writing $\tilde{p}$ for $p_n^*\left(c,\tilde{q}\right)$, its discriminant can be computed as

$$\begin{aligned}
&\left(\left(2n\gamma^*-2\right)\tilde{p}^4+\left(4-2n\gamma^*-2\gamma^*\right)\tilde{p}^3+\left(3n\gamma^*-3\right)\tilde{p}^2+\left(1-n\gamma^*\right)\tilde{p}\right)^2\\
&-4\left(\left(2\gamma^*-2n\gamma^*\right)\tilde{p}^3+\left(3n\gamma^*-3\gamma^*\right)\tilde{p}^2+\left(3\gamma^*-3n\gamma^*\right)\tilde{p}-\gamma^*+n\gamma^*\right)\left(\tilde{p}^3\left(\gamma^*-1\right)-\tilde{p}^4\left(n\gamma^*-1\right)\right)\\
\mathrel{\mathring{=}}\;&\tilde{p}^2\left(1-\tilde{p}\right)^2\times\left[\begin{array}{c}\left(4n^2\gamma^{*2}-8n\gamma^*+4\right)\tilde{p}^4+\left(-8n^2\gamma^{*2}+16n\gamma^*-8\right)\tilde{p}^3\\+\left(8n^2\gamma^{*2}+4n\gamma^{*2}-20n\gamma^*-4\gamma^{*2}+4\gamma^*+8\right)\tilde{p}^2\\+\left(12n\gamma^*-4n\gamma^{*2}-4n^2\gamma^{*2}+4\gamma^{*2}-4\gamma^*-4\right)\tilde{p}+n^2\gamma^{*2}-2n\gamma^*+1\end{array}\right]\\
=\;&\tilde{p}^2\left(1-\tilde{p}\right)^2\times\left[\begin{array}{c}4\left(n\gamma^*-1\right)^2\tilde{p}^4-8\left(n\gamma^*-1\right)^2\tilde{p}^3+\left(8\left(n\gamma^*-1\right)^2-4\gamma^*\left(1-\gamma^*\right)\left(n-1\right)\right)\tilde{p}^2\\+\left(-4\left(n\gamma^*-1\right)^2+4\gamma^*\left(1-\gamma^*\right)\left(n-1\right)\right)\tilde{p}+\left(n\gamma^*-1\right)^2\end{array}\right]\\
=\;&\tilde{p}^2\left(1-\tilde{p}\right)^2\left[\left(-2\tilde{p}+2\tilde{p}^2+1\right)^2\left(n\gamma^*-1\right)^2+4\tilde{p}\left(1-\tilde{p}\right)\gamma^*\left(1-\gamma^*\right)\left(n-1\right)\right]>0,
\end{aligned}$$

a contradiction.

Also, if it were the case that $q_1,q_2\in\left[\bar{q},0.5\right)$, then, using Claim 19, $p_n^*\left(c,q_1\right)=p_n^*\left(c,q_2\right)$ would imply $p_n^*\left(c,1-q_1\right)=1-p_n^*\left(c,q_1\right)=1-p_n^*\left(c,q_2\right)=p_n^*\left(c,1-q_2\right)$, where $1-q_1\neq 1-q_2$. Thus, by the argument given above, we again arrive at a contradiction.

Therefore, $p_n^*\left(c,\cdot\right)$ is indeed one to one. ■

Proof of Lemma 4. Suppose $q\in\left(0.5,1-\bar{q}\right]$. First note that, in this case, $\phi\left(q\right)<q$, so that $\left(\phi\left(q\right),q\right)$ is an nondegenerate interval indeed. In fact, since $0<1/q-1<1$, we have $\sqrt{1/q-1}>1/q-1$, so that $\phi\left(q\right)=\left(1+\sqrt{1/q-1}\right)^{-1}<\left(1+1/q-1\right)^{-1}=q$. Also, note that $\phi\left(q\right)=\left(1+\sqrt{1/q-1}\right)^{-1}>\left(1+\sqrt{1/0.5-1}\right)^{-1}=0.5$ (if that were not the case, the present proof could actually be done by a simple call to Proposition 4).

Now, as already noted in the proof of Proposition 4, we have

$$f_n\left(q,q,\gamma^*\right)\mathrel{\mathring{=}}-\left(1-\gamma^*\right)q^2\left(2q-1\right)\left(1-q\right)^2\sim 1-2q<0,$$

where $\gamma^* = \gamma^*(n,c)$. Also, by (33),

$$f_n(\phi(q), q, \gamma^*) \mathrel{\mathring{=}} (n-1)\gamma^* \frac{q^2(1-q)(2q-1)\sqrt{\frac{1}{q}-1}}{1+4q(1-q)+4q\sqrt{\frac{1}{q}-1}} > 0.$$

Thus, by the Intermediate Value Theorem, $f_n(\cdot, q, \gamma^*)$ has a root between $\phi(q)$ and q which, by Claim 20, is necessarily $p_n^*(c,q)$.

If $q \in [\bar{q}, 0.5)$, then $1-q \in (0.5, 1-\bar{q}]$, so that we have already proved that $p_n^*(c, 1-q) \in (\phi(1-q), 1-q)$. Now, $p_n^*(c,1-q) = 1-p_n^*(c,q)$ by Claim 19, and, from $\sqrt{1/(1-q)-1}\sqrt{1/q-1} = \sqrt{q/(1-q)}\sqrt{(1-q)/q} = 1$, we get

$$\begin{aligned} \phi(1-q) &= \left(1+\sqrt{\frac{1}{1-q}-1}\right)^{-1} = \left(1+\frac{1}{\sqrt{\frac{1}{q}-1}}\right)^{-1} = \frac{\sqrt{\frac{1}{q}-1}}{1+\sqrt{\frac{1}{q}-1}} = 1-\frac{1}{1+\sqrt{\frac{1}{q}-1}} \\ &= 1-\ \phi(q). \end{aligned}$$

Therefore, $1-p_n^*(c,q) \in (1-\phi(q), 1-q)$, i.e., $p_n^*(c,q) \in (q, \phi(q))$. ■

Proof of Proposition 5. i) Under conditions $c \in (0,1)$ and $n \geq n_0(c)$, as argued in the proof of Lemma 11, $p_n^*(c,\cdot)$ is a smooth, and hence continuous, function (this in itself could also be obtained through Berge's Theorem). Since it it also one to one by that lemma, it is strictly monotone (the real analysis lemma we hereby employ follows immediately from the Intermediate Value Function). Finally, as to whether it is strictly increasing or strictly decreasing: since, by Proposition 4, $p_n^*(c,\bar{q}) < 0.5 = p_n^*(c, 0.5)$, it is strictly increasing.

ii) Using the notation established in Lemma 11, we have, from the Implicit Function Theorem and the Chain Rule,

$$\frac{\partial}{\partial c} p_n^*(c,q) = -\frac{\frac{\partial}{\partial \gamma} f_n(p_n^*(c,q), q, \gamma^*(n,c))}{\frac{\partial}{\partial p} f_n(p_n^*(c,q), q, \gamma^*(n,c))} \frac{\partial}{\partial c}\gamma^*(n,c). \tag{36}$$

By Lemma 3 and Proposition 2, $\partial\gamma^*(n,c)/\partial c = \partial\gamma_B(n,c,0.5)/\partial c < 0$. Also, as explained in the proof of Lemma 11, $\partial f_n(p^*, q, \gamma^*)/\partial p < 0$, where p^* and γ^* stand for $p_n^*(c,q)$ and $\gamma^*(n,c)$. Therefore, $\partial p_n^*(c,q)/\partial c \sim -\partial f_n(p^*, q, \gamma^*)/\partial\gamma$.

Now, if we write expression (33) as $f_n(p,q,\gamma) = \kappa_n(p,q) + \lambda_n(p,q)\gamma$, since $f_n(p^*, q, \gamma^*) = 0$ by Claim 20, we have

$$\begin{aligned} & \frac{\partial}{\partial \gamma} f_n(p^*, q, \gamma^*) = \lambda_n(p^*, q) = -\frac{\kappa_n(p^*, q)}{\gamma^*} \sim -\kappa_n(p^*, q) \\ = \ & (1-2q)p^{*4} + (1-4q)p^{*3} + 3qp^{*2} - qp^* = p^*(1-p^*)\left((1-2q)p^{*2} + 2qp^* - q\right) \\ \sim \ & (1-2q)p^{*2} + 2qp^* - q = Q(p^*), \end{aligned}$$

where $Q(x) := (1-2q)x^2 + 2qx - q$. Thus, $\partial p_n^*(c,q)/\partial c \sim -Q(p^*)$, and, in order to conclude that $\partial p_n^*(c,q)/\partial c \lesseqgtr 0$ if $q \gtreqless 0.5$, it suffices to argue that $Q(p^*) \sim 2q-1$.

If $q = 0.5$, then $Q(p^*) = p^* - 0.5$, which is 0 by Proposition 4.

If $q \neq 0.5$, then $Q(x)$ is a quadratic in x, and $\phi(q)$ happens to be one of its roots:

$$\begin{aligned}
Q(\phi(q)) &= (1-2q)(\phi(q))^2 + 2q\phi(q) - q = \frac{1-2q}{\left(1+\sqrt{\frac{1}{q}-1}\right)^2} + \frac{2q}{1+\sqrt{\frac{1}{q}-1}} - q \\
&= \frac{1-2q+2q\left(1+\sqrt{\frac{1}{q}-1}\right) - q\left(1+\sqrt{\frac{1}{q}-1}\right)^2}{\left(1+\sqrt{\frac{1}{q}-1}\right)^2} \\
&= \frac{1+2q\sqrt{\frac{1}{q}-1} - q\left(1+2\sqrt{\frac{1}{q}-1}+\frac{1}{q}-1\right)}{\left(1+\sqrt{\frac{1}{q}-1}\right)^2} = 0.
\end{aligned}$$

Call its second root $\phi_2(q)$.

In the $q > 0.5$ case, Q is concave and $\phi_2(q) > 1 > \phi(q)$, since $Q(1) = 1-2q+2q-q > 0$. Thus, $Q(x) > 0, \forall x \in (\phi(q), q) \subseteq (\phi(q), 1)$, so that, by Lemma 4, $Q(p^*) > 0$.

In the $q < 0.5$ case, Q is convex and $\phi_2(q) < 0 < \phi(q)$, since $Q(0) = -q < 0$. Thus, $Q(x) < 0, \forall x \in (q, \phi(q)) \subseteq (0, \phi(q))$, so that, by Lemma 4, $Q(p^*) < 0$ and we are done. ■

Before going on to prove Proposition 6, we must establish a bit more notation. Given $c \in (0,1)$, we know, from Lemma 2, that $\lim_{n\to\infty} \gamma^*(n,c) = 0$ and that $\lim_{n\to\infty} n\gamma^*(n,c) = m(c) > 0$. Given $q \in [\overline{q}, 1-\overline{q}]$, we can define, for all $p \in [\overline{q}, 1-\overline{q}]$,

$$\hat{U}(c,p,q) = \left[-\frac{m(c)}{2}\frac{q}{p} - \frac{(m(c))^2}{4}\frac{(p-q)^2}{p^2}\right] + \left[-\frac{m(c)}{2}\frac{1-q}{1-p} - \frac{(m(c))^2}{4}\frac{(p-q)^2}{(1-p)^2}\right]. \tag{37}$$

Due to those limits given in Lemma 2, it is just a matter of comparing this expression to (32) to note that $\hat{U}(c,p,q)$ is simply the pointwise limit of $(U(n,c,p,q))_{n\in\mathbb{N}}$, viewed as a sequence of functions of p.

It will be important in the arguments below to note that this convergence is actually uniform. Since p takes its values on the compact set $[\overline{q}, 1-\overline{q}]$, which is bounded away from 0 and 1, not only the coefficients of the terms $p^{-1}, p^{-2}, (1-p)^{-1}, (1-p)^2$ appearing in (32) are bounded, but also these very terms. Since the coefficients converge uniformly simply by being constant in p and pointwise convergent, and the terms converge uniformly to themselves since they have no dependence on n, we have that $(U(n,c,p,q))_{n\in\mathbb{N}}$ converges uniformly to $\hat{U}(c,p,q)$.[25]

[25] This is just a matter of repeatedly using exercise 2 in Rudin (1976, ch. 7).

Proof of Proposition 6. i) Assume $q = 0.5$. By Proposition 4, $(p_n^*(c, 0.5))_{n\in\mathbb{N}}$ is an eventually constant at 0.5 sequence. Therefore, $\lim_{n\to\infty} p_n^*(c, q) = 0.5 = q$.

ii) Assume $q > 0.5$, and let $n \geq n_0(c)$. Let us first solve an auxiliary problem, that of maximizing $\hat{U}(c, p, q)$ for $p \in [\bar{q}, 1 - \bar{q}]$. It may be checked, just as in (34), that, for all $p \in (\bar{q}, 1 - \bar{q})$,

$$\frac{\partial}{\partial p}\hat{U}(c, p, q) = \frac{1}{2}\frac{m(c)}{p^3(1-p)^3}\hat{f}(c, p, q),$$

where

$$\begin{aligned}\hat{f}(c, p, q) := {} & (2m(c)q - m(c) - 2q + 1)p^4 + \left(4q - 2m(c)q - 2m(c)q^2 - 1\right)p^3 \\ & + \left(3m(c)q - 3q + 3m(c)q^2\right)p^2 + \left(q - m(c)q - 3m(c)q^2\right)p + m(c)q^2.\end{aligned}$$

Since

$$\hat{f}(c, 0.5, q) \mathrel{\mathring{=}} \frac{1}{16}(2q-1)(m(c)+1) > 0$$

and

$$\hat{f}(c, q, q) \mathrel{\mathring{=}} -q^2(2q-1)(1-q)^2 < 0,$$

by the Intermediate Value Theorem, there exists a root of $\hat{f}(c, \cdot, q)$ in $(0.5, q)$, which we shall call $p_\infty^*(c, q)$.

Inspired by the roles of Lemmas 9 and 10 in the proof of Proposition 4, let us split our analysis in two.

On one hand, if $m(c) \leq 1$, then we have $\partial^2\hat{U}(c, p, q)/\partial p^2 < 0, \forall p \in (\bar{q}, 1 - \bar{q})$. In order to see this, it is just a matter of applying the $\partial^2/\partial p^2$ operator to each term in square brackets in (37) in turn. The first one gives

$$\frac{\partial^2}{\partial p^2}\left(-\frac{m(c)}{2}\frac{q}{p} - \frac{(m(c))^2}{4}\frac{(p-q)^2}{p^2}\right) \mathrel{\mathring{=}} -\frac{1}{2}\frac{m(c)}{p^4}q(2p - 2m(c)p + 3m(c)q) \sim (2p - 3q)m(c) - 2p.$$

This is negative because it is affine in $m(c)$, it is equal to $-2p < 0$ when $m(c) = 0$, and equal to $-3q < 0$ when $m(c) = 1$. Now the second one:

$$\begin{aligned}& \frac{\partial^2}{\partial p^2}\left(-\frac{m(c)}{2}\frac{1-q}{1-p} - \frac{(m(c))^2}{4}\frac{(p-q)^2}{(1-p)^2}\right) \\ \mathrel{\mathring{=}} {} & -\frac{1}{2}\frac{m(c)}{(1-p)^4}(1-q)(m(c) - 2p + 2m(c)p - 3m(c)q + 2) \sim (3q - 2p - 1)m(c) + 2p - 2.\end{aligned}$$

Again, this is negative because it is affine in $m(c)$, it is equal to $2p - 2 < 0$ when $m(c) = 0$, and equal to $3q - 3 < 0$ when $m(c) = 1$.

On the other hand, if $m(c) > 1$, then $\hat{f}(c,p,q)$ is strictly decreasing in p for $p \in [\bar{q}, 1-\bar{q}]$. In fact, given any $p \in (\bar{q}, 1-\bar{q})$, we have

$$\begin{aligned}\frac{\partial}{\partial p}\hat{f}(c,p,q) &\mathrel{\mathring{=}} \left(-3q^2 - 4p^3 - q + 6pq^2 - 6p^2q + 8p^3q - 6p^2q^2 + 6pq\right) m(c) \\ &\quad + q + 12p^2q - 8p^3q - 6pq - 3p^2 + 4p^3,\end{aligned}$$

once more an affine expression in $m(c)$. At $m(c) = 1$, this equals

$$-3\left(-2pq^2 - 2p^2q + 2p^2q^2 + p^2 + q^2\right) = -3\left(p^2(1-q)^2 + q^2(1-p)^2\right) < 0.$$

As for the behavior of $\partial\hat{f}(c,p,q)/\partial p$ when $m(c) \to +\infty$, note that $\partial\hat{f}(c,p,q)/\partial p = g(p,q)\,m(c) + q + 12p^2q - 8p^3q - 6pq - 3p^2 + 4p^3$, where we have borrowed the g notation from the proof of Lemma 10. Since we show there that $g(p,q) < 0$, we have $\lim_{m(c)\to+\infty} \partial\hat{f}(c,p,q)/\partial p = -\infty$, whence $\partial\hat{f}(c,p,q)/\partial p < 0, \forall p \in (\bar{q}, 1-\bar{q})$.

Thus, in either case, $p^*_\infty(c,q)$ is the one and only solution to $\max_{p\in[\bar{q},1-\bar{q}]} \hat{U}(c,p,q)$.

Finally, since $U(n,c,p,q)$ converges uniformly to $\hat{U}(c,p,q)$ for $p \in [\bar{q}, 1-\bar{q}]$ and we have shown that the $\arg\max_{p\in[\bar{q},1-\bar{q}]} \hat{U}(c,p,q)$ is a singleton ($\{p^*_\infty(c,q)\}$), it follows from Theorem 2.2 of Schochetman (1990) that $\lim_{n\to\infty} p^*_n(c,q) = p^*_\infty(c,q)$, which, as we have already shown, belongs to the $(0.5, q)$ interval.

iii) Assume $q < 0.5$. Then, by the argument above, $0.5 < \lim_{n\to\infty} p^*_n(c, 1-q) < 1-q$. By Claim 19, $\lim_{n\to\infty} p^*_n(c,1-q) = \lim_{n\to\infty}(1 - p^*_n(c,q)) = 1 - \lim_{n\to\infty} p^*_n(c,q)$. Therefore, $0.5 < 1 - \lim_{n\to\infty} p^*_n(c,q) < 1-q$, or yet $0.5 > \lim_{n\to\infty} p^*_n(c,q) > q$, as wished. ■

Before tackling the proof of Proposition 7, we state a couple of lemmas more directly linked to Probability Theory.

Proof of Lemma 6. By Levy's Continuity Theorem, it suffices to show that, for each $t \in \mathbb{R}$, $\varphi_{\text{MultiDiff}(n, q\gamma_B(n,c,p^*_n(c,q)), (1-q)\gamma_R(n,c,p^*_n(c,q)))}(t)$ converges, as $n \to \infty$, to $\varphi_{\text{Skellam}((q/p^*_\infty(c,q))m(c)/2, ((1-q)/(1-p^*_\infty(c,q)))m(c)/2)}(t)$. In turn, the characteristic function of $\text{Skellam}(m_B, m_R)$ is given by $\varphi_{\text{Skellam}(m_B,m_R)}(t) = \exp\left(m_B\left(e^{it}-1\right) + m_R\left(e^{-it}-1\right)\right)$.[26]

Let $n \geq n_0(c)$, so that, by Lemmas 3 and 1, $\gamma_B(n,c,p^*_n(c,q)) = \gamma^*(n,c)/(2p^*_n(c,q))$ and

[26] This can be verified in Skellam (1946), where the probability generating function of this distribution is provided: $G(t) = \exp\left(m_B t + m_R t^{-1} - m_B - m_R\right)$. Then it is just a matter of writing $\varphi_{\text{Skellam}(m_B,m_R)}(t) = G\left(e^{it}\right)$.

$\gamma_R(n,c,p_n^*(c,q)) = \gamma^*(n,c)/(2(1-p_n^*(c,q)))$. For any $t \in \mathbb{R}$, we thus have

$$\begin{aligned}
& \varphi_{\text{MultiDiff}(n,q\gamma_B(n,c,p_n^*(c,q)),(1-q)\gamma_R(n,c,p_n^*(c,q)))}(t) \\
= \; & \left(1+q\gamma_B(n,c,p_n^*(c,q))\left(e^{it}-1\right)+(1-q)\gamma_R(n,c,p_n^*(c,q))\left(e^{-it}-1\right)\right)^n \\
= \; & \left(1+\frac{q}{p_n^*(c,q)}\frac{\gamma^*(n,c)}{2}\left(e^{it}-1\right)+\frac{1-q}{1-p_n^*(c,q)}\frac{\gamma^*(n,c)}{2}\left(e^{-it}-1\right)\right)^n \\
= \; & \left(1+\frac{\frac{q}{p_n^*(c,q)}\frac{n\gamma^*(n,c)}{2}\left(e^{it}-1\right)+\frac{1-q}{1-p_n^*(c,q)}\frac{n\gamma^*(n,c)}{2}\left(e^{-it}-1\right)}{n}\right)^n .
\end{aligned}$$

At the same time, note that

$$\begin{aligned}
& \lim_{n\to\infty}\left(\frac{q}{p_n^*(c,q)}\frac{n\gamma^*(n,c)}{2}\left(e^{it}-1\right)+\frac{1-q}{1-p_n^*(c,q)}\frac{n\gamma^*(n,c)}{2}\left(e^{-it}-1\right)\right) \\
= \; & \frac{q}{p_\infty^*(c,q)}\frac{m(c)}{2}\left(e^{it}-1\right)+\frac{1-q}{1-p_\infty^*(c,q)}\frac{m(c)}{2}\left(e^{-it}-1\right),
\end{aligned}$$

by Proposition 6. Therefore,

$$\begin{aligned}
& \lim_{n\to\infty}\varphi_{\text{MultiDiff}(n,q\gamma_B(n,c,p_n^*(c,q)),(1-q)\gamma_R(n,c,p_n^*(c,q)))}(t) \\
= \; & \exp\left(\frac{q}{p_\infty^*(c,q)}\frac{m(c)}{2}\left(e^{it}-1\right)+\frac{1-q}{1-p_\infty^*(c,q)}\frac{m(c)}{2}\left(e^{-it}-1\right)\right) \\
= \; & \varphi_{\text{Skellam}\left(\frac{q}{p_\infty^*(c,q)}\frac{m(c)}{2},\frac{1-q}{1-p_\infty^*(c,q)}\frac{m(c)}{2}\right)}(t).
\end{aligned}$$

■

Claim 21 *A coin that lands heads up with probability $s \in [0,1]$ is tossed $l \geq 1$ times. Given any $k \in \{1,\ldots,l\}$, the probability of obtaining at least k heads out of these l tosses is strictly increasing in s.*

Proof. Let $F_{l,s}$ denote the cumulative distribution function of Binomial(l,s). Then for any $k \in \{1,\ldots,l\}$, the probability of obtaining at least k heads is

$$1-F_{l,s}(k-1) = \int_0^s y^{k-1}(1-y)^{l-k}\,dy \Big/ \int_0^1 y^{k-1}(1-y)^{l-k}\,dy,$$

obviously strictly increasing in s.[27] ■

Proof of Lemma 5. Given these hypotheses, Lemmas 3 and 1 are applicable. If the pollster

[27] The preceding formula is provided in Wadsworth and Bryan (1974, p. 51).

is truthful and reports q, the distribution of (b, r, a) in (5) becomes

$$\begin{aligned} & \text{Multinomial}\left(n, q\gamma_B(n,c,q), (1-q)\gamma_R(n,c,q), 1-q\gamma_B(n,c,q)-(1-q)\gamma_R(n,c,q)\right) \\ = & \text{Multinomial}\left(n, \frac{\gamma^*(n,c)}{2}, \frac{\gamma^*(n,c)}{2}, 1-\gamma^*(n,c)\right), \end{aligned}$$

symmetric in (b, r). Therefore, $\Pr(B \text{ wins} \mid n, c, q, q) = \Pr(R \text{ wins} \mid n, c, q, q) = 0.5$.

i) If $q = 0.5$, then Proposition 4 gives $p_n^*(c, q) = q$. Therefore, as shown above, $\Pr(B \text{ wins} \mid n, c, p_n^*(c, q), q) = \Pr(B \text{ wins} \mid n, c, q, q) = 0.5$.

ii) If $q > 0.5$, let us first argue that, conditional on the number of abstentions a, $\Pr(b > r \mid n, c, p_n^*(c, q), q, a) > \Pr(b < r \mid n, c, p_n^*(c, q), q, a)$ (unless $a = n$, in which case both probabilities vanish). In fact, given $a \in \{0, \ldots, n-1\}$, b will be distributed as $\text{Binomial}(n-a, s)$, where

$$\begin{aligned} s &= \frac{q\gamma_B(n,c,p_n^*(c,q))}{q\gamma_B(n,c,p_n^*(c,q)) + (1-q)\gamma_R(n,c,p_n^*(c,q))} = \frac{q\frac{\gamma^*(n,c)}{2p_n^*(c,q)}}{q\frac{\gamma^*(n,c)}{2p_n^*(c,q)} + (1-q)\frac{\gamma^*(n,c)}{2(1-p_n^*(c,q))}} \\ &= \frac{\frac{q}{p_n^*(c,q)}}{\frac{q}{p_n^*(c,q)} + \frac{1-q}{1-p_n^*(c,q)}} = \frac{1}{1 + \frac{\frac{1}{q}-1}{\frac{1}{p_n^*(c,q)}-1}} > \frac{1}{1+1} = 0.5, \end{aligned}$$

where we first used Lemmas 3 and 1 to get rid of the γ_B and γ_R terms, and then applied Proposition 4 to conclude that, since $0 < p_n^*(c, q) < q < 1$, $0 < (1/q - 1) / (1/p_n^*(c, q) - 1) < 1$. Note that this probability parameter of the distribution of b would be exactly 0.5 if pre-election poll results could not be misreported:

$$\frac{q\gamma_B(n,c,q)}{q\gamma_B(n,c,q) + (1-q)\gamma_R(n,c,q)} = \frac{q\frac{\gamma^*(n,c)}{2q}}{q\frac{\gamma^*(n,c)}{2q} + (1-q)\frac{\gamma^*(n,c)}{2(1-q)}} = \frac{1}{1+1} = 0.5.$$

Since, conditional on a, the event $b > r$ could also be written as $b \geq \lfloor (n-a)/2 \rfloor + 1$ (which is at least 1), Claim 21 then yields that the probability of this event under misreporting is larger that it would be under truthful reporting of pre-election poll results: $\Pr(b > r \mid n, c, p_n^*(c, q), q, a) > \Pr(b > r \mid n, c, q, q, a)$.

Similarly, conditional on a, $r \sim \text{Binomial}(n-a, 1-s)$, where $1-s < 0.5$, and the event $b < r$ is the same as $r \geq \lfloor (n-a)/2 \rfloor + 1$, so Claim 21 gives $\Pr(b < r \mid n, c, p_n^*(c, q), q, a) < \Pr(b < r \mid n, c, q, q, a)$.

Finally, note that $\Pr(b > r \mid n, c, q, q, a) = \Pr(b < r \mid n, c, q, q, a)$ (if we toss an unbiased coin $n-a$ times, the probability of obtaining more heads than tails equals the probability of obtaining more tails than heads). Therefore, for all $\bar{a} \in \{0, \ldots, n-1\}$,

$$\begin{aligned} & \Pr(b > r \mid n, c, p_n^*(c,q), q, \bar{a}) > \Pr(b > r \mid n, c, q, q, \bar{a}) \\ = & \Pr(b < r \mid n, c, q, q, \bar{a}) > \Pr(b < r \mid n, c, p_n^*(c,q), q, \bar{a}), \end{aligned}$$

while $\Pr\left(b > r \mid n, c, p_n^*(c,q), q, n\right) = \Pr\left(b < r \mid n, c, p_n^*(c,q), q, n\right) = 0$. By the Law of Iterated Expectations,

$$\begin{aligned} & \Pr\left(b > r \mid n, c, p_n^*(c,q), q\right) = \mathrm{E}\left(\Pr\left(b > r \mid n, c, p_n^*(c,q), q, a\right)\right) \\ > \; & \mathrm{E}\left(\Pr\left(b < r \mid n, c, p_n^*(c,q), q, a\right)\right) = \Pr\left(b < r \mid n, c, p_n^*(c,q), q\right), \end{aligned}$$

where expectations are taken with respect to the distribution of a (it is a marginal distribution of (5), $a \sim \text{Binomial}\left(n, 1 - q\gamma_B\left(n, c, p_n^*(c,q)\right) - (1-q)\gamma_R\left(n, c, p_n^*(c,q)\right)\right)$).

Thus,

$$\begin{aligned} & \Pr\left(B \text{ wins} \mid n, c, p_n^*(c,q), q\right) = \Pr\left(b > r \mid n, c, p_n^*(c,q), q\right) + \frac{\Pr\left(b = r \mid n, c, p_n^*(c,q), q\right)}{2} \\ > \; & \Pr\left(b < r \mid n, c, p_n^*(c,q), q\right) + \frac{\Pr\left(b = r \mid n, c, p_n^*(c,q), q\right)}{2} = \Pr\left(R \text{ wins} \mid n, c, p_n^*(c,q), q\right) \\ = \; & 1 - \Pr\left(B \text{ wins} \mid n, c, p_n^*(c,q), q\right), \end{aligned}$$

that is, $\Pr\left(B \text{ wins} \mid n, c, p_n^*(c,q), q\right) > 0.5$.

iii) If $q < 0.5$, then the proof is entirely analogous, the difference being that now Proposition 4 implies $s < 0.5$, since $0 < q < p_n^*(c,q) < 1$. ■

Lemma 12 *Given $q \in (0.5, 1-\overline{q}]$ and $\mu > 0$, let $Z_p \sim \text{Skellam}\left((q/p)\mu, ((1-q)/(1-p))\mu\right)$, for all $p \in (0.5, q]$. Then $\Pr(Z_p > 0) + \Pr(Z_p = 0)/2$ is strictly decreasing in p.*

Proof. Because Z_p is the difference of two independent Poisson-distributed random variables, we may write

$$\Pr(Z_p > 0) = \sum_{i=1}^{\infty}\sum_{j=0}^{i-1} e^{-\frac{q}{p}\mu}\frac{\left(\frac{q}{p}\mu\right)^i}{i!} e^{-\frac{1-q}{1-p}\mu}\frac{\left(\frac{1-q}{1-p}\mu\right)^j}{j!}$$

and

$$\Pr(Z_p = 0) = 1 - \Pr(Z_p > 0) - \Pr(Z_p < 0) = 1 - \Pr(Z_p > 0) - \Pr(-Z_p > 0).$$

Let us call the first and second parameters of the Skellam distribution m_B and m_R, respectively. Now we shall analyze the partial derivatives of the above probabilities with respect to m_B and m_R.

$$\begin{aligned} & \frac{\partial}{\partial m_B}\Pr(Z_p > 0) = -\sum_{i=1}^{\infty}\sum_{j=0}^{i-1} e^{-m_B}\frac{(m_B)^i}{i!}e^{-m_R}\frac{(m_R)^j}{j!} + \sum_{i=1}^{\infty}\sum_{j=0}^{i-1} e^{-m_B}\frac{(m_B)^{i-1}}{(i-1)!}e^{-m_R}\frac{(m_R)^j}{j!} \\ = \; & -\Pr(Z_p > 0) + \sum_{i'=0}^{\infty}\sum_{j=0}^{i'} e^{-m_B}\frac{(m_B)^{i'}}{(i')!}e^{-m_R}\frac{(m_R)^j}{j!} = -\Pr(Z_p > 0) + \Pr(Z_p \geq 0) = \Pr(Z_p = 0). \end{aligned}$$

Since we can also write

$$\Pr\left(Z_p > 0\right) = \sum_{i=1}^{\infty} e^{-\frac{q}{p}\mu} \frac{\left(\frac{q}{p}\mu\right)^i}{i!} e^{-\frac{1-q}{1-p}\mu} \left(1 + \sum_{j=1}^{i-1} \frac{\left(\frac{1-q}{1-p}\mu\right)^j}{j!}\right),$$

we have

$$\begin{aligned}
\frac{\partial}{\partial m_R} \Pr\left(Z_p > 0\right) &= -\sum_{i=1}^{\infty} e^{-\frac{q}{p}\mu} \frac{\left(\frac{q}{p}\mu\right)^i}{i!} e^{-\frac{1-q}{1-p}\mu} \left(1 + \sum_{j=1}^{i-1} \frac{\left(\frac{1-q}{1-p}\mu\right)^j}{j!}\right) \\
&\quad + \sum_{i=1}^{\infty} e^{-\frac{q}{p}\mu} \frac{\left(\frac{q}{p}\mu\right)^i}{i!} e^{-\frac{1-q}{1-p}\mu} \sum_{j=1}^{i-1} \frac{\left(\frac{1-q}{1-p}\mu\right)^{j-1}}{(j-1)!} \\
&= -\Pr\left(Z_p > 0\right) + \sum_{i=1}^{\infty} e^{-\frac{q}{p}\mu} \frac{\left(\frac{q}{p}\mu\right)^i}{i!} e^{-\frac{1-q}{1-p}\mu} \sum_{j'=0}^{i-2} \frac{\left(\frac{1-q}{1-p}\mu\right)^{j'}}{j'!} \\
&= -\Pr\left(Z_p > 0\right) + \Pr\left(Z_p \geq 2\right) = -\Pr\left(Z_p = 1\right).
\end{aligned}$$

By the symmetry property of the Skellam distribution ($Z \sim \mathrm{Skellam}\left(m_B, m_R\right) \Leftrightarrow -Z \sim \mathrm{Skellam}\left(m_R, m_B\right)$), we have

$$\frac{\partial}{\partial m_B} \Pr\left(-Z_p > 0\right) = -\Pr\left(-Z_p = 1\right) = -\Pr\left(Z_p = -1\right)$$

and

$$\frac{\partial}{\partial m_R} \Pr\left(-Z_p > 0\right) = \Pr\left(-Z_p = 0\right) = \Pr\left(Z_p = 0\right).$$

Therefore,

$$\begin{aligned}
\frac{\partial}{\partial p} \Pr\left(Z_p > 0\right) &= \frac{\partial \Pr\left(Z_p > 0\right)}{\partial m_B} \frac{\partial m_B}{\partial p} + \frac{\partial \Pr\left(Z_p > 0\right)}{\partial m_R} \frac{\partial m_R}{\partial p} \\
&= \Pr\left(Z_p = 0\right) \left(-\frac{1}{2} \frac{m}{p^2} q\right) - \Pr\left(Z_p = 1\right) \frac{1}{2} \frac{m}{(1-p)^2} (1-q) \\
&= -\frac{m}{2} \left(\frac{q}{p^2} \Pr\left(Z_p = 0\right) + \frac{1-q}{(1-p)^2} \Pr\left(Z_p = 1\right)\right)
\end{aligned}$$

and

$$\begin{aligned}\frac{\partial}{\partial p}\Pr\left(Z_p<0\right) &= \frac{\partial \Pr\left(Z_p<0\right)}{\partial m_B}\frac{\partial m_B}{\partial p}+\frac{\partial \Pr\left(Z_p<0\right)}{\partial m_R}\frac{\partial m_R}{\partial p}\\ &= -\Pr\left(Z_p=-1\right)\left(-\frac{1}{2}\frac{m}{p^2}q\right)+\Pr\left(Z_p=0\right)\frac{1}{2}\frac{m}{(1-p)^2}(1-q)\\ &= \frac{m}{2}\left(\frac{q}{p^2}\Pr\left(Z_p=-1\right)+\frac{1-q}{(1-p)^2}\Pr\left(Z_p=0\right)\right).\end{aligned}$$

Now note that

$$\begin{aligned}\Pr\left(Z_p>0\right)+\frac{1}{2}\Pr\left(Z_p=0\right) &= \Pr\left(Z_p>0\right)+\frac{1}{2}\left(1-\Pr\left(Z_p>0\right)-\Pr\left(Z_p<0\right)\right)\\ &= \frac{1}{2}\left(1+\Pr\left(Z_p>0\right)-\Pr\left(Z_p<0\right)\right).\end{aligned}$$

Thus,

$$\begin{aligned}&\frac{d}{dp}\left(\Pr\left(Z_p>0\right)+\frac{1}{2}\Pr\left(Z_p=0\right)\right)=\frac{1}{2}\left(\frac{\partial}{\partial p}\Pr\left(Z_p>0\right)-\frac{\partial}{\partial p}\Pr\left(Z_p<0\right)\right)\\ &= \frac{1}{2}\left(\begin{array}{c}-\frac{m}{2}\left(\frac{q}{p^2}\Pr\left(Z_p=0\right)+\frac{1-q}{(1-p)^2}\Pr\left(Z_p=1\right)\right)\\ -\frac{m}{2}\left(\frac{q}{p^2}\Pr\left(Z_p=-1\right)+\frac{1-q}{(1-p)^2}\Pr\left(Z_p=0\right)\right)\end{array}\right)\\ &= -\frac{m}{4}\left(\left(\frac{q}{p^2}+\frac{1-q}{(1-p)^2}\right)\Pr\left(Z_p=0\right)+\frac{1-q}{(1-p)^2}\Pr\left(Z_p=1\right)+\frac{q}{p^2}\Pr\left(Z_p=-1\right)\right)<0.\end{aligned}$$

■

Proof of Lemma 7. Since $\Pr\left(B \text{ wins} \mid n,c,p_n^*(c,q),q\right) = \Pr\left(b-r>0 \mid n,c,p_n^*(c,q),q\right) + \Pr\left(b-r=0 \mid n,c,p_n^*(c,q),q\right)/2$, Lemma 6 yields $\lim_{n\to\infty}\Pr\left(B \text{ wins} \mid n,c,p_n^*(c,q),q\right) = \Pr(Z>0)+\Pr(Z=0)/2$, where $Z \sim \text{Skellam}\left(\left(q/p_\infty^*(c,q)\right)m(c)/2,\left((1-q)/\left(1-p_\infty^*(c,q)\right)\right)m(c)/2\right)$. Thus, at least the existence of the limit is assured. Let us see how it compares to 0.5.

i) If $q=0.5$, then the thesis follows directly from Lemma 5, since $\Pr\left(B \text{ wins} \mid n,c,p_n^*(c,q),q\right) = 0.5, \forall n \geq n_0(c)$. Alternatively (and this second argument will be useful below), it could also be seen to follow from the fact above, since in this case, by Lemma 6, $p_\infty^*(c,q)=q$, and Z will be distributed as the symmetric $\text{Skellam}\left(m(c)/2,m(c)/2\right)$, so that $\Pr(Z>0)+\Pr(Z=0)/2 = 1-\Pr(Z<0)-\Pr(Z=0)/2 = 1-\Pr(Z>0)-\Pr(Z=0)/2$. Thus, $\lim_{n\to\infty}\Pr\left(B \text{ wins} \mid n,c,p_n^*(c,q),q\right) = \Pr(Z>0)+\Pr(Z=0)/2=0.5$.

ii) If $q>0.5$, then, by Lemma 6, $p_\infty^*(c,q)\in(0.5,q)$. Therefore, by Lemma 12, $\Pr(Z>0)+\Pr(Z=0)/2$ will be strictly larger than $\Pr(Z_q>0)+\Pr(Z_q=0)/2$, where $Z_q \sim \text{Skellam}\left((q/q)m(c)/2,\left((1-q)/(1-q)\right)m(c)/2\right)=\text{Skellam}\left(m(c)/2,m(c)/2\right)$. As explained in part (i), $\Pr(Z_q > 0) + \Pr(Z_q = 0)/2 = 0.5$, so that $\lim_{n\to\infty} \Pr(B \text{ wins} \mid n, c, p_n^*(c,q), q) = \Pr(Z > 0) + \Pr(Z = 0)/2 > \Pr(Z_q > 0) + \Pr(Z_q = 0)/2 = 0.5$.

iii) If $q < 0.5$, then $\Pr(B \text{ wins} \mid n, c, p_n^*(c,q), q) = \Pr(R \text{ wins} \mid n, c, 1 - p_n^*(c,q), 1-q) = 1 - \Pr(B \text{ wins} \mid n, c, p_n^*(c, 1-q), 1-q)$, where we have used Claim 19. Since $1 - q > 0.5$, part (ii) yields $\lim_{n\to\infty} \Pr(B \text{ wins} \mid n, c, p_n^*(c,q), q) = 1 - \lim_{n\to\infty} \Pr(B \text{ wins} \mid n, c, p_n^*(c, 1-q), 1-q) < 1 - 0.5 = 0.5$. ■

Proof of Proposition 7. i) Fix any $n \geq n_0(c)$, so that, by (18), Lemmas 3 and 1,

$$\begin{aligned}
&\mathcal{C}(n,c,q,q) - \mathcal{C}(n,c,p_n^*(c,q),q) \\
= \;& nc\left(q\frac{\gamma^*(n,c)}{2q} + (1-q)\frac{\gamma^*(n,c)}{2(1-q)}\right) - nc\left(q\frac{\gamma^*(n,c)}{2p_n^*(c,q)} + (1-q)\frac{\gamma^*(n,c)}{2(1-p_n^*(c,q))}\right) \\
= \;& nc\gamma^*(n,c)\left(1 - \frac{1}{2}\left(\frac{q}{p_n^*(c,q)} + \frac{1-q}{1-p_n^*(c,q)}\right)\right) \\
\sim \;& 2p_n^*(c,q)(1-p_n^*(c,q)) - ((1-p_n^*(c,q))q + p_n^*(c,q)(1-q)) \\
= \;& (2p_n^*(c,q) - 1)(q - p_n^*(c,q)).
\end{aligned}$$

If $q \gtrless 0.5$, then $p_n^*(c,q) \gtrless 0.5$ and $p_n^*(c,q) \lessgtr q$ due to Proposition 4. In either case, $2p_n^*(c,q) - 1$ and $q - p_n^*(c,q)$ have the same sign, so that $\mathcal{C}(n,c,q,q) - \mathcal{C}(n,c,p_n^*(c,q),q)$ is indeed positive.

ii) Let $J(c,p,q) := (2q-1)(2\Pr(Z_{p,q} > 0) + \Pr(Z_{p,q} = 0) - 1)$, where $Z_{p,q} \sim \text{Skellam}((q/p)\,m(c)/2, ((1-q)/(1-p))\,m(c)/2)$. As done in the proof of Lemma 7, note that (17) and Lemma 6 yield

$$\begin{aligned}
&\lim_{n\to\infty} \frac{I(n,c,p_n^*(c,q),q)}{n} = (2q-1)\left(2\lim_{n\to\infty} \Pr(B \text{ wins} \mid n,c,p_n^*(c,q),q) - 1\right) \\
= \;& (2q-1)\left(2\left(\Pr\left(Z_{p_\infty^*(c,q),q} > 0\right) + \Pr\left(Z_{p_\infty^*(c,q),q} = 0\right)/2\right) - 1\right) = J(c, p_\infty^*(c,q), q).
\end{aligned}$$

This implies that also $\lim_{n\to\infty}(\mathcal{I}(n,c,p_n^*(c,q),q)/n) = J(c,p_\infty^*(c,q),q)$. In fact, a simple variation of the argument given in section 5 would show that $\lim_{n\to\infty}((\mathcal{I}(n,c,p_n^*(c,q),q) - I(n,c,p_n^*(c,q),q))/n) = 0$ (just redefine x_{n,n_B} as $\Pr(B \text{ wins} \mid n,c,p_n^*(c,q),q,n_B)$ and y_n as $\Pr(B \text{ wins} \mid n,c,p_n^*(c,q),q)$). Then, given $\varepsilon > 0$, it is only a matter of choosing $n_1 \in \mathbb{N}$ large enough so that $n \geq n_1$ implies both $|(\mathcal{I}(n,c,p_n^*(c,q),q) - I(n,c,p_n^*(c,q),q))/n| < \varepsilon/2$ and $|I(n,c,p_n^*(c,q),q)/n - J(c,p_\infty^*(c,q),q)| < \varepsilon/2$, and then applying the triangle inequality.

Since $I(n,c,q,q) = (2q-1)(2\Pr(B \text{ wins} \mid n,c,q,q) - 1) = 0$ for all $n \geq n_0(c)$ by Lemma 5, we have $\lim_{n\to\infty}(I(n,c,q,q)/n) = 0$ and $\lim_{n\to\infty}(\mathcal{I}(n,c,q,q)/n) = 0$ (take x_{n,n_B} as

$\Pr(B \text{ wins} \mid n, c, q, q, n_B)$ and y_n as $\Pr(B \text{ wins} \mid n, c, q, q))$. Because

$$\begin{aligned} & J(c,q,q) = (2q-1)\left(2\Pr(Z_{q,q}>0) + \Pr(Z_{q,q}=0) - 1\right) \\ = \; & (2q-1)\left(\Pr(Z_{q,q}>0) + \Pr(Z_{q,q}<0) + \Pr(Z_{q,q}=0) - 1\right) = (2q-1)(1-1) = 0 \end{aligned}$$

by the symmetry of $\text{Skellam}(m(c)/2, m(c)/2)$, we can also write $\lim_{n\to\infty}(\mathcal{I}(n,c,q,q)/n) = J(c,q,q)$.

Therefore,

$$\lim_{n\to\infty} \frac{\mathcal{I}(n,c,p_n^*(c,q),q) - \mathcal{I}(n,c,q,q)}{n} = J(c,p_\infty^*(c,q),q) - J(c,q,q).$$

If $q > 0.5$, then $p_\infty^*(c,q) \in (0.5, q)$ by Proposition 6, so that

$$\begin{aligned} & J(c,p_\infty^*(c,q),q) - J(c,q,q) \\ = \; & 2(2q-1)\left(\Pr\left(Z_{p_\infty^*(c,q),q}>0\right) + \frac{\Pr\left(Z_{p_\infty^*(c,q),q}=0\right)}{2} - \left(\Pr(Z_{q,q}>0) + \frac{\Pr(Z_{q,q}=0)}{2}\right)\right) > 0, \end{aligned}$$

by Lemma 12.

In order to see that this same inequality will hold for the $q < 0.5$ case, first note that $J(c,p,q) \equiv J(c,1-p,1-q)$. In fact,

$$J(c,1-p,1-q) = (2(1-q)-1)\left(2\Pr(Z_{1-p,1-q}>0) + \Pr(Z_{1-p,1-q}=0) - 1\right),$$

where $Z_{1-p,1-q} \sim \text{Skellam}(((1-q)/(1-p))\,m(c)/2, (q/p)\,m(c)/2)$. By the symmetry property of the Skellam distribution, we then have $-Z_{1-p,1-q} \sim \text{Skellam}((q/p)\,m(c)/2, ((1-q)/(1-p))\,m(c)/2)$, the exact same distribution of $Z_{p,q}$. Thus,

$$\begin{aligned} J(c,1-p,1-q) &= (1-2q)\left(2\Pr(-Z_{1-p,1-q}<0) + \Pr(-Z_{1-p,1-q}=0) - 1\right) \\ &= (1-2q)\left(2\Pr(Z_{p,q}<0) + \Pr(Z_{p,q}=0) - 1\right) \\ &= (1-2q)\left(2\left(1-\Pr(Z_{p,q}>0) - \Pr(Z_{p,q}=0)\right) + \Pr(Z_{p,q}=0) - 1\right) \\ &= (1-2q)\left(-2\Pr(Z_{p,q}>0) - \Pr(Z_{p,q}=0) + 1\right) \\ &= (2q-1)\left(2\Pr(Z_{p,q}>0) + \Pr(Z_{p,q}=0) - 1\right) = J(c,p,q). \end{aligned}$$

Having noted this, since $1 - p_\infty^*(c,q) = p_\infty^*(c,1-q)$ by Claim 19, we once more have

$$\begin{aligned} & J(c,p_\infty^*(c,q),q) - J(c,q,q) = J(c,p_\infty^*(c,q),q) = J(c,1-p_\infty^*(c,q),1-q) \\ = \; & J(c,p_\infty^*(c,1-q),1-q) = J(c,p_\infty^*(c,1-q),1-q) - J(c,1-q,1-q) > 0, \end{aligned}$$

since $1-q>0.5$.

Thus, in either case, we have $\lim_{n\to\infty}\left(\left(\mathcal{I}\left(n,c,p_n^*\left(c,q\right),q\right)-\mathcal{I}\left(n,c,q,q\right)\right)/n\right)>0$, so that $\mathcal{I}\left(n,c,p_n^*\left(c,q\right),q\right)>\mathcal{I}\left(n,c,q,q\right)$ for sufficiently large n.

iii) It follows immediately from parts (i) and (ii) that, for sufficiently large n, $\mathcal{W}\left(n,c,p_n^*\left(c,q\right),q\right)>\mathcal{W}\left(n,c,q,q\right)$. ■

Proof of Proposition 8. i) Note from (18) and Lemmas 3 and 1 that, for all $n\geq n_0\left(c\right)$,

$$\frac{\mathcal{C}\left(n,c,q,q\right)}{n}=\frac{n\frac{\gamma^*\left(n,c\right)}{2}c+n\frac{\gamma^*\left(n,c\right)}{2}c}{n}=\gamma^*\left(n,c\right)c.$$

Hence, Lemma 2 yields $\lim_{n\to\infty}\left(\mathcal{C}\left(n,c,q,q\right)/n\right)=0$.

Since $\mathcal{C}\left(n,c,p_n^*\left(c,q\right),q\right)<\mathcal{C}\left(n,c,q,q\right)$ for sufficiently large n (Proposition 7) and $\mathcal{C}\left(n,c,p_n^*\left(c,q\right),q\right)\geq 0$, the Squeeze Theorem gives $\lim_{n\to\infty}\left(\mathcal{C}\left(n,c,p_n^*\left(c,q\right),q\right)/n\right)=0$ too.

ii) It was already shown in the proof of part (ii) of Proposition 7 that, both for the $q>0.5$ and the $q<0.5$ cases, $\lim\limits_{n\to\infty}\left(\mathcal{I}\left(n,c,p_n^*\left(c,q\right),q\right)/n\right)=J\left(c,p_\infty^*\left(c,q\right),q\right)>0=J\left(c,q,q\right)=\lim\limits_{n\to\infty}\left(\mathcal{I}\left(n,c,q,q\right)/n\right)$.

iii) This follows immediately from parts (i) and (ii). ■