Canadian Journal of Political Science (2025), 1–27
doi:10.1017/S0008423924000465

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE/ÉTUDE ORIGINALE

Working the Crowd: Citizen Forecasting, Sophistication and Diversity in Canadian Federal and Provincial Elections

Philippe Mongrain[1], Nadjim Fréchet[2], Brian Thompson Collart[3] and Yannick Dufresne[4]

[1]University of Antwerp, Sint-Jacobstraat 2-4, Antwerp, 2000, Belgium, [2]Université de Montréal, 3150, rue Jean-Brillant, Montréal, H3T 1N8, QC, Canada, [3]Université Laval, 1030, avenue des Sciences-Humaines, Québec, G1V 0A6, QC, Canada and [4]Université Laval, 1030, avenue des Sciences-Humaines, Québec, G1V 0A6, QC, Canada

Corresponding author: Philippe Mongrain; Email: philippe.mongrain@uantwerpen.be

Abstract

According to the "miracle of aggregation" principle, in the absence of systematic biases, errors in individual judgments within a population should cancel each other out and lead to a correct decision at the aggregate level. This article explores potential individual- and group-level correlates of the accuracy of citizens' electoral expectations and investigates how potential markers of political sophistication—namely, educational attainment and political interest—could be used to improve upon the raw aggregation of citizens' forecasts using massive survey datasets collected during six Canadian national and provincial election campaigns between 2011 and 2022 ($n = 279{,}003$). We find that while educational attainment and interest increase the probability of a correct forecast at the individual level, delegating the forecasting task based on these variables does not necessarily lead to improvements in the accuracy of aggregate-level predictions. At the group level, we fail to uncover any evidence that sociological or informational diversity increases forecasting accuracy.

Résumé

Selon le principe du « miracle de l'agrégation », en l'absence de biais systématiques, les erreurs dans les jugements individuels au sein d'une population devraient s'annuler mutuellement et conduire à une décision correcte au niveau agrégé. Cet article considère différents facteurs, tant au niveau individuel qu'au niveau du groupe, qui pourraient avoir une influence sur la précision des attentes électorales des citoyens. Par ailleurs, nous évaluons comment des marqueurs potentiels de sophistication politique—à savoir le niveau d'éducation et l'intérêt politique—pourraient être mobilisés pour améliorer l'agrégation brute des attentes des citoyens. Pour ce faire, nous employons des ensembles de données d'enquêtes massives collectées au cours de six campagnes électorales nationales et provinciales canadiennes entre 2011 et 2022 (n = 279 003). Nous constatons que si le niveau d'éducation et d'intérêt politique augmentent bel et bien la probabilité d'une prévision correcte au niveau individuel, la délégation de la tâche prévisionnelle en fonction de ces variables n'entraîne pas nécessairement une amélioration de la précision des attentes électorales au niveau agrégé. Enfin, au niveau du groupe, aucun élément ne laisse croire que la diversité sociologique ou informationnelle augmente la précision des attentes exprimées par les citoyens.

Keywords: aggregation; expectations; wisdom of crowds

mots-clés: agrégation; attentes; sagesse des foules

Introduction

Citizens' expectations can be a formidable forecasting tool once aggregated. One of the most concrete examples of this phenomenon was provided by British anthropologist and statistician Francis Galton more than a century ago. In the fall of 1906, Galton attended the West of England Fat Stock and Poultry Exhibition in Plymouth. As Galton walked through the fair, he stumbled upon a weight-judging competition. The crowd was asked to guess the dressed weight of an ox that was put on display before them. By paying a small fee, people could enter the contest. They were provided with a card on which they wrote their best estimate of the animal's weight along with their name and address. Prizes were to be given to the most accurate participants. As Galton (1907: 450) later wrote,

> [t]he competitors included butchers and farmers, some of whom were highly expert in judging the weight of cattle; others were probably guided by such information as they might pick up, and by their own fancies. The average competitor was probably as well fitted for making a just estimate of the dressed weight of the ox, as an average voter is of judging the merits of most political issues on which he votes, and the variety among the voters to judge justly was probably much the same in either case.

Once the contest was over, Galton was able to gather the cards to study the participants' estimates. The median guess of the 787 participants was 1,207 lbs. The actual weight of the dressed ox proved to be 1,198 lbs. In other words, the common judgment of the contestants was too high by less than 1 per cent of the ox's total weight. However, when looked upon separately, individual estimates were, in most cases, much less accurate. This came as a surprise to Galton, who had hypothesized that the mix of a few experts and numerous unskilled fair-goers would lead, on average, to a rather poor appraisal of the ox's true weight (for a re-examination of Galton's paper, see Wallis, 2014). Galton's country fair visit serves as the classic illustration of the "wisdom of crowds" (WOC) principle. According to this principle, "under the right circumstances, groups are remarkably intelligent, and are often smarter than the smartest people in them. […] Even if most of the people within a group are not especially well-informed or rational, it can still reach a collectively wise decision" (Surowiecki, 2004: xiii–xiv). In fact, it is even argued that the WOC principle extends to individuals themselves: repeated estimations aggregated across a single person are generally more accurate than single estimations ("inner-crowd wisdom"), although aggregating the answers of different individuals ("outer-crowd wisdom") tend to give better results (Fiechter and Kornell, 2021; van Dolder and van den Assem, 2018). In a classic experiment, akin to the ox weighting competition studied by Galton, economist Jack Treynor had his students guess the number of jelly beans in a jar. Aggregation, once again, proved a successful strategy. Treynor (1987: 50) noted that "[a]pparently it doesn't take knowledge of beans, jars or packing factors for a group of students to make an accurate estimate of the number of beans in a jar. All it takes is independence."

Election campaigns, not unlike sporting events, are carnivals of expectations. Taking a quick look at the history of election forecasting, one will find wagers on papal elections as soon as the beginning of the sixteenth century, massive straw polling operations by general interest magazines at the turn of the twentieth century and even predictions based on "partisan-flavoured" sodas and ice cream sales (see, for example, Erikson and Tedin, 2016; Herbst, 1993; Rhode and Strumpf, 2013). We can add to these more extravagant or commercially oriented attempts at forecasting election outcomes the treasure trove of voter intention surveys that have been conducted since the advent of modern public opinion polling in the 1930s, vote and seat share forecasts by modellers and aggregators, and the countless hours of speculation by pundits and journalists about parties' and candidates' electoral prospects. Expectations are clearly an essential part of election campaigns in competitive democracies. Although their motivations and level of objectivity may differ, pundits, pollsters, researchers, politicians and voters all engage in speculations about who will win and by how much.

We often turn to pollsters and panels of experts to answer these questions. However, *as a group*, citizens have been credited with being just as or even more accurate than traditional forecasting methods (Gaissmaier and Marewski, 2023; Graefe, 2016; Murr, 2017). The main objective of the present article is to explore both individual- and group-level correlates of citizens' forecasting accuracy in order to draw conclusions on how these potential predictors could be exploited to improve forecasts based on aggregated citizens' expectations. This objective is in line with Graefe's (2016: 227) statement that "[f]uture research should focus on developing methods to identify the most accurate forecasters in a sample." To achieve this goal, we use massive datasets from the Ipsos Canada Election Surveys (n = 134,236), the Local Parliament Project (n = 37,380), the 2019 Canadian Election Study (n = 41,843) and the online voting-prediction tool Datagotchi (n = 65,544).

The Wisdom of Crowds

As explained by de Oliveira and Nisbett (2017: 2066, italics added), crowd wisdom "is typically observed when estimates are independent and randomly chosen to be aggregated by some method like averaging. This often allows people's errors to cancel out during the averaging process, as *each person's guess is comprised of truth plus some positive or negative error*." Three models are often invoked to explain the WOC principle (see Brennan, 2021: 376–77; Špecián, 2022: 56–63), namely, (1) the Miracle of Aggregation, (2) Condorcet's Jury Theorem and (3) Hong and Page's (2004) Diversity Trumps Ability Theorem (see also Page, 2007). All of these models have received a fair amount of criticism. Nevertheless, there is

https://doi.org/10.1017/S0008423924000465 Published online by Cambridge University Press

ample empirical evidence that groups of citizens predict election outcomes better than individual citizens taken separately (Murr, 2017).

Page and Shapiro (1992, 1999) have claimed that skeptical views about the knowledge and the reasoning capacities of the public are not well-founded despite decades of research underlining the political ignorance and inattentiveness of individual citizens. Instead, they show that public opinion on policies is rational *as a whole* partly because randomly-distributed errors in individual judgments tend to cancel out in the aggregate. Therefore, citizens can collectively formulate reasonable opinions without most individuals possessing a vast knowledge of political matters. This has often been referred to as the "miracle of aggregation." Many scholars have criticized this so-called "miracle." According to Page and Shapiro's critics, simple statistical aggregation only offers a partial remedy to the somewhat low levels of political knowledge among the public (Althaus, 1998; Bartels, 1996; Brennan, 2021; Caplan, 2007, 2009; Gilens, 2019; Kuklinski and Quirk, 2000).

The miracle of aggregation is one manifestation of the WOC. The term "wisdom of crowds" was popularized by American journalist James Surowiecki (2004) and has been used by multiple authors to explain the accuracy of citizens' election forecasts (Miller et al., 2012; Murr, 2011). Although it has older roots (see Landemore, 2012: 1), the WOC principle is mostly derived from Condorcet's (1785) jury theorem, which states that the probability of a group coming to a correct decision tends toward unity as the group increases in size. For this theorem to be true, four conditions must be met: (1) the group has to make a choice between two alternatives (one correct and one incorrect) according to a majority rule, (2) each individual has to make his or her decision independently of others (see Lorenz et al., 2011), (3) the probability of voting for the correct alternative has to be the same for every member of the jury (uniformity in competence) and (4) this probability must be above 50 per cent. Condorcet's original conditions have since been relaxed by many authors. The theorem holds even when all members do not have the same probability of making the right decision and under certain forms of correlated voting (see, for example, Becker et al., 2017; Boland, 1989; Grofman et al., 1983; Ladha, 1992). It is also possible to extend Condorcet's theorem to situations where there are more than two options (List and Goodin, 2001). In fact, individuals within a group need *not* even predict better than chance on average for the group to beat the average citizen, as this can be achieved by weighting individuals' judgment on the basis of their competence (Shapley and Grofman, 1984).

More substantively, Larrick et al. (2012) have established two conditions for crowds to be wise, that is, (1) individuals within the group need some minimal knowledge or expertise about the issue at hand and (2) they need to hold diverse perspectives—an idea at the heart of the "diversity trumps ability theorem" (DTA) (Hong and Page, 2001, 2004). This theorem states that "groups of ordinary individuals that are inclusive, and thus cognitively diverse, will outperform narrower groups of individuals that have superior ability" (Quirk, 2014: 129). The DTA theorem rests on four main conditions: (1) the problem at hand must be sufficiently difficult, (2) all problem solvers need to have some degree of ability in solving the problem, (3) the group of problem solvers must be diverse and (4) the group of problem solvers has to be reasonably big and must be drawn from a large enough population (Page, 2007: 158–65). Landemore (2012, 261) has even proposed to generalize Hong and Page's DTA theorem into a "numbers trump ability theorem" since diversity should be a natural consequence of increasing group sizes (see Quirk, 2014, however), although one might expect diminishing returns from each additional individual over a specific threshold (Jacobson et al., 2011).

Cognition and Affect

In one of the first empirical analyses of the factors influencing predictive judgment, McGregor (1938: 182) stated that "[a]n individual's pre-existent attitudes, wishes, and knowledge concerning a given social situation provide a frame of reference that will influence the formation of the premises upon which his predictions concerning events related to that situation will be based." We find, in this frame of reference, two of the main ingredients of the expectation-formation process: (1) predispositions or preferences and (2) information. In other words, voters' expectations about electoral outcomes are simultaneously influenced by affect (mostly partisan biases) and cognition (information effects) (Dolan and Holbrook, 2001).

Partisan Biases

It has long been recognized that preferences exert a major influence on expectations (Rehm and Gadenne, 2013: 91–92). Apart maybe from purely cognitive limitations, motivated reasoning is probably the single most important threat to forecasting accuracy. Without fail, research on voters' expectations shows that partisan preferences are strongly correlated with their expectations of election winners (Mongrain, 2021a). For example, Hayes (1936) observed that the majority of people who intended to vote for incumbent president and Republican candidate Herbert Hoover in the US presidential election of 1932 also expected him to win; just as most Democratic supporters were keen on predicting a victory for Franklin D. Roosevelt, who ultimately won in a landslide. Hayes (1936: 186) also noted, among other things, that Socialist voters "were presumably in the best position to guess the winner of the race dispassionately" because they could hardly expect anything from their candidate. Since marginal parties have no *realistic* chance of gaining power, their reasoning should be less driven by wishful thinking. Relatedly, individuals who describe themselves as independents should be less susceptible to motivational biases. Furthermore, in observing that upper-class Republicans were less prone to forecast a Hoover victory than their lower-class co-partisans, Hayes (1936: 187) turned to the "greater penetration of the upper groups by the more reliable indicators of the real outcome" as a potential explanation. This is in line with studies arguing that greater information inputs or knowledge can moderate the effect of wishful thinking on expectations. Another potential mechanism explaining the high correlation between voter intention or partisanship and expectations is social homophily. Because individuals tend to spend time with people sharing political views similar to their own, pre-existing biases are often reinforced by social contacts (for a review, see Mongrain, 2023). A certain degree of diversity and disagreement is thus deemed desirable as "partisan bubbles" tend to filter out disagreeable and politically uncongenial information (Leiter et al., 2020; Uhlaner and Grofman, 1986).

https://doi.org/10.1017/S0008423924000465 Published online by Cambridge University Press

Taking into consideration motivational biases is essential to the study of voters' expectations: failing to adequately control for individual preferences will necessarily lead to biased estimates. This is particularly true in the case of election forecasting: in essence, elections are contests of competing values and ideologies. Therefore, objectivity is hard to achieve because individuals are rarely free of "ego-involvement." As mentioned by Cantril (1938: 389, footnote 22), on a host of social events, "the average judgment of a group of individuals […] cannot be compared qualitatively with the average judgment of a group on the length of a line or the number of beans in a jar." For this reason, Treynor's (1987) statement according to which a group does not need "knowledge of beans" to make a decent guess when eyeballing a candy jar might not apply to politics; poorly informed voters are presumably more likely to rely on affect (that is, their partisan preferences) to guide their judgment. The vast majority of studies that followed the pioneering work of the 1930s confirmed the close association of partisan preferences and expectations (see, for example, Dolan and Holbrook, 2001; Krizan et al., 2010; Mongrain, 2021a).

Information Effects

Intuitively, one might expect expertise or knowledge to have much to do with forecasting skills. The evidence, however, is rather mixed. In their pioneering work on citizen forecasting, Lewis-Beck and Skalaban (1989) and Lewis-Beck and Tien (1999) concluded that education and contextual factors, such as closeness to the election and (perceived) tightness of the race, were more important than political interest or involvement in predicting voters' accuracy. Lewis-Beck and Tien (1999: 180) argued that "[m]ore educated people are better able to understand the political world," but also that "[t]heir more extensive social networks link them naturally to more information." Dolan and Holbrook (2001) found that political knowledge was significantly related to accurate election predictions, and that greater sophistication reduced the influence of wishful thinking on citizens' expectations. Miller et al. (2012) concluded that self-ratings of political and election knowledge increased forecasting accuracy. These results support the intuitive expectation established above: knowledge does appear to be positively associated with forecasting ability.[1]

More recently, Morisi and Leeper (2024) have investigated the impact of individual and exogenous informational characteristics on citizen forecasting accuracy during the 2016 Brexit referendum. Sophistication, which was measured using respondents' level of education and political attentiveness, was directly associated with greater accuracy in predicting the result of the Brexit referendum and indirectly by reducing the influence of partisan biases. The effects noted by Morisi and Leeper (2024) were, however, relatively small. Additionally, the level of attention paid to politics seemed much more important than education as a predictor of forecasting accuracy.

A few works have also suggested that the size and nature of an individual's social network could play a role in providing politically-relevant information and, thereby, contribute to forecasting accuracy. Larger personal networks, frequent political discussion with family, friends and colleagues, ideological or partisan heterogeneity as well as politically knowledgeable contacts were deemed potentially beneficial to the quality of individuals' prospective judgment about election outcomes (Leiter et al., 2018, 2020; but see Mongrain, 2023).

Improving Forecasts

As already mentioned, the average judgment within a group is usually closer to the truth than that of a randomly-chosen individual. This raises one important question: can we improve the outcome of statistical aggregation by putting a premium on competence or sophistication? Research on decision-making and prospective judgment has not always been kind to experts. The usefulness of expertise was already questioned more than eighty years ago by McGregor (1938) who concluded that professors were no more proficient than their students at predicting social events. The research conducted by Tetlock (2017) on expert opinion is often summarized with the author's observation that the average expert barely did better than a "dart-throwing chimp" at making accurate predictions in a variety of domains. In fact, according to Hammond (1996: 278), "in nearly every study of experts carried out within the judgment and decision-making approach, experience has been shown to be unrelated to the empirical accuracy of expert judgments" (but see Jacobson et al., 2011).

Nevertheless, there is empirical evidence pointing in the opposite direction. Murr (2015) has shown that delegating and weighting forecasts according to respondents' level of competence improved citizens' prediction of US presidential outcomes. Competence was measured by identifying characteristics of accurate forecasters in past elections and calculating the predicted probability of a correct forecast in the current election (thus giving more weight to respondents sharing these characteristics). Delegation then works by eliminating individuals below a certain level of competence and keeping only those above that same threshold (see Kazmann, 1973). It is, in essence, an "epistocratic" approach to decision-making as it restricts the forecasting task to the most competent members of the group (a "select crowd"; see Budescu and Chen, 2015).

In another study, Mongrain (2021b) distinguishes between two views of the WOC principle in citizen forecasting. He refers to these as the "democratic view" and the "technocratic view." The "technocratic" approach is somewhere halfway between a democratic rule of full inclusion (equality) and an epistocratic rule of competence-based discrimination (quality). Using district-level data from multiple elections in Canada, France, Germany and Great Britain, the author created two indices, one based on respondents' factual knowledge of politics and one based on respondents' own past forecasting performance (from panel survey data). Weighting by these indices produced modest, but noticeable, increases in the number of correctly predicted district races. One of the major limitations of Mongrain's (2021b) study, however, was the very small number of respondents in each district (notwithstanding the complete lack of data for many districts).

Given the above discussion of citizens' forecasts and crowd wisdom, this article puts forward the following two hypotheses:

H1: Politically sophisticated voters will be more likely to correctly forecast the outcome of an election in their district.

H2: At the aggregate level, socially and cognitively diverse groups will be more likely to correctly forecast the outcome of an election in their district.

https://doi.org/10.1017/S0008423924000465 Published online by Cambridge University Press

Data and Methods

To test these hypotheses, we use data from nine surveys conducted during Canadian national and provincial election campaigns between 2011 and 2022. More precisely, the analyses rely on data collected from various Ipsos Canada Election Surveys, the Local Parliament Project (LPP), the 2019 Canadian Election Study (CES) and Datagotchi, a gamified knowledge-transfer and data collection app. Ipsos surveys for the 2011 and 2015 federal elections as well as the 2011 and 2014 Ontario general elections were conducted in the last few days of the campaign and/or on election day (exit polls). To measure expectations about district-level outcomes, respondents were asked the following question: "If you had to bet $1000.00 of your own money, which party's candidate do you think will win in your riding during this election?"[2] The 2011 federal election is a particularly interesting case for the study of citizen forecasting. The Liberal Party dropped in third place for the first time in the country's history, while the New Democratic Party (NDP) had its best performance since its creation by winning a total of 103 seats out of 308, including 59 of Quebec's 75 seats (this was later referred to as the "Orange Wave"). The NDP formed the Official Opposition for the first time, in large part owing to its success in Quebec, which "ha[d] always been exceedingly difficult terrain for the CCF-NDP" in past elections (Whitehorn, 1997: 105). The NDP's sudden surge in voter intentions in the last few days of the campaign provides a perfect test of citizens' reactivity to changing campaign dynamics. The 2019 federal election is another interesting case: the incumbent Liberal Party received slightly less votes than the Conservative Party (33.1 vs 34.3%) but nonetheless won more seats (157 vs 121). District-level forecasts, which provide seats rather than national vote share estimates, might be particularly useful in that kind of situation.

The 2015 CES of the LPP also questioned its respondents about their electoral expectations. More precisely, LPP respondents were asked to rate the likelihood of winning, on a 0–100 scale, for each party in their local district ("Thinking now about where you live, how likely is each party to win your constituency?"). Data from the 2015 Ipsos survey and the 2015 LPP survey were combined following harmonization. More precisely, winning probabilities from LPP respondents were used to identify the most likely winner. Forecasts were coded 1 when the party with the highest likelihood of winning given by a respondent matched the actual winner and 0 otherwise.

Data from the 2019 CES were collected through a phone survey as well as a web survey. The phone survey included the following question: "In your own local riding, which party has the best chance of winning?" Respondents were not given a predefined list of possible outcomes. Those who named more than one party were invited to provide their best guess. Internet respondents were asked to rate the likelihood (on a 0–100 scale) of each party winning in their riding. As for the LPP, respondents' probability estimates for each party were used to create a binary indicator of forecasting accuracy (0 = incorrect, 1 = correct) in order to merge answers from the phone and web surveys.[3]

Finally, the Datagotchi data were collected during the 2022 Quebec general election. The following question was used to measure respondents' expectations: "In your opinion, which party has the best odds of winning in your riding?" The previous election in 2018 was the first election since 1970 to be won by a party other than the federalist Liberal Party of Quebec (PLQ) or the sovereigntist Parti Québécois (PQ). The Coalition Avenir Québec (CAQ), a nationalist center-right party, formed a majority government in 2018 by winning 74 seats (out of 125) in the Quebec National Assembly with 37.4 per cent of the popular vote. In 2022, the CAQ increased its majority by gaining a total of 90 seats with 41 per cent of the vote. During the entire duration of the campaign, between 28 August and 2 October, the CAQ remained high above the other parties in terms of voter intentions (at around 40% compared to less than 20% for its closest competitor). Although, the CAQ's victory was hardly surprising, the outcomes of local (riding) races were much less certain. Section C of the online appendix displays voter intention data for each election.

Table 1 provides a brief overview of each election's outcome as well as the percentage of respondents who made a correct district-level forecast in these elections. The table also displays the percentage of districts falling in different ranges of sample sizes and the average number of respondents per district in each election. As can be seen, with the exception of the 2015 Canadian federal election, a clear majority of voters correctly predicted the outcome in their own riding. Therefore, in most cases, the average voter did significantly better than would have a simple coin toss (especially if we consider the fact that more than two candidates might be competitive in many districts—see, for example, Gaines, 1999; Johnston and Cutler, 2009).

The chosen surveys are characterized by very large samples totalling thousands and, in most cases, tens of thousands of observations with respondents in every (or almost every) district. Therefore, these are ideal datasets to study the benefits of aggregation as we have relatively large subsamples in each cluster (district). On the question of sample size, one could reasonably ask how large a group needs to be in order to reap the benefits of collective wisdom and aggregation. Research on experts' estimates and forecasts tend to show that beyond a relatively small number of inputs, improvement in collective accuracy rapidly declines. The exact point of diminishing returns varies, but it is often found between five and twelve individuals only (see, for example, Hemming et al., 2018; Hogarth, 1978; Hora, 2004). In a study of experts' predictions regarding various geopolitical questions, Satopää et al. (2014: 353) established that the optimal group size was around

Table 1. Overview of Election Outcomes and Citizens' Forecasts

		Results			Sample Size				
Election	% Correct	Winner	% Vote	% Seats	% < 25	% 25-49	% 50-99	% ≥ 100	Avg n
Canada 2011	57.66	CPC	39.62	53.90	0.97	0.32	3.90	94.81	280.76
Canada 2015	50.34	LPC	39.47	54.44	0.90	4.18	28.96	65.97	114.42
Canada 2019	58.74	LPC	33.12	46.45	1.48	9.47	74.85	14.20	76.31
Ontario 2011	60.24	OLP	37.65	49.53	0.00	0.00	2.83	97.17	227.68
Ontario 2014	63.71	OLP	38.67	54.21	0.94	9.43	75.47	14.15	77.10
Quebec 2022	73.60	CAQ	40.98	72.00	0.00	1.60	3.20	95.20	385.78

Notes. Election results retrieved from Elections Canada, Elections Ontario, and Elections Quebec. CAQ = Coalition Avenir Quebec. CPC = Conservative Party of Canada. LPC = Liberal Party of Canada. OLP = Ontario Liberal Party.

https://doi.org/10.1017/S0008423924000465 Published online by Cambridge University Press

fifty. In his study of citizens' constituency-level forecasts in the 2010 British general election, Murr (2011: 782) concluded "that in most cases about 20 respondents suffice to have a much greater than 50 per cent chance of getting it right." Overall, to the extent that other requirements are met, the WOC does not appear to require particularly large crowds to function properly. In order to assess the optimal group size for district-level forecasts, random samples of sizes varying from 1 to 30 (that is, 1, 2, 3… 30) were drawn from each district with at least sixty respondents. In other words, we executed multiple random draws of successively larger numbers within each district. For each sample size, the sampling procedure was repeated ten times with replacement. The aggregated forecasts from the ten trials were averaged to get the percentage of correctly predicted seats at each sample size. The results are shown in Figure 1. We see rather clear improvements in collective accuracy as within-district samples increase in size. It seems, however, that the rate of improvement considerably weakens beyond approximately ten to fifteen respondents. The trend across elections shown in Figure 1 strengthens the argument that we have more than enough observations per district (see Table 1) to observe and test WOC effects.

To test the two hypotheses established above, we proceed in two steps. First, we estimate multilevel random effects logistic regression models for individual-level forecasts

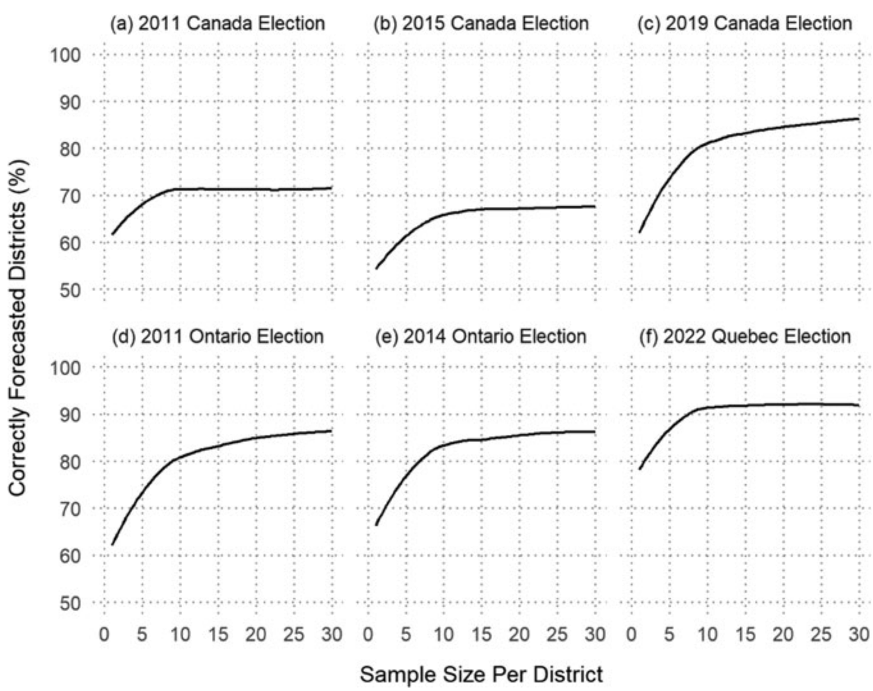

Figure 1. Percentage of Correctly Predicted Districts at Varying Sample Sizes.

Notes. The aggregation of respondents' expectations is based on data from 304 (out of 308) districts in the 2011 Canadian federal election, 310 (out of 338) in the 2015 Canadian federal election, 253 (out of 338) in the 2019 Canadian federal election, 106 (out of 107) in the 2011 Ontario general election, 81 (out of 107) in the 2014 Ontario general election, and 123 (out of 125) in the 2022 Quebec general election.

https://doi.org/10.1017/S0008423924000465 Published online by Cambridge University Press

of district (riding) election outcomes. The dependent variable—voters' riding-level forecasts—is a binary indicator coded 1 for correct forecasts and 0 otherwise. We use educational attainment and political interest as proxies for political sophistication (see Morisi and Leeper, 2024; Turper and Aarts, 2017). Education has been found to cultivate political interest and information seeking (Grönlund and Milner, 2006; Le and Nguyen, 2021) and there is ample evidence that well-educated citizens are less susceptible to wishful thinking and generally better at anticipating election outcomes (Dolan and Holbrook, 2001; Meffert et al., 2011; Morisi and Leeper, 2024). However, according to Elo and Rapeli (2010), interest is the most suitable proxy for political knowledge when compared to other measures such as self-assessed knowledge or the accuracy of party placements on a left-right scale. Luskin (1990: 351) concluded that political "[s]ophistication depends, above all, on motivation (interest, occupation and, indirectly, parental interest). It also depends on ability (intelligence). But the big informational variables (education and [media usage]) have little effect." Respondents with a university degree were coded 1 and all other respondents were coded 0. Interest with politics or the election was measured on a 0–10 scale (which was rescaled from 0 to 1). Note that respondents' interest was only recorded in the 2015 LPP and 2019 CES. At the individual level, every model includes basic sociodemographic controls (sex, age and household income) and vote choice, which serves as a proxy for partisanship since respondents were not questioned about their party identification (PID) in most surveys. For the 2015 LPP and 2019 CES, we were able to use partisan identification instead of vote choice in the additional models with both education and political interest as measures of sophistication.

Consequently, there are two sets of regressions. In the first set, all models include vote choice (1 = intend to vote for winner, 0 = otherwise) as a proxy for partisan preference and education as a measure of sophistication. In the second set, models include PID (1 = identify with one of the losing parties, 2 = no PID, 3 = identify with the winning party) instead of vote choice[4] and both education and political interest as measures of sophistication. The first set of models also include an interaction term between vote choice and education, the expectation being that motivational biases will be weaker among highly educated voters (more concretely, the gap in the probability of making a correct forecast should be smaller between highly educated losers and winners than between losers and winners with a lower level of education). Following a similar logic, the second set of models include an interaction term between PID and political interest. All models have random slopes for the interaction terms (since we can assume election outcomes are more evident in certain districts than in others, education or interest might play a lesser/greater role in enhancing or decreasing partisan biases).

In addition to individual-level characteristics, in both sets of regressions, models include one measure of competitiveness and one measure of change as district-level variables, namely the (standardized) margin of victory (the difference in vote share between the local winner and the second-place candidate) and a dummy variable for reelection (1 if the incumbent party candidate was reelected and 0 otherwise). Since federal electoral districts were reviewed in 2012, the incumbent party reelection variable was replaced by a variable denoting boundary changes for the 2015 Canadian election. As mentioned by Murr (2011: 778, italics in original), "*boundary changes* may greatly change the size and composition of a constituency rendering past

https://doi.org/10.1017/S0008423924000465 Published online by Cambridge University Press

election results useless for predictions." As voters are nested within groups (districts), it is essential to take into account the level of competitiveness in each riding as well as other potential unobserved group-level factors. Finally, response date (the standardized number of days between the interview and election day) was also accounted for in models for the 2015 and 2019 federal elections as well as the 2022 Quebec general election as respondents were interviewed over a period of several weeks. The closer we get to election day, the easier it should be to make a correct forecast.[5]

In the second step, we estimate logistic regression models using districts (groups) instead of individual respondents as the unit of analysis. Following Murr (2011), these models look at the influence of informational (cognitive) and sociological diversity on the accuracy of aggregated forecasts using entropy-based diversity indices. Informational diversity is captured through respondents' education, level of political interest, vote choice (since respondents with different political orientations can be attentive or exposed to different information sources) and response date (as more information, and presumably more accurate information, becomes available as election day gets closer), while sociological diversity is captured through respondents' sex, age and income level.[6] The diversity measures (D) were computed as shown in Equation 1:

$$D_j = \ -\sum_{i=1}^{n} P_{ij}(lnP_{ij}) \tag{1}$$

where P_i is the proportion of respondents within district j who possess the i^{th} diversity characteristic and n is the number of characteristics considered (for example, $n = 6$ if there are six age categories).

Therefore, the diversity index is the negative sum of the products of each characteristic's proportion in a district and the natural log of its proportion. Higher values of the index indicate greater diversity. In order to properly capture diversity, we kept each variable's original scale when appropriate. For example, education was not dichotomized on the basis of university education; to obtain a fine-grained measure of educational diversity, we recorded the proportion of individuals in each category (for example, less than high school, high school diploma, some college and so forth). These models also include group size (the number of respondents within each district), (standardized) margin of victory and incumbent party reelection (or boundary changes) as covariates. Following Murr (2015), group size was logged to normalize its distribution. Group diversity was measured at the district level since it was the smallest geographical unit associated to respondents in all cases, with the exception of the 2022 Quebec general election. In the 2022 Datagotchi survey, the first three digits of respondents' postal code (that is, the Forward Sortation Area [FSA]) were also available. Therefore, we measured diversity both at the district and the FSA levels among Datagotchi respondents.

Results

Figure 2 shows the percentage of correct forecasts at the individual and group levels for each election. Group-level (or district-level) forecasts correspond to the aggregation of individual forecasts inside every district. Consider the 2022 Quebec general election. Whereas about 73.6 per cent of citizens correctly predicted which party

https://doi.org/10.1017/S0008423924000465 Published online by Cambridge University Press

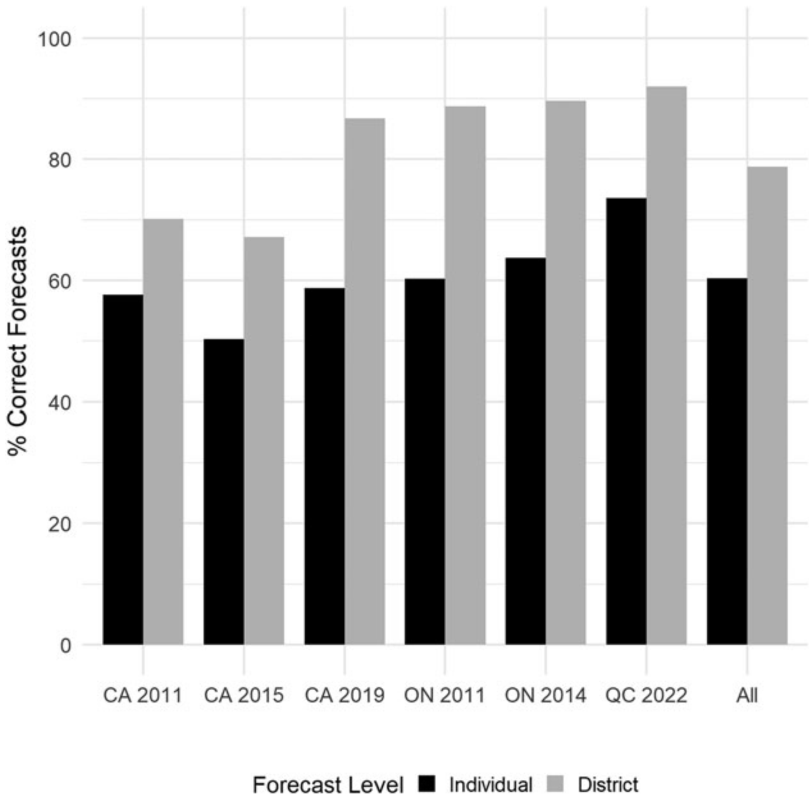

Figure 2. Accuracy of Individual- and District-Level Forecasts.

would win in their local riding, about 92 per cent of groups did—an increase of 18.4 percentage points. Across all elections, around 60.4 per cent of respondents were able to correctly identify the winning candidate in their district. The success rate for group-level forecasts in all five elections was 78.8 per cent, an increase of over 18 percentage points.

One could argue that the patterns observed in Figure 1 and Figure 2 are driven by the greater number of respondents on the winning side (we do observe that respondents who intend to vote for the winning candidate represent a plurality of voters in most districts). As suggested by previous research, there is a clear association between voters' preferences and electoral expectations. Individuals supporting the winning party or candidate tend to display greater accuracy simply because they "benefit" from their biases (just as losers' biases tend to act against them). Therefore, we reproduced Figure 1 and Figure 2 for losers and winners (according to their voter intention) separately. These additional analyses are available in section D of the appendix. We find the same patterns in both groups (that is, larger groups tend to provide more accurate forecasts and aggregation beats the average voter), although the accuracy of individual- and district-level forecasts are considerably higher among winners. As such, the partisan biases of winners appear to

https://doi.org/10.1017/S0008423924000465 Published online by Cambridge University Press

partially "compensate" for the partisan biases of losers. When it comes to citizens' electoral forecasts, one might wonder if there is wisdom in the crowd or if there is mostly bias (wishful thinking) in the crowd. It does seem like a significant portion of the "wisdom" in the crowd stems from winners' partisan biases.

The number of predicted seats for each party in each election from aggregated citizens' forecasts as well as the actual outcomes can be found in Table 2. As can be seen, the mean absolute error (MAE) ranges from a low of 0.6 (Ontario 2011) to a high of 13.1 (Canada 2011) percentage points. We also computed the symmetric percentage error (SPE) and log accuracy ratio, or log error (LE), to give a better sense of the relative magnitude of errors across elections (see Tofallis, 2015). SPE and LE values close to 0 indicate more accurate forecasts (see section E of the appendix for details): for example, the mean absolute SPE (sMAPE), which as an upper-limit of 100, ranges from a low of 1.5 (Ontario 2011) to a high of 41.7 (Canada 2011). In the three provincial elections, voters as a group correctly ranked each party. In the 2019 federal election, voters correctly predicted the overall outcome, although they overestimated the Conservatives' seat share and underestimated that of the Bloc Québécois. Despite the fact that the Liberal Party lost the popular vote, aggregated citizens' expectations correctly gave the Liberals a plurality of seats. In the 2011 and 2015 federal elections, voters proved collectively unable to anticipate the outcome. Although aggregated expectations correctly predicted a victory of the Conservative Party in 2011, they pointed toward a minority government (less than 155 seats). Furthermore, citizens' expectations for the 2011 federal election did not hint at the possibility of an "Orange Wave" for the NDP. In 2015, not only did voters collectively fail to foresee the victory of the Liberal Party, but they considerably overestimated the seat share of the NDP. If citizens' performance across the six elections can be deemed respectable overall, there is clearly room for improvement.

Table 3 displays the results of the two sets of regression models described above. Starting with the first set of models, we find, as expected, that respondents who voted for the winning candidate in their district were more likely to make a correct forecast than those who voted for one of the losing candidates. On average, and all else being equal, a vote for the winning candidate leads to an increase in the probability of a correct forecast at the district level of about 20 percentage points by a minimum in the 2022 Quebec general election and of 51 percentage points by a maximum in the 2015 Canadian federal elections.[7] This is not surprising in light of the vast literature on motivational biases. Tests of first and second differences show that the interaction term between vote choice and education is statistically significant in all cases, with the exception of the 2015 federal election. In other words, the gap between losers and winners in the predicted probability of making a correct forecast is smaller among respondents with a university degree than it is among others. While education does not seem to matter for winners, it makes a small, but noticeable difference among supporters of losing parties (between 7–16 percentage points depending on the election). The results in Table 3 also highlight the importance of task difficulty. Larger margins of victory and incumbent reelection (no change) are associated with a higher likelihood of making a correct forecast. For example, a one standard deviation increase in the margin of victory increases the odds by 7–15 percentage points on average

https://doi.org/10.1017/S0008423924000465 Published online by Cambridge University Press

Table 2. Predicted Number of Seats from Aggregated Citizens' Forecasts

	Forecast		Actual		Error			
Election	n	%	n	%	n	p.p.	SPE	LE
Canada 2011 [(a)]								
Bloc Québécois	46	14.94	4	1.30	42	13.64	84.00	2.44
Conservative Party	145	47.08	166	53.90	−21	−6.82	6.75	−0.14
Liberal Party	74	24.03	34	11.04	40	12.99	37.04	0.78
New Democratic Party	45	14.61	103	33.44	−58	−18.83	39.19	−0.83
MAE					40.25	13.07		
sMAPE							41.74	
MALE								1.05
Canada 2015 [(a)]								
Bloc Québécois	0	0	10	2.99	−10	−2.99	100.00	Und
Conservative Party	135	40.30	99	29.55	36	10.75	15.38	0.31
Green Party	1	0.30	1	0.30	0	0	0.00	0.00
Liberal Party	99	29.55	181[(b)]	54.03	−82	−24.48	29.29	−0.60
New Democratic Party	103	30.75	44	13.13	59	17.62	40.14	0.85
MAE					37.40	11.17		
sMAPE							36.96	
MALE								0.44
Canada 2019 [(a)]								
Bloc Québécois	24	7.1	32	9.47	−8	−2.37	14.29	−0.29
Conservative Party	134	39.64	121	35.80	13	3.84	5.10	0.10
Green Party	3	0.89	3	0.89	0	0.00	0.00	0.00
Liberal Party	156	46.15	157	46.45	−1	−0.30	0.32	−0.01
New Democratic Party	27	7.99	39	11.54	−12	−3.55	18.18	−0.37
People's Party	1	0.30	0	0.00	1	0.30	100.00	Und
MAE					5.83	1.73		
sMAPE							22.98	
MALE								0.15
Ontario 2011								
Conservative Party	38	35.85	37	34.91	1	0.94	1.33	0.03
Liberal Party	53	50.00	53	50.00	0	0.00	0.00	0.00
New Democratic Party	15	14.15	16[(b)]	15.09	−1	−0.94	3.23	−0.06

(*Continued*)

Table 2. (*Continued.*)

Election	Forecast		Actual		Error			
	n	%	n	%	n	p.p.	SPE	LE
MAE					0.67	0.63		
sMAPE							1.52	
MALE								−0.01
Ontario 2014								
Conservative Party	36	33.96	28	26.42	8	7.54	12.50	0.25
Liberal Party	48	45.28	58	54.72	−10	−9.44	9.43	−0.19
New Democratic Party	22	20.75	20[(b)]	18.87	2	1.88	4.76	0.10
MAE					6.67	6.29		
sMAPE							8.90	
MALE								0.18
Quebec 2022								
Coalition Avenir Québec	86	68.8	90	72.00	−4	−3.2	2.27	−0.05
Parti Québécois	2	1.60	3	2.40	−1	−0.80	20.00	−0.41
Parti Libéral	22	17.60	21	16.80	1	0.80	2.33	0.05
Québec Solidaire	13	10.40	11	8.80	2	1.60	8.33	0.17
Parti Conservateur	2	1.60	0	0.00	2	1.60	100.00	Und
MAE					2	1.60		
sMAPE							26.59	
MALE								0.17

Notes. (a) In the 2011, 2015, and 2019 Canadian federal elections, a tie was predicted in certain districts; between the Bloc Québécois and the NDP in Abitibi—Baie-James—Nunavik—Eeyou (32.47 percent each) and between the Conservative Party and the NDP in Nunavut (50 percent each) in 2011; between the Conservative Party and the Liberal Party in Eglinton—Lawrence (41.59% each) and Scarborough Centre (39.25 percent each), and between the Liberal Party and the NDP in Winnipeg Centre (39.13% each) in 2015; between the Bloc Québécois and Conservative Party in Abitibi—Baie-James—Nunavik—Eeyou (23.81% each), between the Conservative Party and the Liberal Party in Fleetwood—Port Kells (32.08% each) and Winnipeg South (40% each), between the Conservative Party, the Green Party, and the People's Party in Nunavut (25% each), between the Bloc Québécois and the NDP in Rimouski-Neigette—Témiscouata—Les Basques (31.91% each), and between the Liberal Party and the NDP in Surrey Centre (44.23% each) in 2019. Therefore, the total number of predicted seats is 310 (instead of 308) in 2011 and 345 (instead of 338) in 2019. There are 338 forecasted seats in 2015 as the missing forecasts for the northern territories (i.e., Northwest Territories, Nunavut, and Yukon) are "compensated" by the three previously mentioned tied district-level races. (b) Because there were no observations for Northwest Territories, Nunavut, and Yukon in the 2015 Canadian federal election, the score of the Liberal Party (which won the seats in those ridings) was adjusted accordingly by removing three seats. Because there were no observations for Timiskaming—Cochrane in both the 2011 and 2014 Ontario general elections, the score of the Ontario New Democratic Party (which won the local seat in both cases) was adjusted accordingly by removing one seat in each case. p.p. = percentage points. MAE = mean absolute error. SPE = symmetric percentage error (maximum = 200; values close to 0 indicate accurate forecasts). sMAPE = symmetric mean absolute percentage error. LE = log accuracy ratio or the natural logarithm of the quotient of the forecasted value and the actual value (values close to 0 indicate accurate forecasts). MALE = mean absolute log error. Und = undefined (i.e., natural logarithm of 0).

https://doi.org/10.1017/S0008423924000465 Published online by Cambridge University Press

Table 3. Predictors of Forecasting Accuracy in District-Level Elections at the Individual Level

	I. Vote & Education						II. PID & Interest	
	CA 11	CA 15	CA 19	ON 11	ON 14	QC 22	CA 15	CA 19
Partisan preference								
Voted for winner	1.93***	2.30***	2.17***	1.81***	2.23***	1.77***		
Party ID								
Loser PID (R)								
No PID							0.87	0.07
Winner PID							3.44***	1.32***
Sophistication								
University degree	0.33***	0.31*	0.44***	0.40***	0.34*	0.72***	0.29***	0.30***
High interest							1.23*	0.24
Sociodemographics								
Male	0.27***	0.15**	0.10**	0.17***	0.08	0.11***	−0.35***	0.09**
55 years and over	0.09***	0.12**	0.42***	0.21***	−0.12	0.24***	0.26**	0.42***
High income	0.10***	0.18***	0.14***	0.17***	0.28***	0.01	0.34***	0.14***
Interaction								
Vote × University	−0.30***	0.09	−0.50***	−0.52**	−0.75**	−0.58***		
No PID × Interest							0.15	1.10***
Winner PID × Interest							−0.44	0.70**
Task difficulty								
Margin of victory (*z*)	0.38***	0.62***	0.60***	0.50***	0.36***	0.51***	0.75***	0.61***
Reelected	2.13***		0.69***	0.69***	1.30***	0.63***		0.61***
Boundary changes		−0.09					−0.24	
Response date (*z*)		−0.21***	−0.08***			−0.04**	−0.16***	−0.07***
Intercept	−2.21***	−0.98***	−1.27***	−1.17***	−1.32***	−0.26*	−1.90***	−1.32***
Random effects								
Intercept	0.41	1.44	0.41	0.32	0.92	0.20	16.43	1.54
Vote choice	0.43	2.14	0.57	0.70	2.57	0.65		
No PID							54.41	2.83
Winner PID							40.48	2.64
University degree	0.40	3.39	0.11	0.34	1.17	0.42		
High interest							31.95	2.51
Vote × University	0.66	6.65	0.60	1.75	3.66	0.87		

(*Continued*)

Table 3. (*Continued.*)

	I. Vote & Education						II. PID & Interest	
	CA 11	CA 15	CA 19	ON 11	ON 14	QC 22	CA 15	CA 19
No PID × Interest							115.79	5.10
Winner PID × Interest							80.56	4.27
Sample size								
Observations	86,264	27,084	21,711	24,063	6,867	48,179	22,441	23,241
Districts	308	335[(a)]	338	106[(b)]	106[(b)]	125	335[(a)]	338

Notes. DV: Individual-level forecasting accuracy (0 = incorrect, 1 = correct). Multilevel random effects logistic regression models. Significance levels: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. (a) No observations for Nunavut, Western Arctic (Northwest Territories) and Yukon. (b) No observations for Timiskaming—Cochrane. Regression analyses adjusted for age, sex, education and household income. Weights were computed using the Public Use Microdata Files (PUMFs) of the 2011 National Household Survey (Statistics Canada, 2014) for the 2011 Canadian federal election and the 2011 Ontario general election; the 2016 Canadian Census (Statistics Canada, 2022) for the 2015 Canadian federal election and the 2014 Ontario general election; and the 2021 Canadian Census (Statistics Canada, 2023) for the 2019 Canadian federal election and the 2022 Quebec general election. R = reference category.

depending on the election, while incumbent reelection increases the odds by an average of 11–48 percentage points.

The second set of models broadly confirms the previous findings in that (1) respondents who share the partisan identity of the winner are substantially more likely to make a correct forecast than those who identify with one of the losing parties and (2) education appears to have a positive but relatively small influence on forecasting accuracy, one that is mostly concentrated among losers. The most interesting results, however, have to do with political interest. On average, and all else being equal, moving from the minimum to the maximum value on the interest scale increases the probability of a correct forecast by 26 and 15 percentage points in the 2015 and 2019 federal elections, respectively. In both elections, interest has a considerably stronger impact among independents (32 and 31 percentage points in 2015 and 2019, respectively) than among both respondents identifying with losing (19 and 6 percentage points) and winning (9 and 15 percentage points) parties.

One way to better visualize and assess the impact of education and interest on forecasting accuracy is to plot the predicted probabilities of making a correct forecast in each district when these variables are set to their minimum values and then to their maximum values, while holding other covariates constant (at their mean or modal value). This is shown in Figure 3 (education only) and Figure 4 (education

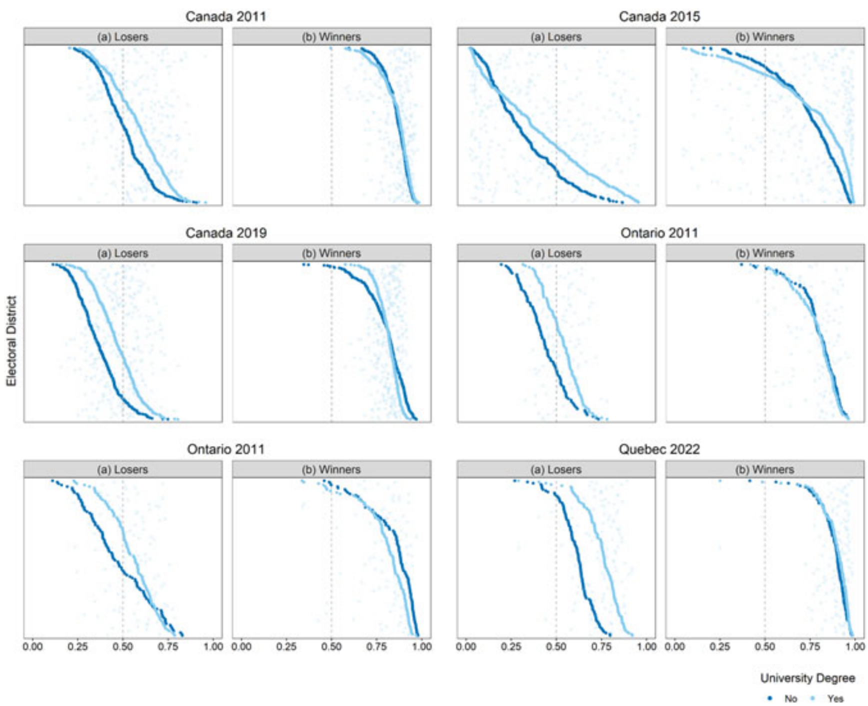

Figure 3. Predicted Probability of Correct Forecast by District According to Education.
Note. Semitransparent dots show the predicted probability for highly-educated voters before being arranged in descending order.

https://doi.org/10.1017/S0008423924000465 Published online by Cambridge University Press

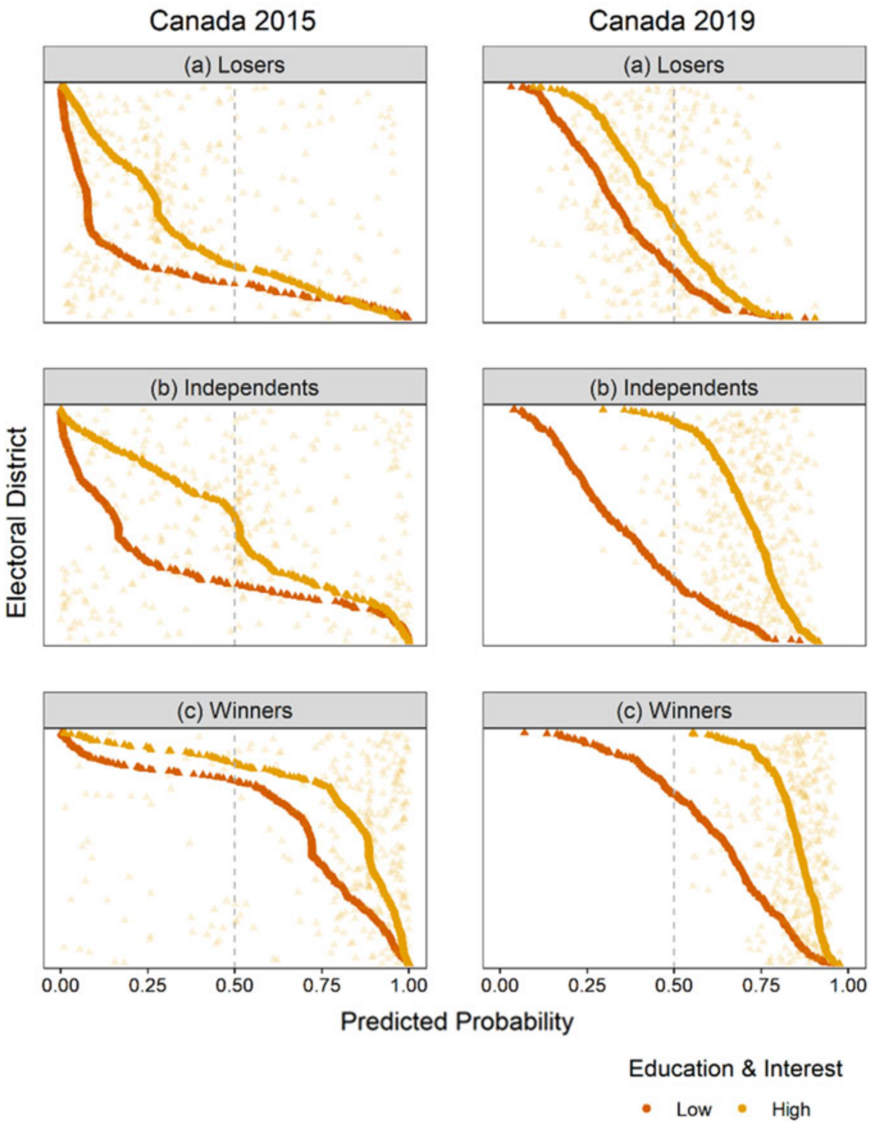

Figure 4. Predicted Probability of Correct Forecast by District According to Education and Interest.
Note. Semitransparent dots show the predicted probability for highly-educated and highly-interested voters before being arranged in descending order.

and interest) for supporters of losing and winning parties in each election. Each dot represents the predicted probability of making a correct forecast in a district. Figure 3 shows the predicted probability of making a correct forecast for voters with lower education (dark blue dots) as well as the predicted probability for voters with higher education (light blue dots). Figure 4 shows the predicted probabilities for respondents with lower education who are uninterested in politics (dark orange

https://doi.org/10.1017/S0008423924000465 Published online by Cambridge University Press

dots) compared to those of respondents with higher education who are also highly interested in politics (light orange dots). These graphs provide a clear illustration of previous findings (that is, education matters for losers but not so much for winners and political interest in conjunction with educational attainment increases accuracy among losers, independents and winners), but they also cast doubt on the usefulness of these potential markers of competence to improve forecasts in the aggregate. Although, on average, education and interest improve the likelihood of a correct forecast across districts, most of the observed improvements stay well below the 50 per cent mark among losers—those who are the least likely to correctly guess the outcome. In fact, delegating (restricting) the forecasting task to respondents with a university degree and a relatively high level of political interest (that is, above 0.6) produces percentages of correctly predicted district outcomes identical to those found for all respondents irrespective of their education or level of interest.

Table 4 shows the impact of diversity on group-level forecasts. Consistent with Murr's (2011) results, diversity does not seem to matter much. Across elections, there are no discernible patterns and most diversity indices have statistically insignificant coefficients. Using an overall measure (index) of diversity does not lead to different conclusions. Note that diversity, as measured in Table 4, is a property of the group of forecasters. However, it can also be conceptualized as a property of respondents' immediate environment (their district). Therefore, we ran additional analyses using measures of sociological diversity derived from census data within each district. These analyses, which are available in section F of the appendix, do not suggest that diverse social environments boost accuracy. Group size (logged)

Table 4. Predictors of Forecasting Accuracy in District-Level Elections at the Group Level

	CA 11[(a)]	CA 15[(a)]	CA 19[(a)]	ON 11[(a)]	ON 14[(a)]	QC 22-I[(a)]	QC 22-II[(b)]
Informational diversity							
Vote choice	−4.32	−1.85	−0.78	1.55	5.62	4.00	0.35
Education	13.01*	−0.48	4.84	−1.50	4.66	−6.48	2.79
Interest		0.42	−0.12				
Response date		−2.99	−2.90			4.97	
Sociological diversity							0.16
Age group	4.39	−1.83	−4.90*	3.52	3.41	−5.49	−6.41
Sex	−8.62*	−1.23	1.73	0.65	3.54	−0.30	−3.87
Income	−9.31	0.70	0.30	0.80	6.66	9.40	2.83
Task difficulty							
Margin of victory (*z*)	1.68**	1.42***	1.98***	5.46**	0.66	2.95**	0.10***
Reelected	7.68***		2.23***	2.21	7.18**	2.37	1.62**
Boundary changes		−0.05					
Decision making							
Logged group size	0.62	0.67	0.96	2.96	−0.21	0.23	0.38
Intercept	−1.40	3.20	−0.82	−13.37	−13.70	−5.13	2.61
Random intercept							1.29
Observations	308	335[(c)]	338	106[(d)]	106[(d)]	125	415
Pseudo-R^2	0.80	0.22	0.39	0.52	0.68	0.54	

Notes. DV: Group-level forecasting accuracy (0 = incorrect, 1 = correct). (a) Logistic regression models. (b) Multilevel random effects logistic regression model (for this model, diversity was measured at the level of the forward sortation areas (FSAs); FSAs are embedded within districts). Significance levels: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. (c) No observations for Nunavut, Western Arctic (Northwest Territories) and Yukon. (d) No observations for Timiskaming—Cochrane.

https://doi.org/10.1017/S0008423924000465 Published online by Cambridge University Press

is also statistically insignificant, which might appear as surprising. However, this is as one should expect considering the fact that the benefits of aggregation in terms of forecasting accuracy drop considerably beyond only a few respondents (as shown in Figure 1). The vast majority of districts have samples well above the 10–15 respondent threshold.

While our findings support Hypothesis 1 as highly educated and politically interested respondents tend to be more accurate than respondents with lower levels of education or interest, there is no convincing evidence in favour of Hypothesis 2. Socially and cognitively diverse crowds do not appear to outperform more uniform groups. However, the fact that educated and interested respondents are more likely to form accurate expectations about election outcomes does not necessarily mean that we can rely exclusively on their judgment to obtain better results in the aggregate.

Conclusion

Our goal in this study has been to assess individual- and group-level explanations of citizens' forecasting accuracy regarding election outcomes. More precisely, we looked at two potential markers of sophistication or competence among individual voters—education and political interest—as well as different measures of group sociodemographic and informational diversity. As expected, both education and political interest are positively related to forecasting accuracy and appear to reduce the influence of partisan preferences on expectations, especially among losers. However, the effect of education is quite small. Unfortunately, for most elections, educational attainment was the only available proxy for political sophistication. Although education has been found to possibly correlate with political attentiveness, it can be seen as a relatively weak proxy for sophistication (Luskin, 1990). As noted by McGregor (1938: 195, emphasis in original), we have to keep in mind that "[i]t is the *nature* of one's information that is determinative, not the *amount*." Interest for politics, which is closer conceptually to political sophistication, seems to play a more determinant role than education in explaining the likelihood of a correct forecast, although the evidence is limited to only two elections. More importantly, our results suggest that discriminating on the basis of education or level of interest will hardly translate into any benefits in the aggregate. Doubts can also be raised about the benefits of diversity in improving group-level accuracy. Like Murr (2011), we found no convincing evidence that a mix of heterogeneous individuals, either in terms of their sociodemographic profiles or the information they possess, helps to increase the group's chances of making a correct forecast. Our findings are also consistent with those of de Oliveira and Nisbett (2017), who concluded that diverse crowds resemble homogeneous ones when making numerical judgments, such as predicting candidates' vote shares.

Therefore, while Larrick et al. (2012) have identified expertise and diversity as necessary properties for groups to be effective forecasters, our results seem to suggest that the aggregation of citizens' electoral expectations is not easily improved upon using measures of political sophistication and informational or sociological diversity. The extant literature does point to other potentially fruitful strategies, that would however require the collection of experimental or panel data, such as group deliberation (Navajas et al., 2018; see also Becker et al., 2017; Mercier and Claidière, 2022), combining respondents' own estimates with their estimates of other people's judgments (Fujisaki et al., 2023), or weighting on prior performance (Hill and Ready-Campbell, 2011).

The present article is not without limitations. First, it is important to mention that crowd wisdom can be mobilized for a variety of tasks, including idea generation, problem solving, or even policy-making. All of these tasks involve a great deal of future-oriented thinking (for example, which decision or idea will be most efficient or yield the biggest returns). However, we cannot make the claim that our results apply to all instances in which collective intelligence is mobilized. Rather, our results speak to a much narrower strand of the literature, namely citizen forecasting of election outcomes. Second, in all six elections, members of parliaments were elected from single-member districts (SMD) according to a first-past-the-post (FPTP) rule. This limits the generalizability of our findings. Elections conducted in multimember districts (MMD) under a party-list proportional representation (PR) system could prove more challenging for voters. Third, the 2019 CES, Datagotchi, Ipsos and LPP surveys are among the very few existing surveys to include large-enough samples at the district level to reliably test the WOC prinicple in the context of elections; relying on these very large datasets comes at a price however. As already mentioned, few items were available to measure respondents' political sophistication. Finally, sophistication is a multifaceted concept (see Oscarsson and Rapeli, 2018), one we admittedly could not fully grasp with educational attainment and interest alone. Although sophistication is not limited to factual knowledge of politics, future research on voters' expectations should consider measuring the influence of *election-specific* knowledge (for example, party slogans, leaders, platforms, polling trends and so forth) on forecasting accuracy (Miller et al., 2012).

Notes

1 Conflicting evidence of knowledge effects on forecasting accuracy and wishful thinking have also been provided for elections in Israel (Babad, 1997), New Zealand (Babad et al., 1992) and Sweden (Sjöberg, 2009).

2 The wording of this item includes a wishful thinking reduction strategy (betting $1,000 of your own money), which probably aims at de-emphasizing respondents' personal preferences. However, it appears that the "imposition of external motivators to reduce biased distortions has little effect" (Marsh and Wallace, 2014: 380). Instructing respondents to ignore their own preferences or offering a financial incentive do little to reduce wishful thinking. Therefore, we do not expect the wording of the Ipsos item to have any influence on respondents' expectations, even more so because they were not asked to bet *real* money on the outcome.

3 When the highest probability was assigned to more than one possible outcome (for example, Party A has a 40 per cent chance of winning, Party B has a 40 per cent chance of winning and Party C has a 20 per cent chance of winning), the respondent's forecast was treated as a "don't know"—and thus incorrect—answer. Only 6.2 per cent of LPP respondents gave the highest probability of winning to the actual winner *and* another party. In the 2019 CES web survey, 6.4 per cent of respondents did the same.

4 As one would expect, there is considerable overlap between voter intention and PID. About three out of four respondents intended to vote for the party they felt closest to.

5 See sections A and B of the Appendix for more details on data sources and variable coding. The replication files are available online at https://doi.org/10.7910/DVN/77NKCG.

6 For the diversity measure, respondents' age was divided in six categories: "18–24," "25–34," "35–44," "45–54," "55–64," "65+."

7 Predicted probabilities were computed by holding continuous variables at their means and categorical variables at their proportions.

Acknowledgements. First, we wish to thank the three anonymous reviewers for their excellent suggestions and comments. We are also grateful to the participants of the "Perceptions of Government" panel at the 2023 Midwest Political Science Association Conference and members of the Media, Movements and Politics research group at the University of Antwerp for their feedback on previous versions of this article.

Supplementary Material. The supplementary material for this article can be found at https://doi.org/10.1017/S0008423924000465.

Declaration of Competing Interests. Competing interests: The authors declare none.

References

Althaus, Scott L. 1998. "Information Effects in Collective Preferences." *American Political Science Review* **92** (3): 545–58.

Babad, Elisha. 1997. "Wishful Thinking Among Voters: Motivational and Cognitive Influences." *International Journal of Public Opinion Research* **9** (2): 105–25.

Babad, Elisha, Michael Hills and Michael O'Driscoll. 1992. "Factors Influencing Wishful Thinking and Predictions of Election Outcomes." *Basic and Applied Social Psychology* **13** (4): 461–76.

Bartels, Larry M. 1996. "Uninformed Votes: Information Effects in Presidential Elections. *American Journal of Political Science*" **40** (1): 194–230.

Becker, Joshua, Devon Brackbill and Darmon Centola. 2017. "Network Dynamics of Social Influence in the Wisdom of Crowds." *Proceedings of the National Academy of Sciences* **114** (26): E5070–76.

Boland, Philip J. 1989. "Majority Systems and the Condorcet Jury Theorem." *The Statistician* **38** (3): 181–89.

Brennan, Jason. 2021. "In Defense of Epistocracy: Enlightened Preference Voting." In *The Routledge Handbook of Political Epistemology*, eds. Michael Hannon and Jeroen de Ridder. Abingdon, United Kingdom: Routledge.

Budescu, David V. and Eva Chen. 2015. "Identifying Expertise to Extract the Wisdom of Crowds." *Management Science* **61** (2): 267–80.

Cantril, Hadley. 1938. "The Prediction of Social Events." *Journal of Abnormal and Social Psychology* **33** (3): 364–89.

Caplan, Bryan. 2007. *The Myth of the Rational Voter: Why Democracies Choose Bad Policies*. Princeton, NJ: Princeton University Press.

Caplan, Bryan. 2009. "Majorities Against Utility: Implications of the Failure of the Miracle of Aggregation." *Social Philosophy and Policy* **26** (1): 198–211.

Condorcet, Marie-Jean-Antoine-Nicolas de Caritat, m. d., 1785. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Paris: Imprimerie Royale.

de Oliveira, Stephanie and Richard E. Nisbett. 2017. "Demographically Diverse Crowds Are Typically Not Much Wiser than Homogeneous Crowds." *Proceedings of the National Academy of Sciences* **115** (9): 2066–71.

Dolan, Kathleen A. and Thomas M. Holbrook. 2001. "Knowing Versus Caring: The Role of Affect and Cognition in Political Perceptions." *Political Psychology* **22** (1): 27–44.

Elo, Kimmo and Lauri Rapeli. 2010. "Determinants of Political Knowledge: The Effects of the Media on Knowledge and Information." *Journal of Elections, Public Opinion and Parties* **20** (1): 133–46.

Erikson, Robert S. and Kent L. Tedin. 2016. *American Public Opinion: Its Origins, Content and Impact*. New York: Routledge.

Fiechter, Joshua L. and Nate Kornell. 2021. "How the Wisdom of Crowds, and of the Crowd Within, Are Affected by Expertise." *Cognitive Research: Principles and Implications* **6** (5): 1–7.

Fujisaki, I., K. Yang, K. Ueda 2023. "On an Effective and Efficient Method for Exploiting the Wisdom of the Inner Crowd." *Scientific Reports* **13** (3608).

Gaines, Brian J. 1999. "Duverger's Law and the Meaning of Canadian Exceptionalism." *Comparative Political Studies* **32** (7): 835–61.

Gaissmaier, Wolfgang and Julian N. Marewski. 2023. "Forecasting Elections with Mere Recognition from Small, Lousy Samples: A Comparison of Collective Recognition, Wisdom of Crowds, and Representative Polls." *Judgment and Decision Making* **6** (1): 73–88.

Galton, Francis. 1907. "Vox Populi." *Nature* **75**: 450–51.

Gilens, Martin. 2019. "Citizen Competence and Democratic Governance." In *New Directions in Public Opinion*, ed. Adam J. Berinsky. New York: Routledge.

Graefe, Andreas. 2016. "Forecasting Proportional Representation Elections from Non-Representative Expectation Surveys." *Electoral Studies* **42**: 222–28.

Grofman, Bernard, Guillermo Owen and Scott L. Feld. 1983. "Thirteen Theorems in Search of the Truth." *Theory and Decision* **15**: 261–78.

Grönlund, Kimmo and Henry Milner. 2006. "The Determinants of Political Knowledge in Comparative Perspective." *Scandinavian Political Studies* **29** (4): 386–406.

Hammond, Kenneth R. 1996. *Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice*. Oxford: Oxford University Press.

Hayes, Jr., Sanuel P. 1936. "The Predictive Ability of Voters." *Journal of Social Psychology* **7** (2): 183–91.

Hemming, Victoria, Mark A. Burgman, Anca M. Hanea, Marissa F. McBride and Bonnie C. Wintle. 2018. "A Practical Guide to Structured Expert Elicitation Using the IDEA Protocol." *Methods in Ecology and Evolution* **9** (1): 169–80.

Herbst, Susan. 1993. *Numbered Voices: How Opinion Polling Has Shaped American Politics*. Chicago, IL: University of Chicago Press.

Hill, Shawndra and Noah Ready-Campbell. 2011. "Expert Stock Picker: The Wisdom of (Experts in) Crowds." *International Journal of Electronic Commerce* **15** (3): 73–102.

Hogarth, Robin M., 1978. "A Note on Aggregating Opinions." *Organizational Behavior and Human Performance* **21** (1): 40–46.

Hong, Lu and Scott E. Page. 2001. "Problem Solving by Heterogeneous Agents." *Journal of Economic Theory* **97** (1): 123–63.

Hong, Lu and Scott E. Page. 2004. "Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers." *Proceedings of the National Academy of Sciences* **101** (46): 16385–89.

Hora, Stephen C. 2004. "Probability Judgments for Continuous Quantities: Linear Combinations and Calibration." *Management Science* **50** (5): 597–604.

Jacobson, Jonas, Jasmine Dobbs-Marsh, Varda Liberman and Julia Minson. 2011. "Predicting Civil Jury Verdicts: How Attorneys Use (and Misuse) a Second Opinion." *Journal of Empirical Legal Studies* **8** (S1): 99–119.

Johnston, Richard and Fred Cutler. 2009. "Canada: The Puzzle of Local Three-Party Competition." In *Duverger's Law of Plurality Voting: The Logic of Party Competition in Canada, India, the United Kingdom and the United States*, eds. Bernard Grofman, André Blais and Shaun Bowler. New York: Springer.

Kazmann, Raphael G. 1973. "Democratic Organization: A Preliminary Mathematical Model." *Public Choice* **16**: 17–26.

Krizan, Zlatan, Jeffrey C. Miller and Omesh Johar. 2010. "Wishful Thinking in the 2008 U.S. Presidential Election." *Psychological Science* **21** (1): 140–46.

Kuklinski, James H. and Paul J. Quirk. 2000. "Reconsidering the Rational Public: Cognition, Heuristics, and Mass Opinion." In *Elements of Reason: Cognition, Choice, and the Bounds of Rationality*, eds. Arthur Lupia, Matthew D. McCubbins and Samuel L. Popkin. New York: Cambridge University Press.

Ladha, Krishna K. 1992. "The Condorcet Jury Theorem, Free Speech, and Correlated Votes." *American Journal of Political Science* **36** (3): 617–34.

Landemore, Hélène. 2012. "Collective Wisdom: Old and New." In *Collective Wisdom: Principles and Mechanisms*, eds. Hélène Landemore and Jon Elster. New York: Cambridge University Press.

Larrick, Richard P., Albert E. Mannes and Jack B. Soll. 2012. "The Social Psychology of the Wisdom of Crowds." In *Social Judgment and Decision Making*, ed. Joachim I. Krueger. New York: Psychology Press.

Le, Kien and My Nguyen. 2021. "Education and Political Engagement." *International Journal of Educational Development* **85**: 17–26.

Leiter, Debra, Andreas Murr, Ericka Rascón Ramírez and Mary Stegmaier. 2018. "Social Networks and Citizen Election Forecasting: The More Friends the Better." *International Journal of Forecasting* **34** (2): 235–48.

Leiter, Debra, Jack L. Reilly and Mary Stegmaier. 2020. "Echoing Certainty in Uncertain Times: Network Partisan Agreement and the Quality of Citizen Forecasts in the 2015 Canadian Election." *Electoral Studies* **63** (102115).

Lewis-Beck, Michael S. and Andrew Skalaban. 1989. "Citizen Forecasting: Can Voters See Into the Future?" *British Journal of Political Science* **19** (1): 419–27.

Lewis-Beck, Michael S. and Charles Tien. 1999. "Voters as Forecasters: A Micromodel of Election Prediction." *International Journal of Forecasting* **15** (2): 175–84.

List, Christian and Robert E. Goodin. 2001. "Epistemic Democracy: Generalizing the Condorcet Jury Theorem." *Journal of Political Philosophy* **9** (3): 277–306.

Lorenz, Jan, Heiko Rauhut, Frank Schweitzer and Dirk Helbing. 2011. "How Social Influence Can Undermine the Wisdom of Crowd Effect." *Proceedings of the National Academy of Sciences* **108** (22): 9020–25.

Luskin, Robert C. 1990. "Explaining Political Sophistication." *Political Behavior* **12** (4): 331–61.

Marsh, Kerry L. and Harry M. Wallace. 2014. "The Influence of Attitudes on Beliefs: Formation and Change." In *The Handbook of Attitudes*, eds. Dolores Albarracin, Blair T. Johnson and Mark P. Zanna. New York: Psychology Press.

McGregor, Douglas. 1938. "The Major Determinants of the Prediction of Social Events." *Journal of Abnormal and Social Psychology* **33** (2): 179–204.

Meffert, Michael F., Sascha Huber, Thomas Gschwend and Franz Urban Pappi. 2011. "More Than Wishful Thinking: Causes and Consequences of Voters' Electoral Expectations About Parties and Coalitions." *Electoral Studies* **30** (4): 804–15.

Mercier, H. and N. Claidière. 2022. "Does Discussion Make Crowds Any Wiser?" *Cognition* **222** (1049122).

Miller, Michael K., Guanchun Wang, Sanjeev R. Kulkarni, Vincent H. Poor and Daniel N. Osherson. 2012. "Citizen Forecasts of the 2008 U.S. Presidential Election." *Politics and Policy* **40** (6): 1019–52.

Mongrain, Philippe. 2021a. "Did You See It Coming? Explaining the Accuracy of Voter Expectations for District and (Sub)national Election Outcomes in Multi-party Systems." *Electoral Studies* **71** (102317).

Mongrain, Philippe. 2021b. "A Technocratic View of Election Forecasting: Weighting Citizens' Forecasts According to Competence." *International Journal of Public Opinion Research* **33** (3): 713–23.

Mongrain, Philippe. 2023. "With a Little Help From My Friends? The Impact of Social Networks on Citizens' Forecasting Ability." *European Journal of Political Research* **62** (4): 1320–46.

Morisi, Davide and Thomas Leeper. 2024. "What Influences Citizen Forecasts? The Effects of Information, Elite Cues, and Social Cues." *Political Behavior* **46**: 21–41.

Murr, Andreas. 2017. "Wisdom of Crowds." In *The SAGE Handbook of Electoral Behaviour*, ed. Kai Arzheimer, Jocelyn Evans and Michael S. Lewis-Beck. Thousand Oaks, CA: SAGE.

Murr, Andreas E. 2015. "The Wisdom of Crowds: Applying Condorcet's Jury Theorem to Forecasting US Presidential Elections." *International Journal of Forecasting* **31** (3): 916–29.

Murr, Andreas Erwin. 2011. "'Wisdom of Crowds'? A Decentralised Election Forecasting Model that Uses Citizens' Local Expectations." *Electoral Studies* **30** (4): 771–83.

Navajas, Joaquin, Tamara Niella, Gerry Garbulsky, Bahador Bahrami and Mariano Sigman. 2018. "Aggregated Knowledge from a Small Number of Debates Outperforms the Wisdom of Large Crowds." *Nature Human Behaviour* **2**: 126–32.

Oscarsson, Henrik and Lauri Rapeli. 2018. "Citizens and Political Sophistication." In *Oxford Research Encyclopedia of Politics*. Oxford: Oxford University Press. https://doi.org/10.1093/acrefore/9780190228637.013.220.

Page, Benjamin I. and Robert Y. Shapiro. 1999. "The Rational Public and Beyond." In *Citizen Competence and Democratic Institutions*, eds. Stephen L. Elkin and Karol E. Soltan. University Park, PA: Pennsylvania State University Press.

Page, Benjamin. I. and Robert Y. Shapiro. 1992. *The Rational Public: Fifty Years of Trends in Americans' Policy Preferences*. Chicago: University of Chicago Press.

Page, Scott E. 2007. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton, NJ: Princeton University Press.

Quirk, Paul J. 2014. "Making It Up on Volume: Are Larger Groups Really Smarter?" *Critical Review* **26** (1–2): 129–50.

Rehm, Jürgen T. and Volker Gadenne. 2013. *Intuitive Predictions and Professional Forecasts: Cognitive Processes and Social Consequences*. Oxford: Pergamon Press.

Rhode, Paul W. and Koleman Strumpf. 2013. "The Long History of Political Betting Markets: An International Perspective." In *The Oxford Handbook of the Economics of Gambling*, eds. Leighton Vaughan-Williams and Donald S. Siegel. Oxford: Oxford University Press.

Satopää, Ville A., Jonathan Baron, Dean P. Foster, Barbara A. Mellers, Philip E. Tetlock and Lyle H. Ungar. 2014. "Combining Multiple Probability Predictions Using a Simple Logit Model." *International Journal of Forecasting* **30** (2): 344–56.

Shapley, Lloyd and Bernard Grofman. 1984. "Optimizing Group Judgmental Accuracy in the Presence of Interdependencies." *Public Choice* **43** (3): 329–43.

Sjöberg, Lennart. 2009. "Are All Crowds Equally Wise? A Comparison of Political Election Forecasts by Experts and the Public." *Journal of Forecasting* **28** (1): 1–18.

Špecián, Petr. 2022. *Behavioral Political Economy and Democratic Theory Fortifying Democracy for the Digital Age*. Abingdon, United Kingdom: Routledge.

Statistics Canada. 2014. "2011 National Household Survey Public Use Microdata File (PUMF): Individuals File." Abacus Data Network, V1. https://hdl.handle.net/11272.1/AB2/PYYXXR.

Statistics Canada. 2022. "2016 Census Public Use Microdata File (PUMF). Individuals File." Abacus Data Network, V2. https://hdl.handle.net/11272.1/AB2/GDJRT8.

Statistics Canada. 2023. "2021 Census Public Use Microdata File (PUMF). Individuals File." Abacus Data Network, V4. https://hdl.handle.net/11272.1/AB2/1WTDOP.

Surowiecki, James. 2004. *The Wisdom of Crowds*. London: Abacus.

Tetlock, Philip E., 2017. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton, NJ: Princeton University Press.

Tofallis, Chris. 2015. "A Better Measure of Relative Prediction Accuracy for Model Selection and Model Estimation." *Journal of the Operational Research Society* **66** (8): 1352–62.

Treynor, Jack L. 1987. "Market Efficiency and the Bean Jar Experiment." *Financial Analysts Journal* **43** (3): 50–53.

Turper, Sedef and Kees Aarts. 2017. "Political Trust and Sophistication: Taking Measurement Seriously." *Social Indicators Research* **130** (1): 415–34.

Uhlaner, Carole J. and Bernard Grofman. 1986. "The Race May Be Close but My Horse Is Going to Win: Wish Fulfillment in the 1980 Presidential Election." *Political Behavior* **8** (2): 101–29.

van Dolder, Dennie, Martijn J. van den Assem, 2018. "The Wisdom of the Inner Crowd in Three Large Natural Experiments." *Nature Human Behaviour* **2**: 21–26.

Wallis, Kenneth F. 2014. "Revisiting Francis Galton's Forecasting Competition." *Statistical Science* **29** (3): 420–24.

Whitehorn, Alan. 1997. "Alexa McDonough and Atlantic Breakthrough for the New Democratic Party." In *The Canadian General Election of 1997*, ed. Alan Frizzell and Jon H. Pammett. Toronto: Dundern Press.

Cite this article: Mongrain, Philippe, Nadjim Fréchet, Brian Thompson Collart and Yannick Dufresne. 2025. "Working the Crowd: Citizen Forecasting, Sophistication and Diversity in Canadian Federal and Provincial Elections." *Canadian Journal of Political Science* 1–27. https://doi.org/10.1017/S0008423924000465