nature human behaviour

LETTERS

https://doi.org/10.1038/s41562-018-0302-y

Asking about social circles improves election predictions

M. Galesic[1,2*], W. Bruine de Bruin[3,4,5], M. Dumas[1,6], A. Kapteyn[5], J. E. Darling[5,7] and E. Meijer[5]

Election outcomes can be difficult to predict. A recent example is the 2016 US presidential election, in which Hillary Clinton lost five states that had been predicted to go for her, and with them the White House. Most election polls ask people about their own voting intentions: whether they will vote and, if so, for which candidate. We show that, compared with own-intention questions, social-circle questions that ask participants about the voting intentions of their social contacts improved predictions of voting in the 2016 US and 2017 French presidential elections. Responses to social-circle questions predicted election outcomes on national, state and individual levels, helped to explain last-minute changes in people's voting intentions and provided information about the dynamics of echo chambers among supporters of different candidates.

Past polls have asked people to indicate who they think will win the election or to judge the probability that each candidate will win. Possibly because people know how their social contacts will vote[1], such election-winner questions have successfully predicted many election outcomes[2,3]. However, election-winner questions have some imperfections. They do not straightforwardly predict actual vote shares because they ask for expectations that a candidate will win and not for the estimated percentage of voters who will vote for the candidate. They produce predictions on the national, but not on the state and individual levels. Furthermore, they might be influenced by occasional inaccurate predictions reported in the media[4].

Social-circle questions can provide useful information in election polls for several reasons. It has been shown that people make relatively accurate judgements about various characteristics of their immediate social circles[5–7]. Averaged across a national sample, respondents' judgements of the percentage of their social contacts with different characteristics (from levels of household wealth to health problems) come close to the actual percentage in the general population[8]. Moreover, reporting about friends' preferences for an unpopular candidate can be less embarrassing than admitting to personally having these preferences[9,10]. People's reports about their social contacts may also reveal the social interactions that shape their beliefs and behaviours, and predict changes in their own intentions over time due to social influence processes[11,12]. Finally, social circle questions provide information about individuals who were not included in the sample of a particular poll, thus implicitly increasing sample size and possibly reducing some of its sampling, non-response and coverage biases[13].

We studied the usefulness of social-circle questions in two different elections: the 2016 US presidential election and the 2017 French presidential election. Held on 8 November 2016, the US election essentially focused on two candidates: Hillary Clinton and Donald Trump. Other candidates were collectively not predicted to win more than about 10% of the vote. By contrast, the French elections involved at least five prominent candidates: François Fillon, Benoît Hamon, Marine Le Pen, Emmanuel Macron and Jean-Luc Mélenchon (in addition to six others). The French election was held in two rounds: the first round on 23 April 2017 and the second, focusing on the two top candidates (Marine Le Pen and Emmanuel Macron), on 7 May 2017.

In each country, we asked two-part questions about participants' social circles: (1) "What percentage of your social contacts is likely to vote in the upcoming election?" and (2) "Of all your social contacts who are likely to vote, what percentage do you think will vote for [candidate]?" (see Methods). In the United States, we asked social-circle questions in two national surveys: the GfK election poll conducted in the week before the election[14] and the University of Southern California (USC) Dornsife/LA Times election poll conducted daily from July 2016 until after the election[15,16]. We compared the social-circle questions with questions about own intentions to vote. After asking whether participants plan to go vote, the GfK poll asked, "If you were to vote in the presidential election that's being held on November 8th, which candidate would you choose?"[14], while the USC poll used a probabilistic question, "If you do vote in the election, what is the percent chance that you will vote for Clinton, Trump or someone else?"[17,18] (see Methods). In addition, the USC poll elicited election-winner expectations: "What is the percent chance that Clinton, Trump or someone else will win?" In France, we asked the social-circle questions in the election poll conducted by the survey research company BVA on a national sample in the week before the first round of the election. We compared the answers to social-circle questions with answers to own-intention questions of the form: "Which candidate are you most likely to vote for?" (see Methods). These data allowed us to investigate five research questions, discussed below.

Our first research question examined whether asking about social circles improved predictions of national election results. Tables 1 and 2 summarize results from the US and French elections, including established measures of prediction error. In the United States, social-circle questions were more accurate than own-intention questions in predicting the whole distribution of vote shares for different candidates (Table 1 and Supplementary Dataset). We found lower values of the error measures Mosteller 3 (ref. [19]) and $\overline{A}$[20,21] for social-circle questions than for own-intention questions. Compared with own-intention questions, social-circle questions predicted the difference in the popular vote for Clinton and Trump

[1]Santa Fe Institute, Santa Fe, NM, USA. [2]Max Planck Institute for Human Development, Berlin, Germany. [3]Centre for Decision Research, Leeds University Business School, University of Leeds, Leeds, UK. [4]Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, USA. [5]Center for Economic and Social Research, University of Southern California, Los Angeles, CA, USA. [6]London School of Economics, London, UK. [7]VHA Health Services Research and Development Center for the Study of Healthcare Innovation, Implementation, and Policy, VHA Greater Los Angeles Health Care System, Los Angeles, CA, USA. *e-mail: galesic@santafe.edu

Table 1 | Results of the 2016 US presidential election, predictions based on survey questions and indicators of predictions' accuracy

	Election results	Aggregate polls	GfK poll		USC poll	
		Own intention	Own intention	Social circle	Own intention	Social circle
Participation rate	54.8%	-	76.5%	72.8%	80.4%	76.4%
Popular vote						
Clinton	48.2%	45.7%	46.2%	50.2%	44.8%	45.5%
Trump	46.1%	41.8%	43.2%	43.7%	46.3%	49.4%
Other	5.7%	12.5%	10.6%	6.1%	8.9%	5.1%
Electoral votes to Clinton (based on the state-level predictions of the popular vote)	232	323	298	293	305	258
Error measures for national predictions (state-level predictions) of the popular vote						
Error of the predicted difference between two main candidates (Mosteller 5)		1.8 (1.3)	0.9 (1.5)	4.4 (5.0)	−3.6 (−0.6)	−6.0 (−1.4)
Average absolute error of the predicted vote share for all candidates (Mosteller 3)		4.5 (2.1)	3.3 (7.1)	1.6 (3.9)	2.3 (6.6)	2.2 (5.9)
Average absolute log ratio of the predicted and actual odds for all candidates ($\overline{A}$)		0.38 (0.14)	0.29 (0.47)	0.08 (0.24)	0.21 (0.45)	0.12 (0.35)

Results of the GfK poll were based on a probabilistic national sample of $N=1,822$ participants interviewed from 3 November 2016 to the morning of 8 November 2016. Results of the USC poll were based on a probabilistic national sample of $N=2,229$ participants interviewed from 31 October 2016 to 7 November 2016. For error measures, lower absolute values are better. For aggregate polls, question wording varied. In the GfK poll, own-intention questions asked which candidate participants would vote for, and in the USC poll, they asked participants to judge the per cent chance of voting for each candidate. Comparison with the actual election results and with the aggregate results of 1,106 national polls (for predictions of the popular vote) and 3,073 state polls (for predictions of the electoral votes), as summarized at FiveThirtyEight[23,35], suggest that both GfK and USC polls have satisfactory accuracy. Note that Clinton eventually received 227 and Trump 304 electoral votes, because some electors had defected.

Table 2 | Results of the 2017 French presidential election, predictions based on survey questions and indicators of predictions' accuracy

	Election round 1				Election round 2			
	Election results	Aggregate polls	BVA poll		Election results	Aggregate polls	BVA poll	
		Own intention	Own intention	Social circle		Own intention	Own intention	Social circle
Participation rate*	75.8%	73.2%	89.6%	74.7%	66.0%	74.0%	81.3%	68.7%
Popular vote								
Macron	24.0%	23.8%	25.9%	24.6%	66.1%	60.8%	62.3%	64.2%
Le Pen	21.3%	22.3%	22.3%	21.8%	33.9%	39.2%	38.5%	35.8%
Fillon	20.0%	19.5%	15.2%	17.3%				
Mélenchon	19.6%	18.9%	19.7%	19.6%				
Hamon	6.4%	7.6%	7.3%	8.8%				
Others	8.7%	7.9%	9.6%	7.9%				
Error measures								
Error of the predicted difference between the main candidates (Macron and Le Pen, Mosteller 5)		−1.2	0.8	0.1		−10.6	−8.4	−3.9
Average absolute error of the predicted vote share for all candidates (Mosteller 3)		0.8	1.6	1.2		5.3	4.2	1.9
Average absolute log ratio of the predicted and actual odds for all candidates ($\overline{A}$)		0.08	0.12	0.11		0.23	0.18	0.09

Results of the BVA poll were based on a national quota sample of $N=1,003$ participants interviewed from 17–22 April 2017. For error measures, lower values are better. For comparison, we provide actual election results, as well as aggregate poll results based on questions about own voting intentions asked in 20 different polls in a week before election round 1, and 18 different polls before round 2 (ref. [36]). Note that, compared with the aggregate polls, own-intention predictions from the BVA poll have satisfactory accuracy. Social-circle predictions always outperform those based on own-intention questions. *Non-participation count includes people who did not vote and those who casted blank ballots.

less well. This is reflected in the error measure Mosteller 5 (ref. [19]), which considers only the two main candidates. Which of these well-established error measures is more important will depend on the number of prominent candidates in a given election and on the aims of the particular poll. In France, social-circle questions performed better than own-intention questions on all error measures and in

both election rounds (Table 2). In both countries, social-circle questions produced more accurate predictions of participation rates. Possibly reflecting media forecasts of a substantial Clinton win, the USC poll's election-winner question erroneously predicted that Clinton would win, giving her a 53.4% chance compared with 42.5% for Trump and 4.1% for other candidates.

Second, we investigated whether social-circle questions improved predictions of state election results. In both US polls, social-circle questions produced more accurate predictions of state winners, as compared with own-intention questions. Consequently, social-circle questions predicted the number of electoral votes for each candidate better than own-intention questions (Table 1 and Supplementary Dataset). USC's social-circle questions were the only ones that predicted Trump winning the majority of the electoral votes. These results were obtained despite average sample sizes of only 27 participants per state for the GfK polls, and 44 per state for the USC polls. The GfK and USC polls predicted 67% and 77% of all states correctly with social-circle questions, respectively, as compared to 65% and 61% with own-intention questions. Social-circle questions were particularly useful for predicting election outcomes in a priori defined 'swing states'[22] (Colorado, Florida, Iowa, Michigan, North Carolina, Nevada, New Hampshire, Ohio, Pennsylvania, Virginia and Wisconsin). For GfK and USC polls, social-circle questions correctly predicted 82% and 73% of swing states, respectively, whereas own-intention questions predicted 46% and 64% of swing states correctly. For further comparison, aggregates of 3,073 state polls (including 60 polls per state on average) correctly predicted 90% of states and 55% of swing states[23]. Social-circle questions were also more successful than both own-intention questions and aggregate polls in predicting winners of the five swing states that unexpectedly went to Trump (Florida, Michigan, North Carolina, Pennsylvania and Wisconsin). They predicted four of these states correctly, compared with three by own-intention questions and zero by aggregate polls. In summary, these results suggest that people possess valuable information about their social circles, which could be used to improve election predictions at national and state levels.

Third, we examined whether social-circle questions benefited predictions of individual voting behaviour. We found that changes in social-circle reports predicted subsequent changes in own voting intentions. For participants who completed the USC surveys in August, September, late October/early November 2016 and immediately after the election ($N = 1{,}263$), social-circle questions contributed to the explanation of their actual voting behaviour over and above own-intention questions (Supplementary Table 1). Figure 1a shows that, up until the week before the election, participants reported that they were on average more likely to vote for Clinton than for Trump, whereas more participants ended up voting for Trump than for Clinton. Figure 1b shows a reversal towards Trump in social-circle reports as early as September 2016, when own-intention questions were still predicting a lead for Clinton. A weighted average of own-intention and social-circle estimates produced more accurate predictions of individual voting behaviour than own intentions alone (with weights being regression coefficients in a model that included both types of questions; see Supplementary Table 1). Of note, election-winner expectations did not contribute to explanations of voting behaviour over and above own-intention and social-circle questions (Supplementary Table 1). Similar patterns were observed throughout the pre-election period, with own-intention and social-circle reports jointly contributing to reports of own intentions in subsequent survey waves (Supplementary Table 2a,b).

Fourth, we analysed whether social-circle questions helped to explain last-minute changes in voting intentions. As mentioned, not all participants ended up voting for the candidate they announced as their favourite in the week before the election (see also Supplementary Table 3). Participants whose own intentions

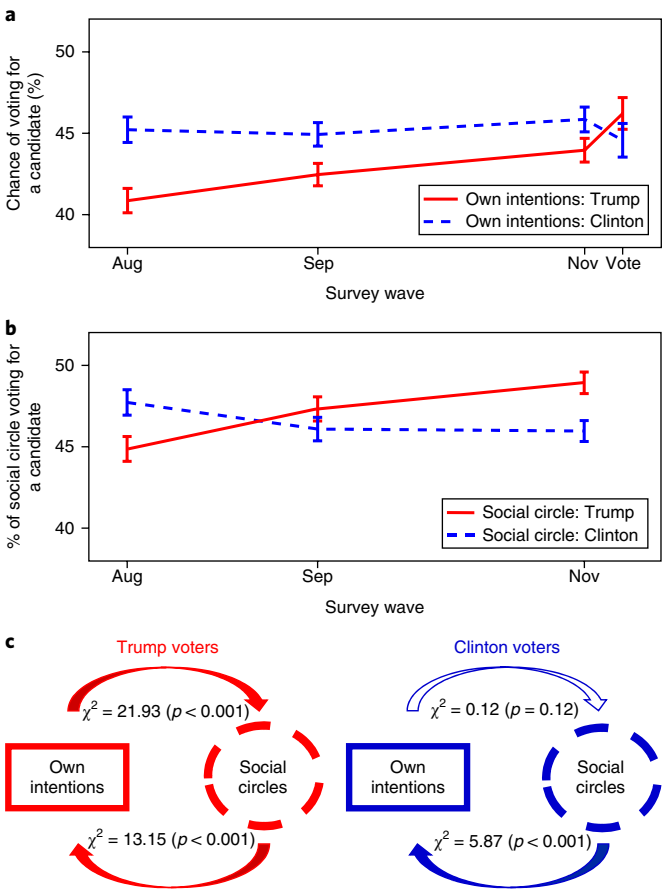

Fig. 1 | Social-circle reports anticipate changes in own voting intentions. **a**,**b**, Shifts in the average individual voting intentions and behaviour (**a**) were announced by shifts in social circles (**b**). Error bars show within-subject 95% confidence intervals[43]. **c**, Granger causality tests suggest that social circles influenced own intentions that were reported weeks later. For Trump voters, own intentions also seemed to influence subsequent social-circle reports. Results are for $N = 1{,}263$ individuals who participated in the USC survey waves in August, September, early November 2016 and immediately after the election. Because we are interested in predictions of individual behaviour in this particular sample, estimates are adjusted for the likelihood of voting (see Methods) and are otherwise unweighted.

mismatched those in their social circles were less likely to eventually vote for their intended candidate (Supplementary Fig. 1). Although some of these participants had less strong intentions to vote for their preferred candidate in the first place, our overall results suggest that social-circle reports foretold a switch in voting intentions before it happened. Generally, changes in participants' social circles over time predicted their later intentions to vote for specific candidates and to vote at all, as revealed by vector autoregression modelling and Granger causality tests[24,25] (Fig. 1c and Supplementary Tables 4 and 5). This pattern of results was found for both Trump and Clinton voters, suggesting that participants' perceptions of how social contacts would vote affected their own beliefs regarding the candidates. In addition, Trump voters appeared to influence later changes in their social circles, whereas Clinton voters did not.

Our final research question examined whether asking about social circles provided insights about the dynamics of echo chambers. Social-circle questions revealed increased homogenization of Trump voters' social circles over time. Figure 2 shows the percentage of like-minded social contacts that Trump and Clinton voters reported in the USC poll. Extreme echo chambers would be seen in

social circles that include nearly 100% like-minded individuals. In August 2016, individuals who eventually voted for Trump and those who eventually voted for Clinton had similarly diverse social circles. Their social circles included on average around 68% and 71% like-minded individuals, respectively. However, over time, social circles of Trump voters included increasingly more like-minded individuals. By contrast, we did not observe a similar increase in the homogeneity of Clinton voters' social circles. Hence, the additional Trump voters in the social circles of our Trump-voting participants were probably coming from people who previously did not plan on voting, were undecided or were planning to vote for third candidates. It is also possible that Trump voters were more inclined to exclude Clinton supporters from their social circles than were Clinton voters to exclude Trump supporters. In any case, the homogenization continued after the election, when Trump voters reported social circles that consisted of on average 77% like-minded individuals, compared with 68% among Clinton voters. Just after the election, 42% of Trump voters had social circles that included ≥90% like-minded individuals, compared with only 30% of such participants among Clinton voters. When further investigating whether the homogeneity of social circles was related to sociodemographic variables (Supplementary Table 6 and Supplementary Figs. 2 and 3), we found moderate relationships with participants' political leanings, age, education and US state of residence. For Trump voters, homogenization of social circles was particularly pronounced among voters ≥65 years of age, in particular in states that voted Republican. Education played an additional role, with social circles of less-educated Trump voters homogenizing more and faster than those of more-educated Trump voters. Age did not predict homogenization for Clinton voters. In addition, in strongly Democrat states, more-educated Clinton voters had somewhat more homogeneous circles than less-educated voters (see Supplementary Figs. 2 and 3).

Taken together, our results make two contributions. First, people's reports about their social circles can improve predictions of election results and enhance understanding of individual voting behaviour. We observed these findings across different poll designs in two countries with different political systems, suggesting that other election polls could potentially benefit from including social-circle questions. Social-circle questions may also be useful in surveys that aim to forecast other beliefs and behaviours. One reason for the usefulness of social-circle questions could be the increased implicit sample size, which was reflected in reduced standard errors of social-circle compared with own-intention questions. For the GfK poll, standard errors for predictions from social circles versus own intentions were 0.78 versus 1.22 for Clinton, 0.78 versus 1.21 for Trump and 0.31 versus 0.77 for other candidates. Similarly, USC poll standard errors for predictions from own intentions were 0.92 versus 1.35 for Clinton, 0.94 versus 1.45 for Trump and 0.35 versus 0.75 for others. In addition, social-circle reports might provide information about people who would otherwise be missing from polls due to coverage, sampling or non-response errors[13]. Thus, social-circle questions could be particularly useful when polls must rely on relatively small samples in some states. Another reason for the usefulness of social-circle questions might be that participants who are reluctant to report that they favour a potentially embarrassing option could nevertheless be willing to report that their social

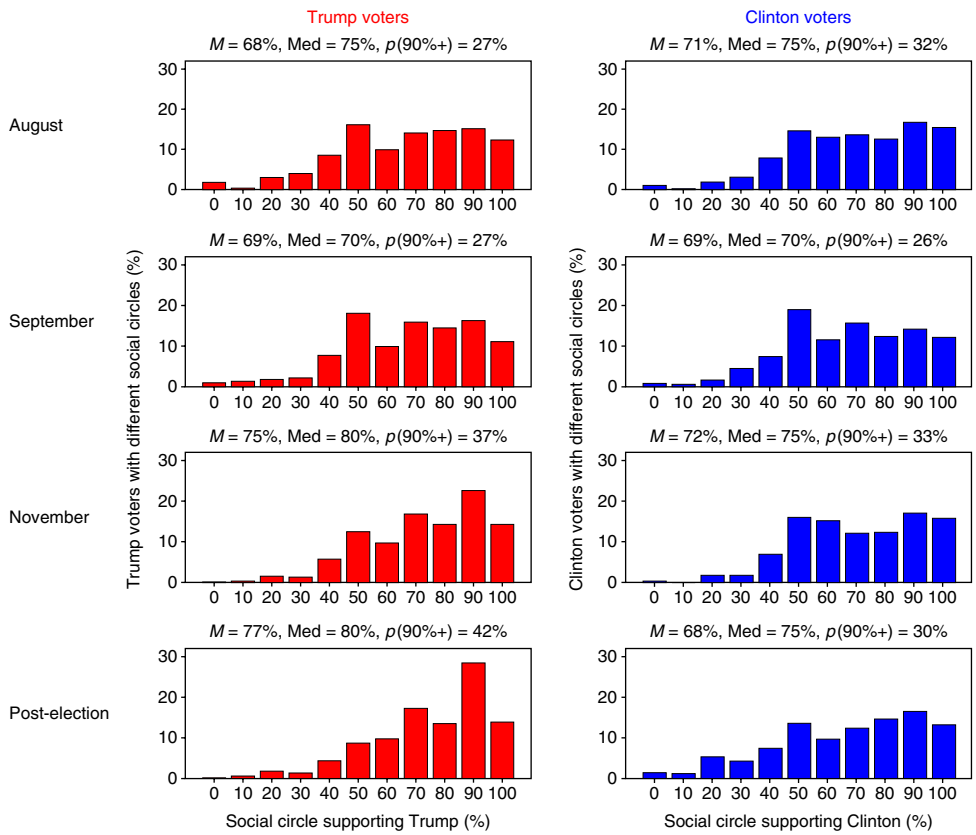

Fig. 2 | The extent of 'echo chambers' among Clinton and Trump voters, over time. Data are for $N = 1{,}263$ individuals who participated in August, September, November, and post-election USC survey waves. Bars show percentage of participants (*y* axis) reporting that a particular percentage of their social contacts (*x* axis) supported their own preferred candidate. The numbers within each panel are the mean (*M*) and median (Med) percentage of social contacts supporting participants' own preferred candidate, as well as the percentage of participants who reported that 90% or more of their social contacts supported their own preferred candidate (p(90%+)).

circle favours it[9,10]. Finally, through processes of social influence, individuals' voting intentions could indeed become more similar to the prevailing opinion in their social circles over time[11,12].

Our second main finding is that asking about social circles can provide insights into the social dynamics that shape individual voting behaviour. We found interesting differences between Trump voters and Clinton voters. Trump voters seemed to be influenced by their peers and influenced them in turn (Fig. 1c and Supplementary Table 5). Clinton voters seemed to be mostly influenced by others while not influencing others themselves. One explanation for this finding is that Trump voters might have been more likely to project their own intentions onto the intentions of their peers, perceiving them as more similar to themselves than they were[26]. It is also possible that they were influencing their friends and family to vote for Trump, or that the composition of their social circles was changing over time to include more Trump supporters. These differences between Trump and Clinton voters are echoed by the finding of increased homogenization of social circles of Trump, but not Clinton, voters (Fig. 2). This pattern of homogenization probably results from several inter-related processes. One is Trump supporters' increasing suspicion of the 'mainstream media'[27] and greater reliance on in-group information sources. Another is 'unfriending' of people with incompatible political opinions, which is practiced by supporters of both candidates[28]. Perceived homogeneity can further increase if people are reluctant to disclose political views that are not in accord with the prevailing opinion among their peers[29]. Our results are in line with a recent analysis of Twitter data that showed significant homogeneity and isolation of Trump voters relative to supporters of other candidates[30].

Overall, social-circle questions are a way of tapping into the 'local' wisdom of crowds[31–33]. Standard election-winner questions attempted to tap into the wisdom of crowds by asking people about their predictions for overall election results[2–4]. This can be problematic because people do not have direct experience with everyone in the general population. Instead, they have to make population inferences based, at least in part, on second-hand information, such as occasional erroneous predictions reported in the media. By contrast, social-circle questions harvest people's direct experiences with their immediate social environments[5–8,34]. It is important to note that the survey sampling design will affect the usefulness of social-circle questions. If social-circle reports come from a biased, non-representative sample of the overall population, their average will probably be a biased estimate of true population values. In well-designed samples of the population of interest, social-circle questions can improve survey estimates, especially when these are otherwise based on small samples or when they pertain to socially sensitive beliefs and behaviours. In addition, social-circle reports can provide valuable information about social interactions that shape individual beliefs and behaviours.

Methods

Aggregate polls. For election predictions based on aggregate polls in the United States, we used data from 1,106 national polls[23] and 3,073 state polls, summarized by the website FiveThirtyEight[35]. In France, we used results of 20 different polls conducted in the week before the first round and 18 before the second round of the election[36].

USC Dornsife/LA Times presidential election poll. *Question texts.* Introduction: "In this interview, we will ask you questions about the upcoming general election for the President of the United States on Tuesday November 8, 2016. All questions ask you to think about the percent chance that something will happen in the future. The percent chance can be thought of as the number of chances out of 100. You can use any number between 0 and 100. For example, numbers such as: 2 and 5 percent may be 'almost no chance', 20 percent or so may mean 'not much chance', a 45 or 55 percent chance may be a 'pretty even chance', 80 percent or so may mean a 'very good chance', and a 95 or 98 percent chance may be 'almost certain.'"

Own-intention questions: "(1) What is the percent chance that you will vote in the Presidential election? (2) If you do vote in the election, what is the percent chance that you will vote for Clinton? And for Trump? And for someone else?" The order of the candidates was randomized for this and other questions in all three polls.

Social-circle questions: "Now we would like you to think of your friends, family, colleagues and other acquaintances of 18 years of age or older that you have communicated with at least briefly within the last month, either face-to-face, or otherwise. We will call these people your social contacts. (1) What percentage of your social contacts is likely to vote in the upcoming election for President? For instance, 0% means that you think none of your social contacts will vote, and 100% means that all of your social contacts will vote. If you are not sure, just try to give your best guess. (2) For the next question, please consider only those of your social contacts who are likely to vote in the upcoming election for U.S. President. Of all your social contacts who are likely to vote, what percentage do you think will vote for Clinton, Trump, or someone else? For instance, 0% would mean that you think no voters in your social circle will vote for that candidate, and 100% means that all voters in your social circle will vote for that candidate. Again, if you are not sure, just try to give your best guess."

Election-winner expectations questions: "What is the percent chance that Clinton will win? And Trump? And other candidates?"

Sample. Participants were members of the Understanding America Study (UAS) at USC's Dornsife Center for Economic and Social Research. This longitudinal study[37] included close to 6,000 US residents who were randomly selected from among all households in the United States using address-based sampling. They were recruited by a combination of mail, phone and web surveys. Members of recruited households who did not have Internet access were provided with tablets and Internet service. In May 2016, all panel members who were US citizens were asked to respond to a pre-election survey. Those who completed the study and agreed to participate constituted the election poll panel.

Starting from 4 July 2016, each member of the poll panel was invited to answer the election poll once a week[38]. Members received the invitation to participate each week on the same day of the week, but they were allowed to respond up to 6 days later (that is, until the day before the next invitation). On average, across waves, study completion rates were 70%. As reported by UAS[38], the average panel recruitment rate, reflecting those individuals who completed the initial mail survey among those who consented to participate in the UAS, was 29.7%. The percentage of active panel members was 13.6%[37]. Combined with the study completion rate, the cumulative response rate for the studies reported here was 9.5%.

Five study waves included all three types of questions of interest here (own-intention, social-circle and election-winner questions): (1) 11–23 July (N=1,782), (2) 8–20 August (N=2,726), (3) 12–24 September (N=2,882), (4) from 31 October to 7 November (N=2,240), and (5) after the election, 9–21 November 2016 (N=3,798). In all waves except wave 4, all questions were asked together. In wave 4 only, social-circle questions were asked in a separate questionnaire from own-likelihood and election-winner questions. Social-circle questions were asked starting from 3 November 2016, and all participants completed all questions within a time period of about 3 days. The pattern of results presented in the main text does not change if we analyse only the 969 participants who, in wave 4, completed all questions on the same day. This would be expected because all pre-election 'surprises' occurred before this survey period, with the last being the 28 October 2016 FBI announcement that they were re-opening the investigation into Clinton's emails.

In this paper, we analysed two subsamples of participants: those who completed wave 4 (excluding a small number of participants who did not answer all questions, resulting in the total N=2,229) and those who completed each of the waves 2, 3, 4 and 5 (N=1,263).

Analyses. Survey weights were constructed by a raking procedure that matched the sample to national population benchmarks for distributions of age by sex, race/ethnicity, sex by education and household size by income, based on the May 2016 Current Population Survey. An additional weighting variable, reflecting whether participants voted in the 2012 election and whom they voted for, was used to achieve representative proportions of voters for different candidates[39]. The weights were used only for the analyses on N=2,229 individuals who participated in the last wave before the election (results shown in Table 1). The analyses on N=1,263 individuals who participated in all survey waves from August 2016 to after the election were done on unweighted data, because the goal of these analyses was to describe that particular sample and not to make inferences about the overall population.

In line with previous studies using probabilistic questions about voting behaviour[17,18], USC predictions of election outcomes from own-intention (or social-circle) questions were derived by (1) multiplying each participant's own (or social circle's) likelihood to vote by his/her (or his/her social circle's) likelihood to vote for each of the candidates, and (2) estimating the ratio of the resulting variable and the average of the participant's own (or social circle's) likelihood to vote across all participants[38].

GfK election survey. *Question texts.* Own-intention questions: "(1) How likely are you to vote in this upcoming election? (2) If you were to vote in the presidential election that's being held on November 8th, which candidate would you choose?" Response options were Clinton, Trump, Johnson, Stein, another candidate, undecided, and would not vote. The question was slightly modified for those who were certain to vote/had already voted: "Thinking about the presidential election

that's being held on November 8th, for whom will/did you vote?" Likelihood to vote was determined by combining participants' answers to question (1) and questions asking whether participants were registered voters in their state of residence and whether they had voted in previous elections.

Social-circle questions used the same wording as in the USC poll.

Sample. Participants were selected from GfK's national, probability-based online KnowledgePanel, which currently includes 55,000 active members[40]. They were primarily recruited using address-based sampling methods, including telephone follow-up for refusal conversions. Adults who were selected to join KnowledgePanel but did not have access to the Internet were provided with Internet access and a web-based device at no cost. For this study, the KnowledgePanel sample included active panel members who were ≥18 years of age and lived in the United States at the time of the study. Participants were selected using a proprietary probability proportional to size sample algorithm. As a result, the final sample reflected the demographic profile of adults ≥18 years of age based on targets derived from the March 2016 Current Population Survey. The sample was also balanced in respect to party identification (Democrats, Republicans and Independent/others) as measured on an earlier panel profile survey, with target proportions based on the average values obtained from eight different probability-based national polls fielded in the two months prior to this study.

A total of 4,181 members of GfK's KnowledgePanel were included here. The field period was 4 November 2016 (1:30 EST) to 8 November 2016 (11:45 EST). Of those who were invited, 2,367 members completed the survey (a 56.5% study completion rate). As reported by GfK[41], the average panel recruitment rate for participants in this study was 13.0%. Of the recruited households, 62.4% completed the initial profile survey. Together with the study completion rate, a cumulative response rate was 4.6%.

Analyses. Standard geodemographic weights were computed for all participants, regardless of voter registration and likelihood to vote, using iterative proportional fitting or raking. National population benchmarks based on the March 2016 Current Population Survey data were used to create weighting targets based on region, age by sex, education, income and race/ethnicity[41].

Predictions based on own likelihood to vote for different candidates, shown in Table 1 and Fig. 1, were weighted answers to this question for the subset of participants who were likely voters. These were defined as all who indicated being registered voters in the state of their residence, were definitely likely to vote or had already voted, or said they would probably vote and also indicated they always or almost always voted in elections. In all, 1,897 of the respondents were determined to be likely voters (80.1% of all participants). Of those, 1,822 answered both the question about their own likelihood to vote for different candidates and the questions about their social-circle likelihood to vote. Predictions based on the social-circle likelihood to vote for different candidates were obtained as described in the section about USC methodology.

BVA French presidential election poll. *Question texts.* Own intentions: Participants were asked about their voting intentions in the first and the second round of the election using the standard BVA methodology: "During the first round of the presidential election, which candidate are you most likely to vote for? (*Lors du premier tour de l'élection présidentielle, quel serait le candidat pour lequel il y aurait le plus de chance que vous votiez?*)". Response options were the 11 candidates, as well as "I will not vote" (used to infer participation rates) and "I will vote blank". For the second round, participants were asked: "Here is the list of candidates who, according to polls, should include the 2 qualified for the second round. Could you indicate how you would rank each of them? (*Voici la liste des candidats parmi lesquels, d'après les sondages, devraient se trouver les 2 qualifiés du second tour. Pourriez-vous indiquer, dans l'ordre de vos préférences, les candidats*…)." Response options were the four top candidates and 'None of them'.

Social-circle questions for the first round of the election asked: "(1) According to you, what share of your social circle will go vote in the first round of the election? (*A votre avis, quelle est la part de votre entourage qui ira voter au premier tour de l'élection*?) (2) Amongst the members of your social circle who should go vote in the first round of the presidential election, how do you expect their votes to be distributed between the different candidates? (*Parmi les membres de votre entourage qui devrait aller voter au premier tour de l'élection présidentielle, comment devraient se répartir les votes en faveur des différents candidat*?)." The options were Dupont-Aignan, Fillon, Hamon, Le Pen, Macron, Mélenchon, other candidates and voting blank. For the second round, participants were asked: "(1) Suppose that Emmanuel Macron and Marine Le Pen are the candidates in the second round. What will be the share of your social circle that will go vote in the second round? (*Supposons qu'Emmanuel Macron et Marine Le Pen soient les candidats du second tour. Quelle est la part de votre entourage qui ira voter au second tour*?)" and (2) a similar question as in the first round that focused on only Le Pen, Macron and not voting.

Sample. In line with standard French polling practices[42], the sample was selected from the BVA online access panel by quota sampling. The quotas were designed to represent the French population by gender, age, partisan affiliation, employment, region and settlement size, following the guidelines of the French National Statistical Institute. Only registered voters were contacted. The survey took place from 17–22 April 2017 to just before the first round of the election on 23 April 2017. According to BVA, of 1,685 people who satisfied the quota and were invited to participate, 59.5% completed the study, resulting in a final sample of 1,003 participants.

Analyses. Post-stratification weights were used to adjust the sample frequencies to the general population according to gender, age, employment, region and settlement size.

Predictions based on own likelihood to vote, shown in Table 2, were weighted answers of all participants (who were all registered voters, by design) to the questions about their intention to vote for different candidates, to vote blank or to not vote, in the first and second round. Predictions based on social-circle questions were obtained as described in the section about USC methodology above, using answers to questions about the percentage of the social circle who will not vote or will vote blank, and who will vote for different candidates among those social contacts who will vote.

All participants gave informed consent. The research was approved by the USC Institutional Review Board (USC poll), and the Federalwide Assurance Signatory Official of the Santa Fe Institute (all polls).

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Code availability. Stata, SPSS and Matlab codes used for different analyses are available from the corresponding author upon request.

Data availability. The BVA and GFK data are available from the corresponding author upon request. The USC poll data, based on the UAS surveys, can be downloaded from https://uasdata.usc.edu/page/Academic+Papers after registering on the UAS site as a data user.

Received: 18 June 2017; Accepted: 10 January 2018;
Published online: 26 February 2018

References

1. Rothschild, D. M. & Wolfers, J. Forecasting elections: voter intentions versus expectations. *SSRN* https://doi.org/10.2139/ssrn.1884644 (2011).
2. Lewis-Beck, M. S. & Skalaban, A. Citizen forecasting: can voters see into the future? *Br. J. Polit. Sci.* **19**, 146–153 (1989).
3. Graefe, A. Accuracy of vote expectation surveys in forecasting elections. *Public Opin. Q.* **78**, 204–232 (2014).
4. Irwin, G. A. & Van Holsteyn, J. J. M. According to the polls: the influence of opinion polls on expectations. *Public Opin. Q.* **66**, 92–104 (2002).
5. Nisbett, R. E. & Kunda, Z. Perception of social distributions. *J. Pers. Soc. Psychol.* **48**, 297–311 (1985).
6. Dawtry, R. J., Sutton, R. M. & Sibley, C. G. Why wealthier people think people are wealthier, and why it matters: from social sampling to attitudes to redistribution. *Psychol. Sci.* **26**, 1389–1400 (2015).
7. Galesic, M., Olsson, H. & Rieskamp, J. A sampling model of social judgment. Psychol. Rev. (in the press).
8. Galesic, M., Olsson, H. & Rieskamp, J. Social sampling explains apparent biases in judgments of social environments. *Psychol. Sci.* **23**, 1515–1523 (2012).
9. Sudman, S. & Bradburn, N. M. *Asking Questions: A Practical Guide to Questionnaire Design* (Jossey-Bass, San Francisco, CA, 1982).
10. Darling, J. & Kapteyn, A. *Hidden Trump Voters?* (AAPOR, 2017).
11. Huckfeldt, R. R. & Sprague, J. *Citizens, Politics, and Social Communication: Information and Influence in an Election Campaign* (Cambridge Univ. Press, Cambridge, 1995).
12. Sinclair, B. *The Social Citizen: Peer Networks and Political Behavior* (Univ. Chicago Press, Chicago, IL, 2012).
13. Groves, R. *Survey Errors and Survey Costs* (Wiley, Hoboken, NJ, 2004).
14. 2016 election survey. *GfK* http://www.gfk.com/products-a-z/us/public-communications-and-social-science/2016-election-survey (2016).
15. The USC Dornsife/LA Times presidential election "daybreak" poll—understanding America study. *CESRUSC* http://cesrusc.org/election/ (2016).
16. Kapteyn, A., Meijer, E. & Weerman, B. *Methodology of the RAND Continuous 2012 Presidential Election Poll* Working Paper No. WR-961 (RAND Corporation, 2012).
17. Delavande, A. & Manski, C. F. Probabilistic polling and voting in the 2008 presidential election: evidence from the American Life Panel. *Public Opin. Q.* **74**, 433–459 (2010).
18. Gutsche, T. L., Kapteyn, A., Meijer, E. & Weerman, B. The RAND Continuous 2012 Presidential Election Poll. *Public Opin. Q.* **78**, 233–254 (2014).
19. Mosteller, F. & Doob, L. W. *The Pre-Election Polls of 1948* (Social Science Research Council, 1949).

20. Martin, E. A., Traugott, M. W. & Kennedy, C. A review and proposal for a new measure of poll accuracy. *Public Opin. Q.* **69**, 342–369 (2005).
21. Arzheimer, K. & Evans, J. A new multinomial accuracy measure for polling bias. *Polit. Anal.* **22**, 31–44 (2014).
22. Silver, N. Election update: the state of the States. *FiveThirtyEight* https://fivethirtyeight.com/features/election-update-the-state-of-the-states/ (2016).
23. National polls. *FiveThirtyEight* https://projects.fivethirtyeight.com/2016-election-forecast/national-polls/ (2016).
24. Abrigo, M. R. & Love, I. Estimation of panel vector autoregression in stata: a package of programs. *Stata J.* **16**, 778–804 (2016).
25. Holtz-Eakin, D., Newey, W. & Rosen, H. S. Estimating vector autoregressions with panel data. *Econometrica* **56**, 1371–1395 (1988).
26. Dawes, R. Statistical criteria for establishing a truly false consensus effect. *J. Exp. Soc. Psychol.* **25**, 1–17 (1989).
27. Swift, A. Americans' trust in mass media sinks to new low. *Gallup* http://www.gallup.com/poll/195542/americans-trust-mass-media-sinks-new-low.aspx (2016).
28. Duggan, M. & Smith, A. The tone of social media discussions around politics. *Pew Research Center* http://www.pewinternet.org/2016/10/25/the-tone-of-social-media-discussions-around-politics/ (2016).
29. Noelle-Neumann, E. The spiral of silence. *J. Commun.* **24**, 43–51 (1974).
30. Thompson, A. Parallel narratives. *Vice News* https://news.vice.com/story/journalists-and-trump-voters-live-in-separate-online-bubbles-mit-analysis-shows (2016).
31. Tetlock, P. E. & Gardner, D. *Superforecasting: The Art and Science of Prediction* (Random House, New York, NY, 2016).
32. Prelec, D. A Bayesian truth serum for subjective data. *Science* **306**, 462–466 (2004).
33. Prelec, D., Seung, H. S. & McCoy, J. A solution to the single-question crowd wisdom problem. *Nature* **541**, 532–535 (2017).
34. Cahaly, R. Trafalgar Group: most accurate polls in battleground states (FL, PA, MI, NC, OH, CO, GA) and most accurate electoral projection 306 Trump – 232 Clinton. *Trafalgar Group* (11 September 2016); http://us13.campaign-archive1.com/?e=9941a03e2e&u=99839c1f5b2cbb6320408fcb8&id=f7b7326d0b
35. Alabama—2016 election forecast. *FiveThirtyEight* https://projects.fivethirtyeight.com/2016-election-forecast/alabama/ (2016).
36. Opinion polling for the French presidential election, 2017. *Wikipedia* https://en.wikipedia.org/wiki/Opinion_polling_for_the_French_presidential_election,_2017 (2017).
37. Sample and recruitment—understanding America Study. *USC Dornsife* https://uasdata.usc.edu/index.php (2016).
38. Meijer, E. *Sample Selection in the Daybreak Poll* (CESRUSC, 2016); http://cesrusc.org/election/samp-estim01.pdf
39. Meijer, E. *Weighting the Daybreak Poll* (CESRUSC, 2016); http://cesrusc.org/election/weights03.pdf
40. KnowledgePanel. *GfK* http://www.gfk.com/products-a-z/us/knowledgepanel-united-states/ (2016).
41. Thomas, R. K., Barlas, F. M., Weber, A. & McPetrie, L. *Election 2016 Survey—National Election Omnibus* (GfK Custom Research, 2016).
42. Norpoth, H. A win for the pollsters: French election predicted accurately. *The Hill* http://thehill.com/blogs/pundits-blog/international-affairs/330272-after-rough-2016-pollsters-get-it-right-in-french (2017).
43. Cousineau, D. Confidence intervals in within-subject designs: a simpler solution to Loftus and Masson's method. *Tutor. Quant. Methods Psychol.* **1**, 42–45 (2005).

Acknowledgements

The project described in this paper relies in part on data from surveys administered by the UAS, which is maintained by the Center for Economic and Social Research (CESR) at the USC. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of USC or UAS. This project was supported in part by grants from the National Science Foundation (MMS-1560592), the Swedish Foundation for the Humanities and the Social Sciences (Riksbankens Jubileumsfond) Program on Science and Proven Experience. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies. We thank H. Olsson for helpful comments on a preliminary version of the paper and A. Todd for editing the manuscript.

Author contributions

M.G. and W.B.d.B. designed the research questions and the social-circle measures. A.K., J.E.D. and E.M. designed the data collection methods and collected the data within the USC study. M.G. and E.M. analysed the US data. M.D. translated and adjusted the materials for data collection in France and analysed the French data. All authors contributed to the writing of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-018-0302-y.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.G.

nature research

Corresponding author(s): *Double-blind peer review submissions: write DBPR and your manuscript number here instead of author names.*

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Experimental design

1. Sample size

Describe how sample size was determined.

Sample sizes were determined by survey companies as sufficient to provide estimates of election results within an acceptable margin of error (typically +/-3 p.p. at the 95% confidence level)

2. Data exclusions

Describe any data exclusions.

The data were collected only for people who were eligible to vote in the elections described in the paper.

3. Replication

Describe the measures taken to verify the reproducibility of the experimental findings.

We have replicated the results for the 2016 U.S. presidential election, obtained in the USC poll, in a different, GfK poll. We have then replicated the U.S. results in a BVA poll preceding the 2017 French presidential election.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

This study did not include an experimental manipulation.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Blinding was not relevant for the present study because it did not include an experimental manipulation.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ Test values indicating whether an effect is present *Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.*
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	The results were analyzed using software packages Matlab, Stata, and SPSS.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.	No unique materials were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	No eukaryotic cell lines were used.
b. Describe the method of cell line authentication used.	No eukaryotic cell lines were used.
c. Report whether the cell lines were tested for mycoplasma contamination.	No eukaryotic cell lines were used.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No commonly misidentified cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.	No (non-human) animals were used.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	Human research participants represented in all sociodemographic aspects the general population of eligible voters in the U.S. and in France.