



## POLLING IN CONFLICT ZONES — NOT FOR DUMMIES



Dr. Conrad Winn  
Dr. Steve Tatham

Policy Paper n°05  
April 2017

# **POLLING IN CONFLICT ZONES — NOT FOR DUMMIES**

## **Dr. Conrad Winn**

Professor of Political Science at Carleton University, Canada and  
visiting faculty at The NATO School, Oberammergau

## **Dr. Steve Tatham**

Director of Defence SCL Group Ltd and  
visiting faculty at The NATO School Oberammergau

March, 2017

## **ABSTRACT**

Experience reveals that military success needs to be accompanied by intellectual and emotional success for lasting effect—what U.S. doctrine calls “cognitive maneuver.” Public opinion polls or surveys are the main tools for understanding how populations think and for testing communications strategies for changing how they think and feel.

Polls in areas of NATO interest face a double challenge. On the one hand, understanding public opinion is vital in conflict zones. Polls are potentially more important to have in conflict areas than in peaceful societies in the west, where alternative sources of information are widely available about what publics think. On the other hand, polls everywhere are fraught with problems and nowhere more than in conflict areas. In conflict zones, public opinion researchers tend to have fewer skills for eliciting candid opinion while respondents have more reasons to avoiding being candid. This means that polling clients in conflict zones need heightened competence to oversee their pollsters successfully.

This paper is partly a guide to clients. For readers willing to accept that even polls in the west can be problematic, we provide a brief reminder or refresher. For readers who want more background on these challenges, the Appendix provides a more detailed review.

“Seven rules for reducing the risk of bad polling in conflict zones” is the core section of our paper.

## INTRODUCTION

By early 2017, Western military forces will have been involved in Afghanistan for fifteen years and in Iraq for thirteen. These protracted conflicts have provoked much thought; the nature of *victory* has had to be reconsidered. Existing doctrine has been rewritten to cover counter-insurgencies, new doctrine written to embrace new challenges. Greater consideration has been given to new actors who are not direct protagonists but who can influence the outcome.

On the battlefield and at home on TV, it has become difficult to determine who is winning. Physical territory appears to come and go. The strategic Afghan town of Musa Qala's was recaptured from the Taliban in 2007, using new ideas of persuasion<sup>1</sup>, yet lost within the decade. Russian troops were not ostensibly in Donbass in Ukraine while Facebook pages told another story. ISIS claimed an Islamic state in Syria and Iraq. People died in its name in Brussels and Baghdad, Nice and Nasiriyah, Dhaka and Jakarta, Sydney and Sousse, Ottawa and Orlando, Tripoli and Tunis, Copenhagen and Cairo and other locations that may or may not have an obvious link, territorial or otherwise, to Syria or Iraq.

Doctrine and experience tells us that military victory has to be accompanied by a cognitive one. Yet operations have shown that behind the veneer of seemingly firm attitudes can exist opinions that may be unformed, ambivalent, or changeable and that may not predict behaviour.

Commanders and policy makers have turned to polling, survey research, and opinion measurement—largely synonyms for each other. Among western forces in Afghanistan, folklore has had it that the country was the most polled nation ever. If all kinds of consumer, employee, and media surveys and polls were considered, Afghanistan might not merit a footnote in the history of polling. But, it may indeed have been the nation in history that was polled most for military purposes.

In Afghanistan, polls were at least as vulnerable to errors of measurement, interpretation, wishful thinking, and ulterior motives as polls in the west. The Asia Foundation's report on its 2010 Afghan survey made repeated reference to the country's optimism with respondents in their survey attributing their optimism most to an improved security situation. The Foundation's report distilled its findings in the following words:

---

<sup>1</sup> The Art of Influence. <http://www.bbc.co.uk/news/resources/1dt-ff9a9c01-faa4-4038-b4e9-83e619460e1f>

In 2010, 47% of respondents say that the country is moving in the right direction. This figure has been increasing since 2008 (38%) and 2009 (42%). The main reason cited for optimism continues to be the perception of good security, mentioned by 38% of respondents who say the country is moving in the right direction. (p. 3)<sup>2</sup>.

One does not need to know *facts on the ground* to realize that this summary was somehow wrong. Based on responses in the Foundation's survey, a more valid summary might have resembled something like the following:

In 2010, more areas were too dangerous for interviewing (pp. 14-15) and respondents' security concerns jumped in importance (pp. 3, 6, 8, 12). Among respondents perceiving the country as moving in the right direction, fewer attribute this to improvements in security than before (pp. 186-7). The top reasons for perceiving the country as moving in the wrong direction are security-related—inadequate security, bad government, and increased corruption, which jumped in importance (pp. 186-7). Pashtuns and high income earners—both influential—are especially concerned about security (pp. 12, 17, 22). Crime is rising while reporting crime is falling (pp. 30, 33). Half the population is afraid of the Afghan National Police, perceived as less helpful for security than in the past (p. 9) as well as unprofessional (p. 38).

Postponed reporting might have been a valid alternative to a more honest summary than the Foundation provided. The survey's internal contradictions could have been explored with follow up research. The survey reported an astonishing 84% of Afghans believing the Afghan National Police (ANP) to be "honest and fair".<sup>3</sup> This contrasted with reality on the ground and the poll's own data showing the ANP perceived as unprofessional (p. 38). Patrol reports and anecdotal experience showed the Afghans regarding the ANP as corrupt and violent.

---

<sup>2</sup> A Survey of Afghan people. Afghanistan in 2010.  
<https://asiafoundation.org/resources/pdfs/Afghanistanin2010survey.pdf>.

<sup>3</sup> On this point, see Andrew McKay and Steve Tatham, *Behavioural Conflict: Why Understanding People and Their Motives Will Prove Decisive in Future Conflict*.

Without follow up research, it is difficult to know why respondents told the Asia Foundation that the ANP was “honest and fair.” Among the many possibilities are that

- ❑ the question was mistranslated or difficult to understand,
- ❑ interviewers were inexperienced or unfamiliar with the project,<sup>4</sup>
- ❑ Afghans could not imagine a less violent and corrupt police force,
- ❑ Afghans perceived violent and corrupt police as an honest way of operating in their setting at the time, or
- ❑ Abject fear of the ANP led too many to fear saying that the ANP was not honest.

The Foundation could have easily afforded follow up research. The costly sample of approximately 6,500 respondents was several times larger than necessary even for the survey’s goal of comparing ethnic groups and regions. The sample of respondents could well have been smaller with the savings allotted to followup, confirmatory research.

Polls in the west are more efficient, rarely sampling much more than a thousand respondents, but they are nonetheless subject to erroneous results and mistaken conclusions too. In the United Kingdom, publicly released polls predicted wrong outcomes in all three highly consequential, recent votes—the referendum on Scotland’s possible exit from the U.K., the British general election, and the “Brexit” referendum on the U.K.’s future in Europe. The U.K. is not alone in producing wrong polling forecasts.

Many factors contribute to accurate polling, as discussed later in this report. One requirement is to understand the context. That is because context or setting is as important a driver of human behaviour as the “stimulus,” i.e. the phenomenon about which the survey question seemingly asks for an opinion.

Sports behaviour illustrates this well. A man going to a football match is less apt to place money in a donations tin for an animal charity if accompanied by fellow *alpha-male* football supporters than if he is with wife and daughters. Context matters.

---

<sup>4</sup> An astonishing 634 interviewers completed 6,468 interviews. Thus, each interviewer completed an average of about 10 interviews—far too few to build experience and monitor. Source: <http://asiafoundation.org/2010/11/09/asia-foundation-releases-2010-afghan-public-opinion-poll/>, accessed July 20, 2016.

The behaviour of goalkeepers in penalty shootouts also illustrates the relevance of context. Israeli research shows that goalkeepers are almost three times more likely to prevent penalty shot success if they remain stationary on the goal-line.<sup>5</sup> This is a rare choice for goalkeepers. Goalkeepers feel obliged to make wild attempts at saves so that their teammates and fans see the effort. The real context of goalkeepers includes *appearing* to want to stop the ball.

The importance of context goes beyond sport. Many studies<sup>6</sup> show people are prepared to drop rubbish and litter in an environment where there is already rubbish and litter on the ground. One of the ways to reduce littering is to remove the *permissive* environment so that littering is no longer seen as normal. Most people oppose littering in polls even where littering remains widespread. Getting people to actually stop littering requires a change in context.

The key to making good use of opinion measurement is to understand how good polling works and how to use its results to best advantage. This requires the creation of intelligent customers. This paper is designed to assist those who use polling in conflict environments in becoming intelligent customers when commissioning or interpreting polling in those environments.

## OPINIONS—NOT A SIMPLE MATTER

Opinions are a view or judgment that may not be rooted in fact or written in stone. The opinions of individuals or of a large group ("public opinion") can change depending upon evolutions in what people learn, whom they listen to, the circumstances of their lives, or changes in the context in which are asked for one.

### The Real Usefulness of Polls

Polls are more useful for revealing directional changes in beliefs and intentions than in revealing precise behaviour. One-time pollster Michael Barone concludes that "the most important function of polls...is not in telling us who is going to win, but in revealing what is on the voters' minds."\* If readers take nothing more from this paper than Barone's quote, the use of polling in the military and conflict context may improve substantially.

\* See his "Why Political Polls Are So Often Wrong," *Wall Street Journal* (November 11, 2015).

---

<sup>5</sup> Bar-Eli, M., Azar, O.H., Ritov, I. & Keidar-Levin, Y., "Action bias among elite soccer goalkeepers: The case of penalty kicks," *Journal of Economic Psychology* 28 (2007), 606-621.

<sup>6</sup> For example, see *The Florida Litter Study: 1998* Conducted by Florida Center For Solid And Hazardous Waste Management Report #98-9

The importance of public opinion have been widely noted in history. “Public opinion is the thermometer a Monarch should constantly consult,” declared Napoleon Bonaparte. U.S. General turned President Dwight D Eisenhower is on record as stating that “Public opinion wins wars.” Winston Churchill’s more restrained assessment was that: “there is no such thing as public opinion, only published opinion.”<sup>7</sup> In a spirit of skepticism, public opinion expert Herbert Gans distinguishes between opinions reported on by pollsters and those reflected in actual behaviour.<sup>8</sup> People may say one thing and do another.<sup>9</sup>

Political leaders behave as if polls are important to know but not necessarily to follow. They commission a lot of polls but turn to other indicators too. In foreign policy, leaders may behave as if polls do not consistently matter.<sup>10</sup> Babak Bahador concluded in his Kosovo study that “the CNN effect was one of the influencing factors behind NATO’s decision to intervene...”<sup>11</sup> By contrast, vocal opposition to the 2003 Iraq invasion<sup>(12)</sup> failed to deter Prime Ministers Blair and Aznar from joining the military campaign.

The two biggest problems in deciding how to use polls to measure public opinion are whom to poll and how to ask questions. Some people matter more than others—they will act on their convictions or lead others. In 1964, the respected political scientist Philip Converse wrote about “pseudo-opinions.”<sup>13</sup> The polling pioneer Daniel Yankelovich developed what *Time* magazine termed a “mushiness index” to distinguish between firm and unstable or superficial opinions.<sup>14</sup> In the west, pollsters’ difficulties in distinguishing between the

---

<sup>7</sup> Quoted in *Time Magazine* Volume 123 (1984), p. 155

<sup>8</sup> On the distinction between natural public opinion and its synthetic measurement in media polls, see Herbert Gans, “Public opinion polls do not always report public opinion,” *Nieman Labs* (April 29, 2013) at <http://www.niemanlab.org/2013/04/public-opinion-polls-do-not-always-report-public-opinion/>.

<sup>9</sup> See Irwin Deutscher, *What We Say, What We Do: Sentiments and Acts* (Glenview:Scott Foresman, 1973) and his co-authored *Sentiments and Acts* (New York: de Gruyter, 1993).

<sup>10</sup> For cross-national evidence that public opinion as measured in polls had few predictable, electoral consequences in the Iraq war, see Richard Sobel, Peter Furia, and Bethany Barratt, eds., *Public Opinion and International Intervention: Lessons from the Iraq War* (Washington, D.C.: Potomac Books, 2012). See also the review in *Middle East Quarterly* (summer, 2014).

<sup>11</sup> See his influential study, *The CNN effect in Action: How the news media pushed the west towards war in Kosovo* (London: Palgrave MacMillan, 2007).

<sup>12</sup> In the case of Spain estimates have suggested that nearly 90% of the population were against Spain’s participation in the invasion whilst in the UK nearly one million people marched through central London in protest. See Rajeev Syalm, “One Million march against war,” *The Daily Telegraph* (16 Feb., 2004).

<sup>13</sup> Philip E. Converse, “The Nature of Belief Systems in Mass Publics,” in *Ideology and Discontent*, David E. Apter, ed. (New York: Free Press, 1964) pp. 206-61; see also Converse, “Attitudes and Non-Attitudes: Continuation of a Dialogue,” in *The Quantitative Analysis of Social Problems*, Edward R. Tufte, ed. (Reading, MA: Addison-Wesley, 1970) pp. 168-89.

<sup>14</sup> See his *Coming to Public Judgment: Making Democracy Work in a Complex World* (Syracuse: Syracuse University Press, 1991), pp. 34ff.



opinions of all voters in a society and those who actually cast a ballot are behind many recent polling failures, as detailed in the Appendix.

The iconic example of an early study showing that polled attitudes do not necessarily predict behaviour is the 1930's research of Richard LaPiere. For his *Attitudes Versus Actions* study of 1934, which appeared in the journal *Social Forces*, LaPiere criss-crossed America by car with a couple of Chinese ethnicity. They turned away once among 251 hotels and restaurants visited. LaPiere surveyed these businesses with the question, "*Will you accept members of the Chinese race in your establishment?*" The available responses were "Yes", "No", and "*Depends upon the circumstances*". Of the 128 that responded 92 per cent answered "No".

Much research followed. Fishbein and Ajzen<sup>15</sup> reached the conclusion that "...there is little evidence for a systemic relation between [behaviour and attitudes towards an object]"<sup>16</sup>. Milton Rokeach contributed the important understanding that behaviour will be predicted more successfully by gauging people's attitudes to a situation, setting, and context instead of just towards an institution, policy, or leader.<sup>17</sup>

A consensus subsequently developed that attitudes are inadequate predictors of behaviour. As the famous psychologist, D. G. Myers put it, "The original thesis that attitudes determine actions was countered in the 1960s by the antithesis that attitudes determine

### New Discoveries about How People Think

Daniel Kahneman is the only psychologist to have won a Nobel Prize. He co-invented the field of decisionmaking-judgment in psychology,

Kahneman sums up the human brain as a cognitive miser—fast and lazy in reaching conclusions but with a fully engaged, clever brain to justify such conclusions when pressed to.

To illustrate scientific findings about the righteous simplemindedness of the human brain, Kahneman quotes comedian Danny Kaye on a woman he disliked: "Her favorite position is beside herself and her favorite sport is jumping to conclusions."

<sup>15</sup> Fishbein M & Ajzen I, *Belief, Attitude, Intention and Behaviour: An Introduction to Theory & Research*, 1975, <http://people.umass.edu/ajzen/f&a1975.html>

<sup>16</sup> Fishbein M & Ajzen I, *Belief, Attitude, Intention and Behaviour: An Introduction to Theory & Research*, 1975, Addison Wesley, p. 335.

<sup>17</sup> Rokeach made a narrower distinction between behaviour as predicted by measuring attitude to a situation as opposed to attitude toward an object, to use psychological language. See Milton Rokeach and Peter Kliejunas, "Behavior as a function of attitude-toward-object and attitude-toward-situation," *Journal of Personality and Social Psychology*, Vol 22(2), (May 1972), 194-201.

virtually nothing.”<sup>18</sup> However this is not to say that attitudes play no part in behaviours; they do.

The difficulty is that attitudes and behaviour are each affected by situational and social factors that polling often misses.

A related dilemma is of the chicken and egg and which came first. Humans are a rationalizing animal, mobilizing brains mainly to justify opinions that were formed with little thinking.<sup>19</sup> The imperfect predictive value of attitudes may mean that greater success in changing behaviour will be achieved by changing settings instead of changing attitudes.

## **AN IRONY OF POLLING**

The irony of polling for military forces in conflict zones is that it is just like polling within internally peaceful, democratic countries—but dangerously more so. Military forces overseas have limited access to polling expertise and few checks on poll results or their interpretation. They also suffer more serious consequences from decisions guided by polls with wrongful findings or interpretations. Compared to military buyers of opinion surveys in conflict zones, civilian buyers at home benefit from:

- ❑ a larger, more qualified, and more vetted pool of suppliers;
- ❑ a setting where respondents are a lot more comfortable sharing their true feelings and thoughts;
- ❑ information on public attitudes from many sources that they can use to cross-check or validate survey results; and
- ❑ less secrecy and greater transparency, whose existence compels suppliers to be more careful, skilled, and honest at every stage of the research process.

It is a truism in business and economics that geographic clusters of suppliers provide customers with quality and pricing normally unavailable elsewhere. Famous clusters include Silicon Valley, Hollywood entertainment, Italian and Hong Kong fashion, and French fragrances as well as auto industry manufacturers in southern Germany, Michigan/Ontario, and central Japan.

---

<sup>18</sup> *Social Psychology* (New York: McGraw-Hill, 2012) 11th ed., p. 112.. Sometimes the power of a social situation leads us to act contrary to our expressed attitudes.

<sup>19</sup> See Daniel Kahneman, *Thinking, Fast and Slow* (New York: Random House, 2013).

Clients benefit from the innovations and efficiencies made possible by co-location. Neighbouring providers alternate among being collaborators, mutual suppliers, and rivals. Clusters flourish partly because of the proximity of corporate customers, industry auditors, and research labs.<sup>20</sup> Clients can validate suppliers' claims by consulting competitors and third parties. Suppliers develop the skills sought after by their nearby customers and expand as a result.

Polling firms also cluster. Mainly London-based, British pollsters develop the skills needed by media and other nearby clients. U.S. research firms excel at the skills needed for surveys in the world's largest market for research on consumers and organizational behaviour. Survey firms located in central Canada have skills that were developed from generations of national preoccupation with ethno-regional division and the Quebec separatists.

Research firms in distant conflict zones lack the conditions that favour the skills and trust benefitting firms in the North Atlantic region. Yet, research firms in conflict zones actually need far greater skills to do their jobs well.

Polls at home have the advantage of interviewing respondents comfortable with polling. Respondents in North Atlantic countries have minimal concern breach of confidentiality.<sup>21</sup> Election poll respondents do not fret that the governing party might put their name on a watch list, re-assess their taxes, or refuse to collect garbage because they prefer an opposition party. Respondents in a retail or bank survey have no fear of losing credit card privileges because of their answers. Poll respondents would never consider the possibility of violence if they expressed the wrong opinion.

Polls are peripheral to the culture of Afghanistan, Iraq, Iran, Libya, and countless other countries beset by conflict. Respondents in such countries might guess that the polling client is a NATO country and provide an answer to garner favour. Far greater skills are

#### Polling Skills—High Need, Low Availability

Polling firms in conflict zones have fewer of the conditions favouring the developing of research skills than in the west but have far greater need for such skills to predict accurately.

---

<sup>20</sup> As Michael E. Porter puts it, "a cluster of independent and informally linked companies and institutions represents a robust organizational form that offers advantages in efficiency, effectiveness, and flexibility." He also writes that "a cluster allows each member to benefit as if it had greater scale or as if it had joined with others without sacrificing its flexibility." See his classic "Clusters and the New Economics of Competition," *Harvard Business Review* (Nov.-Dec, 1998).

<sup>21</sup> In some countries, it would be a legal offence to do so.

needed in conflict zones to camouflage client identity and gauge authentic attitudes and likely behaviours.

Countries with internal conflict may not be uniformly accessible. Exceptional methodological skills may be necessary to guesstimate the views of unreachable sub-populations.

Civilian clients at home can settle for pollsters with fewer skills because excellent skills are less essential. All segments are reachable even if some may require more expensive to reach than others (eg. mobile phone users). Politicians or retailers might lose market share from wrong decisions based on faulty polls but not their lives or the lives of their loyalists.

Civilian clients at home are also less vulnerable to faulty polls because of other sources of information for cross-checking. They can turn to data on revenue, participation/attendance, customer feedback, letters-to-the-editor, or letters of complaint. Civilian organizations have *big data*<sup>22</sup> as a cross-check or even free polls from the media.

Buyers of polls at home benefit from greater transparency and greater ability to assess their suppliers. In North America, governments will release virtually any poll upon request. Forces in conflict zones treat polls as if they were military secrets with the result that they are subject to essentially no external scrutiny.

The greater transparency of civilian polls at home compels suppliers to be more professional. They are vulnerable to industry association audits and word-of-mouth. Company leaders tend to have graduate degrees from the same universities, even the same professors. They exchange services with each other. Staff and journalists meet at conferences, and talk.

## **THUMBNAIL REVIEW OF LESSONS FROM POLLS IN PEACEFUL WESTERN SETTINGS GOING WRONG**

### **Polls in Civilian Settings at Peace Are Not Uniformly Reliable**

The most famous error in polling is symbolized by a picture of U.S. President-elect Harry Truman laughing as he shows off a poll-driven, newspaper headline on November 3, 1948 wrongly announcing that his Republican opponent had just won.

---

<sup>22</sup> In opinion or marketing research, big data normally refers to relational databases containing information on such matters as purchasing patterns, data from credit card accounts, and other sources that go beyond survey data, that can be used to validate survey data or complement findings from polls. See Bernard Marr, *Big Data* (New York: Wiley, 2015).

Western polls are still imperfect. U.S. polls under-estimated Barack Obama's 2012 prospects and Donald Trump's in 2016. U.K. polls under-estimated support for the Conservatives in the 2015 election and for exodus from the European Community in 2016. The reasons for the failures of polls in western countries and implications for polls in conflict areas are discussed briefly immediately below and at greater length in the Appendix.

### **Coverage Bias**

Coverage bias refers to the physical unreachability of respondents. Together with declining response rates, coverage bias is said to explain forecasting error in media polls in the west. Pollsters themselves point their fingers at cellphones.<sup>23</sup>

Coverage bias cannot explain the overestimation of support for remaining in Europe in the Brexit referendum, where the young (under 25) favoured strongly remaining in Europe.<sup>24</sup> By the logic of coverage bias, the British polls should have over-estimated support for leaving Europe, not the reverse.

As discussed in the Appendix, the problem for western polls is less that they are often wrong and more that it is too difficult to know why. If we cannot explain error, we may not be able to prevent it either.

The mistakes of western polls in western countries have implications for NATO polls in conflict zones. It is vital to eliminate all forms of bias, especially coverage bias, to the extent possible in NATO polls. Beyond this truism, it is a judgment call whether NATO buyers should

- ☐ buy no polls at all,
- ☐ buy fewer polls,
- ☐ secure multiple sources of information to help validate polls results, and/or
- ☐ involve independent polling experts to provide counsel.

---

<sup>23</sup> Cliff Zukin as quoted in the Week Staff, "The Problem with Polls," *The Week* (April 10, 2016) accessed at <http://theweek.com/articles/617109/problem-polls> on June 29, 2016.

<sup>24</sup> See Yougov data as reported in the *New Statesman* (June 23, 2016) accessed at <http://www.newstatesman.com/politics/staggers/2016/06/how-did-different-demographic-groups-vote-eu-referendum> on June 26, 2016.

## **Sampling Bias**

Sampling bias involves failing to reach and interview all categories of people who should be respondents in their proper proportions, for example under-sampling older males in western countries or females in Muslim countries. The problem is that the media and their pollsters do not report if or how they weight data to compensate for improper representation in the survey sample.

Sampling bias has been a cause of U.K. and U.S. polling failure. Historically, British polls under-estimated the Conservative vote because they under-sampled older, middle class makes who would vote Conservative. In the last general election, polls under-sampled Conservatives because Conservative respondents, it turned out, needed more effort than average to be interviewed. Meanwhile, they over-sampled young people, who preferred Labour but did not actually vote. After the 2012 US election, Gallup acknowledged that its failure to predict Obama's election derived from undersampling Democratic voting states.<sup>25</sup>

For NATO polls in conflict regions, the sampling failures of polls in western countries signals the value of more effort to interview randomly selected respondents. Many pollsters carry out corrective weighting privately. Corrective weighting is easy to implement when only one or two segments is under-represented. The hard part is knowing what to weight because turnout (the tendency to cast a ballot) is increasingly diverse among social segments and social segments are can be diverse in their preferences. The British experience is that corrective weighting is less effective than actually finding ways of interviewing segments who are difficult to make contact with.

## **Voter Turnout Bias**

Oversampling youth and mismeasuring the turnout of young voters are different sides of the same coin. U.K. pollsters might have been able to survive their flaw in interviewing so many young voters if they had downweighted sufficiently their influence and hence their support for Labour.

For NATO units in conflict zones, a consequence of this discussion of turnout is that opinion surveys need to go beyond gauging attitudes and preferences and seek to measure

---

<sup>25</sup> Max Blumenthal, "Gallup Poll Reveals 4 Reasons It Got the 2012 Election Wrong," Huffington Post (June 5, 2013), accessed at [http://www.huffingtonpost.com/2013/06/04/gallup-poll-2012\\_n\\_3384882.html](http://www.huffingtonpost.com/2013/06/04/gallup-poll-2012_n_3384882.html) on June 27, 2016.

how the attitudes of respondents on a given subject might lead to potential behaviours. Doing so requires ingenuity in questionnaire design. Good questionnaires

- ❑ make respondents comfortable sharing opinions and attitudes that are risky for them to share while also
- ❑ revealing intended behaviours that might seem inconsistent with the attitudes they express.

### **Declining Response Rates and Adjustment Bias**

Response rates are a known problem but so little research has been done<sup>26</sup> that it is impossible to know their true impact. For decades, pollsters compensated successfully for response rates by weighting. This guaranteed that the demography of the sample reflected the demography of the population from which it was drawn.<sup>27</sup>

Weighting ceased to guarantee accuracy in major part because (a) demographic characteristics declined as predictors of behaviour and (b) little objective information was available about the turnout for different demographic or political segments.

For polls in NATO areas of responsibility, the problem of weighting and adjustment bias may call for two kinds of remedies:

- ❑ examining multiple analyses of survey data using different weighting formulas to estimate how overall results could change, and
- ❑ using multiple sources of evidence to cross-check poll findings.

### **Instrument Bias—Cues and Questions**

In democratic elections, it is accepted that politicians normally answer the questions that they wish they had been asked instead of the ones they were asked. The answers of respondents in polls are even more complex, especially in conflict zones.

---

<sup>26</sup> Perhaps the most important such study was that of Natcen in the U.K. See John Curtice, *The Benefits of Random Sampling Lessons from the 2015 UK General Election* (London: National Centre for Social Research, 2015). See discussion in Appendix, below.

<sup>27</sup> Election day polls are often disastrous. As late as 2004, a Canadian election day poll produced accurate results as a consequence of careful weighting. See <http://www.compas.ca/040629-GlobalTVEDayPollPart2-E.html>. Founder of COMPAS Research, Conrad Winn was involved in that study.

Respondents' answers may reflect many things, for example what

1. respondents truly want,
2. they would want in principle,
3. they can reasonably expect or perceive as inevitable,
4. they think could please the interviewer,
5. they think could please the polling client,
6. would be consistent with the poll's apparent purpose,
7. might provide protection against potentially vengeful, aggressive people or agencies,
8. might be ignited by earlier questions in the questionnaire,
9. might flow from social pressure, or
10. might be triggered by the name of the polling firm.

Respondents' answers might also reflect social pressure, a desire for conformity, or perceptions of inevitability.<sup>28</sup>

Any NATO questionnaire must be constructed with these considerations in mind to lessen the risk of misleading results. An ideal interview experience would match company name and interviews to each region to maximize perceived neutrality. The interview would be constructed to give the impression that NATO is not necessarily the sponsor. Great care must be given to question wording and question order. Ideally, a questionnaire should allow respondents to distinguish between what they ideally want, what they might reasonably desire, or what they likely expect to see.

### **Reality vs. Measurement**

Like other important phenomena, public opinion is inferred from measurement and not observed directly. Democratic leaders are known to make many decisions based on polling data. Submissions to cabinets from domestic departments of democratic governments often include polling reports and recommended communications strategies.

---

<sup>28</sup> On this point: Elizabeth Noelle-Neumann, *The Spiral of Silence* (Chicago: University of Chicago Press, 1984).



Serious problems can arise when pollsters measure opinions that do not matter and fail to measure opinions that do. Accurately measured opinions may not matter if

- ❑ opinions are not held strongly,
- ❑ opinions do not guide behaviour (e.g. western voters who are passionately committed to capital punishment for capital crimes but won't let this guide their vote because of a belief that government is about more than capital punishment),
- ❑ opinions are held by authoritarian people who yearn for leadership from others.

Many distinctions are vital for assessing the authenticity and relevance of the opinions measured in polls.

Sadly, NATO forces and the often inexperienced polling suppliers working in conflict zones may not yet have the skills for assessing the authenticity of the opinions measured in their polls. Conflict zones are not in the best of times or places.

The Asia Foundation's Afghan questionnaire happens to contain many items that are too vague, abstract, or theoretical to allow a confident interpretation of respondent answers. The questionnaire seeks agreement or disagreement with the principle that "Everyone should have equal rights under the law, regardless of their gender, ethnicity, or religion." Without more detailed questions, it is impossible to know to what extent the four-fifths of Afghans who agree with this statement would make exceptions for different ethnic, religious, or sexual groups under various circumstances. Prejudiced people do not see themselves as prejudiced.

## **SEVEN RULES FOR REDUCING THE RISK OF BAD POLLING IN CONFLICT ZONES**

### **Rule 1: Implement Non-Reactivity and Masking**

Any poll anywhere can influence the answers provided by respondents through its perceived client and purposes. This is especially true in conflict zones, where surveys are rare and where respondents have reasons to fear if their true opinions were known. To reassure respondents enough for them to share their genuine opinions requires careful efforts at non-reactivity, an academic term that resembles but is more than "camouflage."

In domestic polls in western countries, polls can influence the answers they get by even before the first question is asked. Respondents' answers can be affected by the name of the polling firm, the identity of the client, the apparent purpose of the project and other aspects of a survey project. It is easy to guess that Scottish accented interviewers phoning Scottish respondents from a hypothetical firm with the term "Scottish" in its name are more apt than others to elicit responses favouring Scotland's exit from the United Kingdom.

#### Non-Reactivity or Camouflage

Design every aspect of a poll to minimize the risk that responses will be influenced by respondents' beliefs about the poll's client and its purposes.

Some polling effects are less obvious. A medical survey measuring aches and pains in general will elicit fewer complaints of digestive difficulties than a survey that focuses on digestive issues in depth. Respondents often adopt the perspective of a research project. In their minds, a study of digestion needs more data on digestive aches and pains than a broader study of aches and pains. By analogous logic, a hypothetical early 2016 survey whose stated purposes involved a focus on financial honesty in politics might well have elicited less support than otherwise for Hillary Clinton than her Democratic opponent, Bernie Sanders, who was more susceptible than Ms. Clinton to criticisms for extreme views than for financial chicanery. A 2016 survey openly sponsored by an organization seemingly committed to gender equality or the advancement of women might have had opposite effects.

In the west, pollsters and their clients have little to worry about with respect to these or other kinds of polling effects. High profile opinion research firms may assign a different, not easily recognized name to their interview operation. In any case, the public is polled often enough that respondents may think relatively little about the underlying purposes of an interview. An election poll could hypothetically be carried out for any candidate, party, lobbyist, advertising company, PR firm, television broadcaster, newspaper, foundation, think tank, academic researcher, or project. Even if polling effects did contaminate results, clients in the west can triangulate findings with other sources of data.

NATO polls require far more thought. Various forms of blindedness should be embedded in the polling project to obscure the identity of the core client and the poll's core purpose to avoid influencing respondents. In double-blindedness, the interviewers do not know the identity of the client or the study's purpose even if the leadership of the research organization does know. In triple-blindedness, the poll is purchased by an intervening third

party; not even the leadership of the polling firm knows the identity of the end client or the project's core purpose.

The questionnaire itself is designed to reinforce non-reactivity. The study's apparent purpose is made seemingly general. Questionnaire content that is at the core of the end client's interest includes questions that could theoretically serve the adversary's interest. These seemingly pro-NATO and pro-adversary questions are themselves surrounded by analogous questions of a more general nature.

The pro-adversary questions discourage respondents from reaching conclusions about the identity of the client. The pro-adversary questions can also help improve the interpretation of answers to the questions of the most direct interest to NATO.

To provide a simple example, suppose a survey wanted to gauge attitudes about corruption in the local regime. Questions about corruption in various types of international agencies, among adversary organizations, and perhaps among foreign countries would help redirect respondents' thoughts about who is sponsoring the poll as well as shed a strong light about how respondents truly feel about corruption and the local regime.

## **Rule 2: Distinguish among Four Separate Types of Root Cause Measures**

A common error in prescribing solutions to local conflicts is a focus on *root causes*, assuming they are knowable and their reversal will end conflict. Identifying true root causes is difficult. Furthermore, efforts to reverse root causes, even true ones, may worsen rather than improve relationships.

Explanation based on objective, scientific understanding requires many studies of large numbers of conflicts of a similar nature. But most conflicts that might attract NATO Forces tend to be unique situations. When situations are few in number, statistical analysis is no longer possible. Identifying root causes becomes a matter of intuitive interpretation based in part on interviewing protagonists. As in democratic politics, it is difficult to distinguish the original motivations from the events or motivations twisted by protagonists for their own advantage.

Assessing root causes is prone to risk, as discussed in the sidebar. A good precaution is to design poll questionnaires to measure four distinct types of variables and perceptions related to root causes:

- A. The true root causes or original attitudes, behaviours, or motivations that led to the behavior NATO wants changed vs.
- B. The claimed, justificatory root causes that may or may not have been true original motivations for the continuing attitudes or behavior NATO wants changed vs.
- C. The current attitudes and rationalizations used to justify the current behavior NATO wants changed vs.
- D. The messages, motivations, incentives, and actions that could change behavior today even if they are totally unrelated to the preceding three types of attitudes.

### Risks of Root Cause Analysis

It is important to be wary of jumping to a conclusion that

- ☐ the root cause was what people say it was;
- ☐ fixing the past will fix the present;
- ☐ repairing a root cause, even a true one, could not worsen relations by inspiring greed or revenge; and
- ☐ greater benefits could not come from messages, actions, or other initiatives unrelated to root causes.

### **Rule 3: Measure the Attitudes of All Respondent Segments, But Cautiously**

It is vital to know the true opinions, frustrations, and likely behaviours of every segment in conflict zones including inhabitants in areas controlled by adversaries. Even if economically motivated to do interviews in danger zones, interviewers should nonetheless be discouraged from doing so. Caring about their welfare builds trust. Discouraging interviewers' risky efforts reduces the chances of fraud involving the submission of the made-up results of interviews that never took place.

Knowing the opinions of people living under the authority of the adversary is nonetheless valuable. Without good measurement, it is challenging to know to what extent these segments have been successfully indoctrinated by the adversary or alienated through protracted exposure to abusive or corrupt conduct.

A partial list of alternatives to sending interviewers into adversary areas would include interviewing:

- ☐ Transient visitors from adversary areas;
- ☐ Residents of border areas, asking about the views of friends, ex-neighbours, and relatives in neighbouring areas controlled by the adversary;

- ❑ Residents of areas once controlled by the adversary, asked to recall their experiences as objectively and fairmindedly as possible; and
- ❑ Residents of safe areas who have tribal, ethnic, occupational, trading, or other relationships with people in adversary areas.

#### **Rule 4: Identify and Prevent Fraud**

The socio-economic setting of conflict zones is conducive to fraud in polling. It is difficult to know the validity of the interviews paid for in the absence of national audit organizations, direct oversight by clients, or the self-monitoring that takes place when many suppliers are clustered together. It is difficult to know if interviews took place, they were conducted among truly randomly chosen people, or if interviews were conducted among paid quasi-professional respondents. Preventing fraud is possible even in the absence of Western-style audits.

#### **Fraud Prevention**

Know and implement the measures that can identify and prevent fraud.

Three preventive measures are to:

- ❑ Require interviewers who use hand-held interview devices (smart phones) to geocode every interview. Genuine interviews have varied geocodes;
- ❑ Embed sleeper questions in every survey. These are questions that are known in advance to be statistically correlated with certain other questions in the study. If the statistical analysis of the data does not produce the expected correlations, fraud is a safe inference.<sup>29</sup> The fraudulent manufacture of responses at the office of the supplier is very unlikely to supply a proper patterning of answers. Doing so successfully would require spending substantial sums on expensive talent with sufficient education and astuteness to make up plausible data; and
- ❑ Compel the supplier to deliver data every 2-4 days over the course of fieldwork. This discourages fraud because the manufacturing of

---

<sup>29</sup> Using this method in a study in a country in South Asia that was not at war, Winn was able to spot fraud early in the process, requiring the supplier to carry out authentic, replacement interviews as a result.

counterfeit responses is much more likely to take place at the end of a fieldwork period instead of over its entire course.

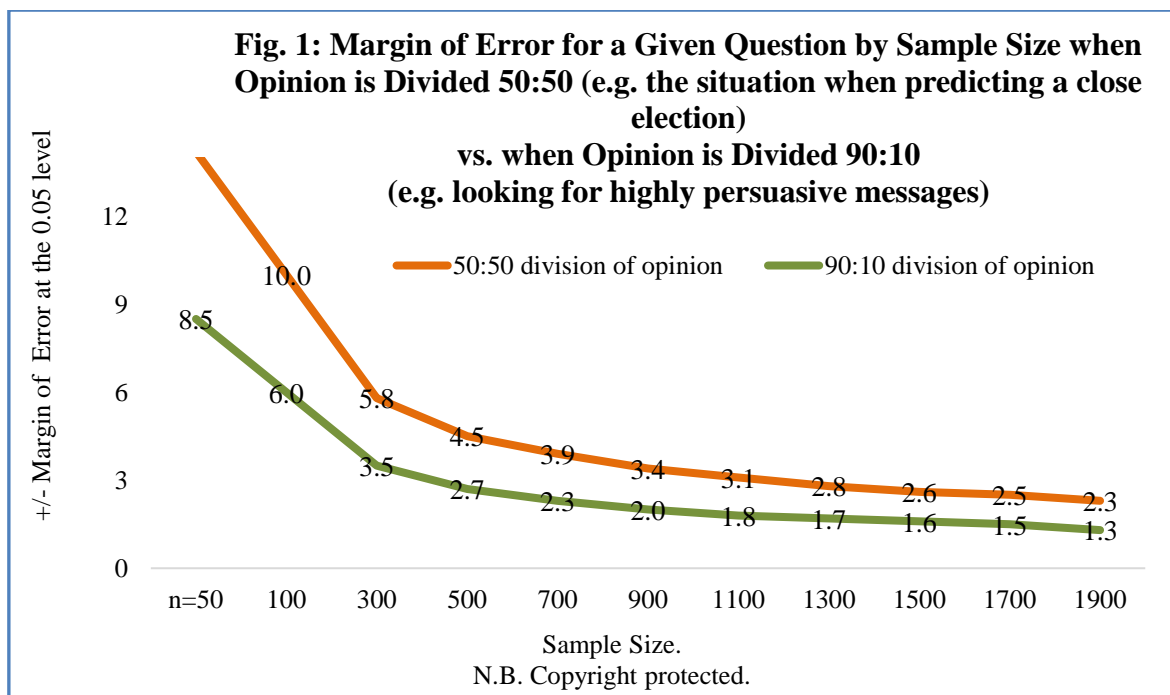
### **Rule 5: Select Sample Size to Suit the Scope of the Study**

The size of the sample should be set by the scope of the study. If the purpose is to predict the outcome of a closely fought referendum or election where ballots are counted honestly, very large samples are needed to be certain of the outcome—perhaps more than 2,000 respondents. In a closely fought election or when opinion is evenly divided, a sample of approximately 2,000 respondents would still have a margin of error of more than  $\pm 2.0$  percentage points, as shown at the end of the red line in fig. 1. That margin of error is adequate when one side leads by, say, 8 percentage points or even less but not in a tight race. In a close contest, estimating the degree of honesty in ballot counting or protagonists' abilities to get supporters to ballot stations may be more valuable than adding apparent precision through larger samples of respondents.

Sample Size Alone Does  
Not Guarantee  
Accuracy

Giant samples that are not vetted for representativeness are less accurate than small ones that are while giant samples that use only abstract, context-less questions are less useful than small samples that provide respondents with context details.

Far smaller samples are normally sufficient to provide a sense of where a nation stands on an issue or how a country would respond to hypothetical messages, reforms, or actions. A sample of 500 is a sweet spot. That is because improvements in sampling accuracy are dramatic between 100 and 500 respondents but improvements level off thereafter, as shown in fig. 1, below.



A sample as small as 100 is often adequate for testing potential messages, reforms, or initiatives. That is because NATO Forces or their host governments should normally embrace potential messages or actions that are almost universally welcome. Suppose a sample survey of 100 respondents shows that the target audience is divided 50:50 in its response to a potential message. Given the margin of error on a sample of 100 is  $\pm 10$  percentage points<sup>30</sup>, one cannot be sure if the true division of opinion among the public is 40:60, 50:50, or 60:40. More precision is not necessary. It matters that the data show the absence of widespread enthusiasm for the potential message. One can be certain that 80 or 90% of the public would not embrace the message.

As respondents increasingly agree on an issue, the margin of error decreases and accuracy improves. When 90% of respondents agree on an issue in a sample of 100, the finding is accurate to within  $\pm 6$  percentage points. This means the true proportion of the public that agrees on this issue is somewhere between 84% and 96% and almost certainly higher than 67 or 77%.

Inexperienced clients may be tempted to purchase giant samples as if giant samples automatically improve accuracy and validity. They only improve accuracy if the sampling process is appropriate and if the questions contain context and are not entirely abstract. A biased or unrepresentative sample of 10,000 respondents is less accurate than a

<sup>30</sup> At the 0.05 level or, in lay language, 19 times out of 20.

representative sample of 100. To the extent that accuracy is essential, engaging an expert third party to oversee the sampling process is money better spent than increasing sample size for un-vetted sampling.

Despite the value of small samples, especially for testing messages, two circumstances require large samples. So long as they are carefully representative, large samples can make sense at the beginning of a communications or messaging program. Large initial samples can make it easier to spot in follow up surveys the initial effects of communications efforts.

Large samples also make sense when there is a plan for detailed comparisons of regions and/or other segments. A rule of thumb is to ensure that 100 or more respondents<sup>31</sup> are interviewed for every cell, segment, or region to be analyzed separately. If one wanted to compare men and women in each of three age groups within each of three regions, one would need to interview at least 1800 respondents.<sup>32</sup> When reporting results for the entire country, each of the regions would need to be weighted to match their true share in the country as a whole. The sample may need to be far larger if the regions and age groups are radically different in size.<sup>33</sup> The sample might even need to be in the thousands if many segments are to be analyzed and they vary greatly in size.

In practice, we have seen few surveys in conflict zones where surveys would not have benefitted from much smaller, hence less expensive, samples. The resources saved from using smaller samples could have been devoted to enhancing the quality of the questionnaire or to a follow up survey asking respondents to explain what they meant by the answers they had provided.

---

<sup>31</sup> The Common Measurements Tool or CMT is the most common standard in customer studies undertaken for OECD governments. It recommends that every cell have sufficient respondents for a margin of error of +/- 10 percentage points at the 0.05 level of significance. In surveys of any target population of any size greater than approximately 5,000, a sample of 100 would satisfy this CMT requirement.

<sup>32</sup> This assumes that the regions and age groups are of equal size and that as many women as men can be readily interviewed.

<sup>33</sup> Imagine a country with 95% of its population in one region and 5% in the other. A national sample would normally need 2,000 respondents to allow inter-regional comparison given the CMT requirement for 100 respondents in the smaller region. If the client organization does not require 2,000 respondents, it could instead over-sample the smaller region, buying 100 interviews in each of the two regions. When reporting on national findings, the responses from the smaller region would be down-weighted to their 5% share of the national population.



## **Rule 6: Follow Good Questionnaire Practices**

Questionnaires matter. As George Gallup, the inventor of modern polling, noted generations ago, large representative surveys provide limited value if the questions are poor. The following are guides to good questionnaire practices:

1. Maximize psychological permission. Design every aspect of the study to make respondents feel comfortable expressing any opinion;
2. Don't assume that preferences, attitudes, values, or opinions necessarily predict behaviour. If behaviour matters, ask about behaviour and not just opinion or preferences;
3. Don't assume that people will behave according to their behavioural preferences. The inhabitants of life and death zones are motivated by survival more than preferences. It may be important to gauge their predictions of winners and losers, government stability, NATO longevity, and related victory-defeat themes;
4. Use number scale questions to the extent possible in place of word response options. Number scale options are faster and cheaper to administer, avoid the problem of words having different meaning to different people, and are better suited to protecting respondent privacy and anonymity;<sup>34</sup>
5. When subjects are delicate, ask the respondent to provide responses on behalf of a hypothetical friend they saw X days previously;
6. When questions are especially delicate, consider scenario questions. In one study of the Free Syrian Army, for example, respondents were asked for their opinions about hypothetical situations in an ostensibly free Syria after the departure of the

---

<sup>34</sup> Anxious respondents are more comfortable with number scale questions. It is easier for them to deny expressing a view of concern to a potential eavesdropper if their opinions are expressed as numbers.

Assad regime. Responses were used to make accurate forecasts of FSA behaviour<sup>35</sup>;

7. When a given theme is important, consider asking analogous questions over the course of the interview so as to gauge stability of opinion and likely behaviour.

### **Rule 7: Design Initial Surveys with the Possibility of Longterm Research Benefits Planned in Advance**

For the short-term, you need to know who believes what, who would be persuaded by what, and related themes.

For the long-term, you need to identify patterns of answers so that you can shorten the questionnaire, saving time and money in future tracking surveys.

## **CONCLUSION**

Polling has come in for significant criticism in the civilian political domain and in the conflict environment. In conflict zones, there is much to be critical of – not least polling companies, which may have prospered doing inadequate jobs, and naive clients who treated polling data as definitive without understanding any of polling's uncertainties.

Commanders have seen data that suggests people support a particular course of action but behave in a different way. We saw this in particular in the case of the so called Perceptions Matrix in Helmand in Afghanistan with commanders often committing troops to task on the basis of flimsy data.

Faced with behaviour that was counter-predicted by polls, Commanders have sometimes concluded that populations are irrational. Better data would normally have shown that populations have almost always acted rationally given their circumstances and the context. Their setting may be more irrational than their response to it.

Blaming surveyed populations for being irrational is like blaming a non-suicidal patient for dying. In the case of the hypothetical patient, the problem is insufficient knowledge of the diagnosis and etiology of disease. In the case of people seeming to behave

---

<sup>35</sup> Winn, unpublished research. Conrad Winn has carried out surveys in conflict zones on several continents for branches of the U.S., Canadian, and U.K. governments.

in ways that contradict their answers in polls, the problem is insufficient knowledge of the complexity of opinion and how to try to measure it.<sup>36</sup>

The repeated failure of polls is a good reason for being wary of its use. Opinion research will retain an allure. Like any other J2 intelligence collection process, opinion research needs to be supervised by clients with expertise and critiqued against other evidence. Staff officers should no more design survey questionnaires than other amateurs. But their increased knowledge will improve the product even if it were only because suppliers felt compelled to hire people with far more training and expertise.

Commanders should be judicious about polling data they pay attention to. Polls are sometimes tools of advocacy with pollsters as seeming *guns for hire*. Buyers should avoid overspending on sample size lest large samples give an unmerited appearance of validity. To reduce their vulnerability to wrong findings and mistaken interpretations, commanders could consider

- ❑ roles for independent or third party experts as auditors and/or evaluators, as well as
- ❑ multiple surveys from different suppliers on the same topic as cross-checks, albeit at far less cost, using far smaller samples.

Finally, we would caution commanders to use polls as just one part of a broader and deeper understanding function that at its heart has a robust social science Target Audience Analysis component.

## **APPENDIX: A DETAILED OVERVIEW OF LESSONS FROM POLLS IN PEACEFUL CIVILIAN SETTINGS**

### **Polls in Civilian Settings at Peace Are Not Uniformly Reliable**

NATO civilians and soldiers may have acquired an excessive confidence in the validity of polling as result of its scientific image and frequency of use in the west. The most famous error in the history of polling is symbolized by an often displayed picture of U.S. President-elect Harry Truman laughing as he shows off a poll-driven, newspaper headline on November 3, 1948 wrongly announcing that he had just won.

---

<sup>36</sup> Steve Tatham teaches the use of social science to gauge behaviour at NATO's Centre of Excellence in Latvia.

Most polls were accurate in subsequent decades, but a near consensus has emerged that election polling cannot be counted on today. In 2012, U.S. polls tended to under-estimate Barack Obama's Presidential prospects and over-estimate Republican Mitt Romney's. In 2014, advocates of Scottish independence were trounced by more than 10 percentage points in a referendum said to be too close to call. In 2015, British polls mis-forecast the general election. In 2016, British polls missed the Brexit referendum while American polls missed the near victory of Bernie Sanders over Hillary Clinton. In the extreme case of the big state of Michigan, polls of Democratic voters showed Clinton winning by an average of 21 percentage points in the March 8<sup>th</sup> state primary, which Sanders won instead.<sup>37</sup>

An analogous pattern emerged in the the June 23, 2016 "Brexit" referendum in the U.K. on membership in the European Community. For decades, telephone polls have been the gold standard because only telephones have been able to sample accurately all voters in western countries. But only two of the eight telephone polls reported in the U.K. media predicted that those who favoured leaving would win. All three telephone polls conducted late enough in the campaign to be almost immune to late shifts of opinion<sup>38</sup> (see table 1) predicted wrongly that those favouring exit from Europe would lose. Poll indications of a victory for staying in Europe were so persuasive that leaving was given only a 12% chance in the massive amount of political betting that took place on the eve of the Referendum.<sup>39</sup>

---

<sup>37</sup> See "Why the polls get it wrong," an op-ed in the *L.A. Times* (March 26, 2016) by Rob Santos, vice president of the American Statistical Association and a past president of the American Association for Public Opinion.

<sup>38</sup> Interviewing began no earlier than the evening of June 20. In the Brexit campaign, telephone polls were less likely to predict a win for leaving Europe than online panel polls, using quasi-professional respondents paid for their participation. During the last two weeks of campaigning, 6 of 10 online polls predicted a win for the side favouring leaving compared to only 2 of six telephone polls. Nonetheless, panels of repeated respondents can never be considered as representative as telephone polls. Furthermore, the social sciences have for years discouraged the repeated involvement of quasi-professional respondents because of the likelihood that their repeated use in polling makes them psychologically less and less representative of genuine voters. In the Brexit campaign, online panel polls offered the advantage of being more likely to predict a victory of leaving Europe. But the online polls also offered the disadvantage of most of the extreme predictions.

<sup>39</sup> BREXIT: Bettor back high turnout and victory for Remain, <https://betting.betfair.com/politics/brexit/eu-referendum-betting-latest-polling-and-odds-june-23-2016-230616-204.html> accessed June 29, 2016.

*Table 1: U.K. Polls Conducted June 20-22 with  
their Predictions for the June 23<sup>rd</sup> Brexit Referendum<sup>40</sup>*

	Predicted winner	Predicted winning margin %	June dates of interviews	Type of survey
IPSOS	Remain	4	20-22	Telephone
YouGov	Remain	2	20-22	Telephone
Opinium	Leave	1	20-22	Online Panel
Survation	Remain	1	20	Telephone

Lack of success in the Brexit referendum followed a poor record leading to the 2015 British election. The U.K. was inundated with polls in this period—approximately 1,900 polls in the 2010-2015 period compared to a total of about 1,600 in the preceding 55 years.<sup>41</sup> In the hours leading up to the 2015 election, a polling consensus was that the outcome was too close to call, and yet the Conservatives won handily.

Faulty polls are not a sufficient reason for NATO to avoid surveys altogether. NATO polls do not normally require quite as much precision as election polls, which only have to be minimally wrong to be incorrect in tight contests. While all three of the late Brexit campaign, telephone polls made wrong predictions, none of them were wrong by much. They were within 3-7 percentage points, not far from their margins of normal sampling error given their sample sizes.

Another reason why forecasting error in election polls may not be immensely important is that knowing precisely how people plan to vote may be less valuable than knowing voters' preoccupations.<sup>42</sup> By this logic, errors in polling forecasts during elections should normally be ignored.

<sup>40</sup> Adapted from BREXIT Poll Tracker The Financial Times. Available online at: <https://ig.ft.com/sites/brexit-polling/>

<sup>41</sup> *The Inquiry into the 2015 pre-election polls: preliminary findings and conclusions* (London: National Centre for Research Methods, Royal Statistical Society, January 19, 2016) accessed at [http://www.ncrm.ac.uk/polling/documents/19Jan\\_slides\\_Final1.pdf](http://www.ncrm.ac.uk/polling/documents/19Jan_slides_Final1.pdf) on June 29, 2016.

<sup>42</sup> The respected commentator and one-time pollster, Michael Barone, writes: "The most important function of polls...is not in telling us who is going to win, but in revealing what is on the voters' minds." See his "Why Political Polls Are So Often Wrong," *Wall Street Journal* (November 11, 2015).

Our view is that polling forecast errors should nonetheless never be ignored. Election and referendum polls are routinely falsifiable. They are tested by reality in a way that surveys of opinions and behavioural intentions cannot normally.<sup>43</sup> Furthermore, the problem is not that individual election polls have usually been far too wrong; it has been that far too many of them have been wrong even if only by small margins in most cases. Finally, the errors common in election and referendum polls reported in the media can provide useful lessons for polls in NATO areas of responsibility.

### **Coverage Bias**

Coverage bias refers to the physical unreachability of respondents even if they are not antagonistic to responding to polls. Coverage bias in western democracies can involve voters who are unreachable or reachable only at a price because they lack a landline or any phone at all. In conflict zones in NATO areas of interest, coverage bias may arise because of risks of violence.

While coverage bias refers to physical barriers to interviewing, sampling bias relates to behavioural barriers to interviewing. Some people are rarely at home, tend not to answer the telephone if at home, or are less likely than others to agree to participate in a survey.

Together with declining response rates, coverage bias is identified routinely as a main reason for the increasing forecasting errors of media polls. Pollsters tend to blame for forecasting failures on "the growth of cellphones and the decline in people willing to answer surveys," in the words of a recent President of the American Association for Public Opinion Research.<sup>44</sup> Cell interviews are more expensive than landline interviews because so many calls have to be made to reach a live person on a cell and to persuade cell respondents to participate.<sup>45</sup>

---

<sup>43</sup> The pioneer of modern public opinion research, George Gallup, saw value in public opinion research but not in election polls, which he undertook only to demonstrate the accuracy of his opinion surveys. See Jill Lepore, "What the turn from polls to data science means for democracy," *The New Yorker* (November 16, 2015).

<sup>44</sup> Cliff Zukin as quoted in the Week Staff, "The Problem with Polls," *The Week* (April 10, 2016) accessed at <http://theweek.com/articles/617109/problem-polls> on June 29, 2016.

<sup>45</sup> Cell interviews are especially expensive in the U.S., where legislation prohibits robo-calls or the use of technology to be involved in the surveying of people on their mobile telephones.

The Washington, D.C. based Pew Research Center incurs the immense extra cost of completing most interviews on mobile phones and hence avoiding coverage bias.<sup>46</sup> Media organizations, which are in financial free fall, cannot afford to do so.<sup>47</sup>

Coverage bias may account for some of the more notorious polling failures but not all. It is plausible to attribute to coverage bias underestimations of support for Bernie Sanders over Hillary Clinton in their fight for the Democratic nomination for President. The young were over-represented among both those who depend only on mobile telephony and those who favoured the more leftwing Sanders.<sup>48</sup> But coverage bias cannot explain the overestimation of support for remaining in Europe in the Brexit referendum, where the young (under 25) were more than twice as likely as the old (65+) to favour remaining in the European Community.<sup>49</sup> By the logic of coverage bias, the British polls should have over-estimated support for leaving Europe instead of under-estimating it. If coverage bias cannot explain the UK Brexit poll failures, coverage bias cannot confidently explain the failure of polls to gauge Sanders' true level of support.

For coverage bias to have been a true factor in the U.S., it had to have operated in the same way in the U.K. unless something else happened. One possibility is that an anti-mobile, anti-youth coverage bias did indeed take place in the U.K. but was obscured by a stronger competing bias that favoured the over-estimation of support for remaining in Europe—perhaps a cuing bias that mis-measured respondents' true opinions, as discussed below.

The problem for western polls is less that they are often wrong and more that it is too difficult to know why so long as they are beset by so many possible biases. If we cannot

---

<sup>46</sup> Cliff Zukin, "What's the Matter With Polling?" *New York Times* (Sunday Review, June 20, 2015). PEW was already completing most of its calls on mobile phones in mid-2015. Professor Zukin is a past president of the American Association for Public Opinion Research.

<sup>47</sup> Most newspapers in western countries are expected to shut down within 15-20 years as subscription income and their volume of paying subscribers continue in free fall with advertising income following suit

<sup>48</sup> According to census-type data from face-to-face, residential interviews for the U.S. government on health issues, Zukin concludes that "a landline-only sample conducted for the 2014 elections would miss about three-fifths of the American public, almost three times as many as it would have missed in 2008." See his "What's the Matter With Polling?" *New York Times*. In the U.S. the autodialing of cell lines is banned under the 1991 Telephone Consumer Protection Act with the result that more expensive live interviewers must be used for the task. See also Jeff Stein, "Why the polls totally underestimated Bernie Sanders in Michigan," *Vox* (March 9, 2016) accessed at <http://www.vox.com/2016/3/9/11186886/bernie-sanders-michigan> on June 27, 2016.

<sup>49</sup> See Yougov data as reported in the *New Statesman* (June 23, 2016) accessed at <http://www.newstatesman.com/politics/staggers/2016/06/how-did-different-demographic-groups-vote-eu-referendum> on June 26, 2016.

estimate why polls may be wrong, we may not have the confidence to estimate when they may be wrong either.

The mistakes of western polls in western countries have implications for NATO polls. A fundamental implication is that it is vital to eliminate all biases, especially coverage bias, to the extent possible in conflict zones. Beyond this truism, it is a judgment call whether NATO buyers should

- ❑ buy no polls at all,
- ❑ buy fewer polls,
- ❑ secure multiple sources of information to help validate whatever polling results become available, and/or
- ❑ involve independent polling experts to provide counsel on managing opinion research suppliers.

### **Sampling Bias**

Sampling bias involves failing to reach and interview all categories of people who should be respondents in their proper proportions in the population, for example undersampling young people or older males in western countries or females in Muslim countries.

Response rates can vary, causing some groups to be undersampled. These biases can be neutralized:

- ❑ in the midst of interviewing—by establishing quotas that match the true distributions of segments in the population; or
- ❑ at the end of interviewing—by algebraic weighting of the sample to upweight the under-represented segment(s) to its (their) proper share.<sup>50</sup>

---

<sup>50</sup> Calculating margin of error requires treating the sample size as if it were a smaller size appropriate to the segment that was upweighted but prior to upweighting. Success in weighting requires that the biases be few in number and the un-interviewed in a category can be assumed to resemble those who were interviewed.



In some instances, polls have become exceptionally accurate as a result of purposeful sample weighting.<sup>51</sup> The problem nonetheless remains that media and their pollsters do not report if or how they do the weighting they may do.

The historic tendency of UK polls to underestimate the Conservative vote was customarily attributed to undersampling of older, middle class males, who tended to vote Conservative more than other segments and were at home less and hence less available for interviews. In 2015, a thorough study by NatCen was carried out over many weeks to get to the bottom of pollsters' underestimation of the Conservative vote. The NatCen study concluded that the polls that year had missed the Conservative lead in the election results because the field work had been rushed. Conservative voters, concluded NatCen, are much more difficult to reach. The Conservative party's substantial lead did not emerge among respondents interviewed in the NatCen survey til at least three attempts and as many as six were made to conduct an interview.<sup>52</sup>

A 55% majority of respondents required 3-6 calls. Among them, Conservatives had a lead of 11 percentage points over Labour while being behind Labour by 6 percentage points among respondents requiring 1-2 calls. Media pollsters had overestimated Labour support because, said the author of the Natcen study, media pollsters do not have the time to undertake so many calls.<sup>53</sup> To that, one may add that media cannot normally afford the cost of such extra calls either.

An equally thorough assessment by the universities-run British Election Study (BES) reached a slightly different conclusion. For the BES team, the key problem is the media pollsters' oversampling of young voters, who tend to be pro-Labour, and undersampling of young non-voters. In the pre-election polls, almost three times more people under 30 claimed an intention to vote in the media polls than had actually voted according to BES' careful

---

<sup>51</sup> This occurred in the increasingly accurate measurement of support for measuring support for separatism in referenda and separatist parties in election in Quebec, Canada, where pollsters began to upweight the highly anti-separatist, older population to its proper share in polls.

<sup>52</sup> John Curtice, *The Benefits of Random Sampling Lessons from the 2015 UK General Election* (London: National Centre for Social Research, 2015).

<sup>53</sup> "Those who were most easily interviewed by BSA interviewers appear to have been more likely to support Labour and less likely to support the Conservatives to a degree that cannot be accounted for by the social profile of these respondents. If indeed this group in any way mimics the kind of person who was most likely to respond to the [media] polls, then we can begin to understand why the polls might have overestimated Labour's strength. Those who are interviewed most easily tend to be distinctive in their political views. Indeed, if we...look at those who were interviewed on either the first or the second call, [our results would have mimicked those of the media polls]." Curtice, *The Benefits of Random Sampling*, p. 16.

post-election interviews (43% vs. 15%). According to a BES spokesperson, “Opinion pollsters are very good at making their samples reflect the general population. But the general population and the electorate are very different things, because around 40% of adults don’t vote.”<sup>54</sup>

The Natcen diagnosis was followed by a report by the National Centre for Research Methods (NCRM).<sup>55</sup> NCRM also blamed sampling. Embracing the Natcen diagnosis, it attributed mistaken forecasts to a combination of poor sampling methods and an over-estimation of the turnout of young voters. Over-sampling of young people was especially problematic because young people were increasingly pro-Labour with older generations increasingly pro-Conservative. A similar dilemma of oversampling youth in 2016 explains why the Brexit polls failed to anticipate a victory for leaving the European community.

The theme of sampling bias normally emerges in discussion only after polling catastrophe. After the 2012 US election, Gallup acknowledged that its failure to predict Obama’s election derived from an undersampling of Democratic voting states.<sup>56</sup> Catastrophic mis-forecasts in the 2016 Michigan primary contest between Clinton and Sanders led to widespread belief that pollsters had under-sampled young and Independent voters.<sup>57</sup> In the lead-up to the 2016 U.S. Presidential election, the *New York Times* led an increasing realization among media that Trump’s prospects had been routinely underestimated because pollsters had been under-sampling older whites.<sup>58</sup>

For NATO polls in conflict regions, the sampling failures of polls in western countries signifies the value of more effort to interview randomly selected respondents. In practice, many pollsters carry out corrective weighting of the samples even if their weighting efforts are not normally reported. Corrective weighting is easier to implement when only one or two segments is under-represented in a given sample.

---

<sup>54</sup> Tom Clark, “New research suggests why general election polls were so inaccurate,” *Guardian* (November 13, 2015), accessed at <http://www.theguardian.com/politics/2015/nov/13/new-research-general-election-polls-inaccurate> on June 29, 2016.

<sup>55</sup> *The Inquiry into the 2015 pre-election polls: preliminary findings and conclusions* (London: National Centre for Research Methods, Royal Statistical Society, January 19, 2016).

<sup>56</sup> Max Blumenthal, “Gallup Poll Reveals 4 Reasons It Got the 2012 Election Wrong,” *Huffington Post* (June 5, 2013), accessed at [http://www.huffingtonpost.com/2013/06/04/gallup-poll-2012\\_n\\_3384882.html](http://www.huffingtonpost.com/2013/06/04/gallup-poll-2012_n_3384882.html) on June 27, 2016.

<sup>57</sup> See Stein, “Why the polls totally underestimated Bernie Sanders in Michigan,” *Vox* (March 9, 2016).

<sup>58</sup> Nate Cohn, “There Are More White Voters Than People Think. That’s Good News for Trump,” *New York Times* (June 9, 2016), accessed at [HTTP://WWW.NYTIMES.COM/2016/06/10/UPSHOT/THERE-ARE-MORE-WHITE-VOTERS-THAN-PEOPLE-THINK-THATS-GOOD-NEWS-FOR-TRUMP.HTML?\\_R=0](http://WWW.NYTIMES.COM/2016/06/10/UPSHOT/THERE-ARE-MORE-WHITE-VOTERS-THAN-PEOPLE-THINK-THATS-GOOD-NEWS-FOR-TRUMP.HTML?_R=0) on June 9, 2016.

Corrective weighting alone is no guarantee. The NatCen study found that their own efforts at corrective weighting were less productive than more call-backs. That is because Conservative voters differed more from other voters in their availability for polling—and perhaps their willingness to be polled—than in their social, occupational, or other demographic characteristics.

### **Voter Turnout Bias**

Oversampling youth and mismeasuring the turnout of young voters are different sides of the same coin. U.K. pollsters might have been able to survive their flaw in interviewing so many young voters if they had downweighted sufficiently their influence and hence their support for Labour because of an expected low turnout. The BES research estimated that 58% of voters 34 years of age and young would cast ballots compared to 83% among those aged 55-64 and 88% among those 65 or older.<sup>59</sup>

Miscalculating turnout does not have to be limited to youth. Over-estimating turnout for Mitt Romney against Barack Obama led to the known over-estimation of Romney's Presidential prospects.<sup>60</sup>

For NATO units contemplating opinion research in conflict zones, a consequence of this discussion of turnout is that NATO surveys need to go beyond gauging attitudes and preferences and seek to measure how the attitudes of respondents on a given subject might lead to potential behaviours.

### **Declining Response Rates and Adjustment Bias**

Like the metaphorical medical doctor who blames his patient for dying, pollsters and their defenders attribute their erroneous results to uncooperative respondents as reflected in falling response rates.<sup>61</sup> However, so little research has been done on response rates<sup>62</sup> that it is impossible to know their true impact. It is accepted wisdom that polls a generation ago were more accurate than today. But response rates had already completed most of their decline from about 80% pre-World War II to under 10% today. So little is known with

---

<sup>59</sup> The Inquiry into the 2015 pre-election polls: preliminary findings and conclusions (London: National Centre for Research Methods, Royal Statistical Society, January 19, 2016), p. 11.

<sup>60</sup> See Blumenthal, "Gallup Poll Reveals 4 Reasons It Got The 2012 Election Wrong."

<sup>61</sup> See Cliff Zukin quoted in the Week Staff, "The Problem with Polls," *The Week* (April 10, 2016) accessed at <http://theweek.com/articles/617109/problem-polls> on June 29, 2016.

<sup>62</sup> Perhaps the most important such study was that of Natcen in the U.K. See John Curtice, *The Benefits of Random Sampling Lessons from the 2015 UK General Election* (London: National Centre for Social Research, 2015).

confidence of the effects of response rates that it would be mistaken to say much more than that high response rates are desirable.

More is known about *differential* response rates but not much more. For decades, pollsters compensated successfully for response rates that diverged among various segments or sub-populations by weighting their results, as discussed above under *Sampling*. Weighting guaranteed that the demography of the sample reflected the demography of the population from which it was drawn.<sup>63</sup>

Weighting ceased to be enough to guarantee accuracy as the following new phenomena took hold:

- ❑ opinions (stated vote intentions) were no longer linked to the demographic characteristics of respondents used to weight samples, the case of the U.K. in the 2015 general election. Hence, ensuring that the demography of the sample matched the demography of the population it was intended to represent no longer guaranteed that the stated intentions of the sample would match the intentions of the broader population; and
- ❑ rates of voter turnout or other behavioural consequences of opinions seemed to diverge a great deal across demographic segments. Hence, even if one could measure the true intentions of the sample being interviewed and even if one could also match perfectly the demography of the sample to the demography of the population, this was no longer good enough because some sectors had become far more behaviourally active than others (e.g. the old with a voter turnout far higher than the young).<sup>64</sup>

In practice, little is known about rates of voter turnout and less still about other behaviours measured in opinion surveys. This is just one more reason why polls are less than entirely reliable for predicting behaviour.

---

<sup>63</sup> Election day polls are often disastrous. As late as 2004, a Canadian election day poll produced accurate results as a consequence of careful weighting. See <http://www.compas.ca/040629-GlobalTVEDayPollPart2-E.html>. One of the co-authors was involved in that study.

<sup>64</sup> The founder of modern polling, George Gallup was prescient in forecasting the turnout problem: ““Such a test is by no means perfect, because a preelection survey must not only measure public opinion in respect to candidates but must also predict just what groups of people will actually take the trouble to cast their ballots.” See Lepore, “What the turn from polls to data science means for democracy.”

Media clients in the west have not required polling suppliers to reveal the weighting formulas actually used to adjust their samples to match national parameters. Without good information on weighting formulas, it is difficult for consumers to estimate how well the data may reflect the opinions of society as a whole. Yet, estimating the opinions of society as a whole is less difficult than estimating the opinions of those who would actually vote.

### **Instrument Bias—Cues and Questions**

In most conversations, honesty is constrained by the speaker's thoughts about how the listener might react. Few conversations are constrained more than the survey or polling interview. Perhaps the most obvious constraint is the listener's sense of obligation to answer a question as it is posed.

In its surveys, the Asia Foundation asked the question, "Generally speaking, do you think things in Afghanistan today are going in the right direction, or do you think they are going in the wrong direction?" Respondents chose one of two options, said they did not know, or said something between the right direction and the wrong direction.

More interesting than the responses may be the questions that the Asia Foundation did not ask. Respondents were not asked and did not volunteer if their answer was based on what they truly wanted, what they could reasonably expect given history, in order to be polite, or on the basis of some other criterion. Respondents were not asked if the question was worth asking; their answers to this hypothetical, secondary question might reveal a lot about their thinking about the primary question. Respondents were not asked who they thought was paying for the survey and if the survey's client was hoping for any particular answer. Finally, respondents were not asked at the outset what a western-sponsored poll should be asking to truly understand people's priorities and concerns.

It is widely understood that a survey's questions can bias or affect the answers by limiting permissible responses. But many other features of a poll can affect the results too. Responses are affected by a poll's perceived purposes and sponsorship. Responses to a given question are affected by the themes of preceding questions (*question order effect*). They are also affected by the name of the firm, not to mention the ethnicity, gender, attire, age, and accent of the interviewer(s). In polling in the west, female interviewers are known to elicit responses more favourable to leftwing politicians and policies favouring the independence

of women.<sup>65</sup> If these various elements of a questionnaire and cues in an interview can affect polling results in the west, they can almost certainly affect polling results in areas under the stress of conflict.

### **Reality vs. Measurement**

Public opinion is like gravity. The existence of both is inferred from their effects but neither has been observed directly by anyone.

Countless players in human history have acknowledged the importance of opinion. Even Soviet Communist party secretary Nikita Khrushchev acknowledged opinion's importance. He insisted that he knew that his country's invasion of Hungary would be disastrous but was compelled to authorize it for fear of public opinion: he would not have been party secretary if he had not. Contemporary Russian leader Vladimir Putin rules thanks to a combination of media manipulation, violence, and a capacity for public opinion measurement<sup>66</sup> that would be the envy of any ruler.

Public opinion and its measurement certainly matters to rulers in genuine democracies. In his studies of U.S. Presidential decision making, Matthew Baum found that public attitudes and public attentiveness were important considerations in all significant decisions on Somalia, not to mention other foreign policy areas, in both the Bush and Clinton administrations.<sup>67</sup>

Polls are commonly used to justify or advocate for a policy. But that use of opinion research is less challenging than knowing what the public truly thinks and how the public will actually act on its apparent opinions. One should probably make distinctions among opinions that

- ❑ are answers provided by a cross-section of society in a survey,
- ❑ are answers provided by the subset of people in a society who are active in politics or have leadership qualities,

---

<sup>65</sup> For the effects of female interviewers in face to face interviews, see Barone, "Why Political Polls Are So Often Wrong." A generation ago, David Northrup of York University's Institute for Behavioural Research conducted experimental research showing that female interviewers elicited responses more favourable to women's issues than did male interviewers.

<sup>66</sup> See "Vladimir Putin's Secret Weapon," *Newsweek Europe* (June 17, 2016).

<sup>67</sup> See Matthew Baum, "How Public Opinion Constrains the Use of Force: The Case of Operation Restore Hope," *Presidential Studies Quarterly* (June, 2004) and his "The Relationships Between Mass Media, Public Opinion, and Foreign Policy: Toward a Theoretical Synthesis," *Annual Review of Political Science* (June, 2008).

- ❑ are answers in polls that reflect opinions that respondents have acted on previously,
- ❑ are answers in polls that reflects opinions that respondents are apt to act on in the future,
- ❑ reflect societal opinions as reflected in past actual behaviour evidenced by patterns of voting, migration, complaints, and other measureable behaviours.<sup>68</sup>

These and other distinctions are vital for assessing the authenticity and relevance of the opinions measured in polls.

Sadly, NATO forces and their imperfect pollsters in conflict ones are largely on their own in assessing the authenticity and relevance of the opinions measured in their polls. Even in the best of times, it is challenging to know to what extent opinions in a poll in any country are authentic rather than expressed because they are socially respectable, held strongly rather than weakly, stable rather than susceptible to persuasion, unchangeable rather than affected by context or setting, actual rather than theoretical, and interpretable in obvious rather than complex ways.

The Asia Foundation's Afghan questionnaire contains many items that are inadequate for truly understanding what respondents are thinking. They are too abstract, undetailed, and lacking in context for their answers to be meaningful. The questionnaire seeks agreement or disagreement with the principle that "Everyone should have equal rights under the law, regardless of their gender, ethnicity, or religion." Without more detailed questions, it is impossible to know to what extent the vast majority of Afghans agreeing with this principle feel it should truly apply to everyone.

---

<sup>68</sup> On the distinction between natural public opinion and its synthetic measurement in media polls, see Herbert Gans, "Public opinion polls do not always report public opinion," *Nieman Labs* (April 29, 2013) at <http://www.niemanlab.org/2013/04/public-opinion-polls-do-not-always-report-public-opinion/>.

# Policy Paper n° 05

## April 2017

The views expressed here are solely those of the authors in their private capacity and do not in any way represent the views of the Latvian National Armed Forces, the Ministry of Defence of Latvia, or any other entity of the Latvian government.

National Defence Academy of Latvia  
Center for Security and Strategic Research  
Ezermalas 8, Riga, Latvia, LV-1014  
Phone: +371 6707-6881 Fax: +371 6707-6888  
<http://www.naa.mil.lv>