(Dis)Assembling Predictive Stability: On the History and Culture of Survey Sampling for Election Forecasts

Lukas Griessl[1]

[1]Department of Sociology, University of Essex

DOI: 10.3217/978-3-85125-932-2-07

Abstract. This essay explores the history of election forecasting alongside the history of survey sampling. In doing so, the following contributes to contemporary scholarship on cultures of prediction, suggesting the notion of predictive stability as a way to conceptualise predictions in social science. In taking an ANT-informed perspective, this essay shows how the development of a stable culture of prediction hinges on the assembling of heterogeneous actors, which stabilisation often takes place in the aftermath of major elections. In order to arrive at this conclusion, the essay will proceed as follows: I will first introduce the topic of cultures of predictions in the social sciences and opinion polling, through which I develop the concept of predictive (in-)stability. After this, I will briefly draw on the history of election forecasting and the history of survey sampling to show that new sampling methods are usually not adopted when their superiority becomes apparent, but when predictive instability of the old ones comes to the fore. In doing so, I will show how the evaluation of pre-election polls informs the way polling is done in general, which in turn, leads to closure regarding the general accepted methodological approaches. This closure is oftentimes reached in the aftermath of major elections.

1 Introduction

How do we know what voters know?[5] Should pre-election polls be 'trusted' as tools of prediction, or feared as mechanisms of distortion, making voters falsely certain of a probable win or loss? Controversies around the metaphysical possibility, as well as the societal role of pre-election polls, have sparked since their first appearance in the early 19th century and have not stopped the progress technologies of election forecasting have made. Especially since the 1950s, social scientists have developed a sophisticated and distinctive methodological apparatus with which to access the social world and to tap into public opinion: The sample survey, therefore, allows to extrapolate statements for a whole population based on limited pieces of it. Particularly in the aftermath of major drawbacks in the 1930s and 1940s, election forecasts based on

[5] I would like to thank the two anonymous reviewers, the organizers, and participants of the conference, in particular of the panel on Cultures of Prediction, as well as Manuel Jung and Pablo Cabrera Alvarez for their helpful thoughts and comments.

opinion polls developed to become a lucrative business. Furthermore, the triumphal procession of election forecasting led to a perception of trust in pollsters, which, however, eroded from time to time. Especially over the past decade, we have seen traditional polls fail in predicting the results of major elections around the world, leading to widespread mistrust in polling data.

The history of election forecasting is intrinsically linked with the history of statistical sampling. When considering both, it is interesting to see how despite consensus regarding the superiority of certain sampling methods, opinion pollsters and election forecasters are generally very slow to adopt those approaches. In particular, we can see that the polling industry mainly adapted new techniques when previous ones fell short in predicting an election, not when their inferiority has been shown theoretically. In order to understand this phenomenon, I will draw on actor-network theory to develop the notion of *predictive (in)stability* to conceptualize how cultures of prediction assemble and resolve. In order to do so, I will first explore the notion of cultures of prediction in relation to social scientific predictions to lay the ground for a subsequent analysis of how such a culture of prediction in the case of election forecasting emerged. I will then discuss the question as to how stable networks emerged and eventually dissolved, for which I draw on the distinction I suggest between opinion polls with predictive content from those without. In a last step, I will develop an understanding of epistemic closure that hinges on the strength of those very networks and their ability to assemble and maintain predictive stability.

2 The Predictive Apparatus

2.1 Some Notes on Actor-Network Theory

The subsequent analysis of the development of the polling apparatus is guided by the methodological tenets of Actor-Network Theory (ANT). Despite being labelled a theory, ANT should rather be understood as a methodological approach to doing social research in heterogeneous settings involving human and non-human entities. Thus, one of the key aspects of it is the overcoming of classical dualisms, such as subject/object, or nature/culture. Instead, heterogeneous actors are assembled in actor-networks, through which practices and associations can gain stability. Such an orientation implies a renunciation of the dichotomy of a knowing subject and an object one wants to further know, instead emphasising the process of mutual knowledge creation. One way to explore how the world becomes represented through statistics and polling points toward the interplay of practices, actors and technologies, a relation that can nicely be illustrated by the concept of ‘statistical chains’, understood as “institutionalized social processes which allow to generate data” (Diaz-Bone and Horvath, 2021: 220).

ANT is therefore interested in how networks transform themselves through the enrolment of new actors or the dissolvement of old ones. One central term is the actant, which Latour uses to account for the role of nature in the make-up of theories in science. ANT refers to the notion as a means to methodologically treat humans and non-humans symmetrical, it is "something that acts or to which activity is granted by others. It implies no special motivation of human individual actors, nor of humans in general. An actant can literally be anything provided it is granted to be the source of an action" (Latour, 1996: 373). The identity of actants is shaped in and through these transformations and the relationships between different actors in the network. Their differences do not hold *a priori* but are effects of actor-networks. Only when such networks become concrete and gain stability, they become actors.

As we shall see later in the essay, the making of knowledge in opinion polling hinges on a variety of actors, which stability is crucial to serve as an apparatus of prediction. When looking at the history of sampling, it will become apparent how different actants, such as newspapers, statisticians, respondents, sampling techniques, landline phones or the public had to enter into particular orderings and relations in order to gain stability. For instance, due to the dissolution of landline phones, traditional sampling techniques partly lost their power as respondents stopped participating, through which the predictive apparatus lost stability. Thus, a new actor, the non-respondent emerged as a threat to the stability of previous ways of polling. This new situation afforded many rearrangements within the predictive apparatus. Following ANTs line of thought that the establishment of scientific facts "comes down […] to placing these actors in a stable network" (Detel, 2001: 14265), I will argue that *predictive stability* is the key dimension for the successful establishment of a culture of prediction in (social) science.

2.2 Prediction in the Social Sciences

All aspects of human life are guided by a sense that we are aware of our limited knowledge of what is yet to come. As social beings, we are oriented toward the future and have a practical interest in gathering knowledge about what will happen. This is often associated with attempts to gather foreknowledge, one of which is to make predictions. Predictions are statements about a future state of the world, for example about the weather or the outcome of an election, to use examples from different domains. Predictions in the natural sciences, however, differ from predictions in the social sciences in important ways, which have to do with the nature of the social and the natural world.

There is a major obstacle to predictions in the social sciences, which anti-naturalist philosophers of (social) sciences have long pointed out: Whereas I can go out to check whether the weather forecast's prediction of a sunny afternoon actually turned out to be correct, predictions in the social sciences contain additional layers of uncertainty,

as the predictive claim can elicit or suppress certain behaviour. As Guala (2015) summarizes, this distinctive feature of human nature has been given various names, such as “self-fulfilling / defeating prophecy” (Merton, 1948), “interactivity” (Hacking, 1999), “reflexivity” (Soros, 2013), or “performativity” (MacKenzie, 2006) and points toward the theory-dependency of social behavior. While the weather doesn’t care about what the weather forecasts say about it, the electorate often cares a great deal about what the current election forecasts say about their future voting behavior. Hacking formulated this as follows: “A cardinal difference between the traditional natural and social sciences is that the classifications employed in the natural sciences are indifferent kinds, while those employed in the social sciences are mostly interactive kinds. The targets of the natural sciences are stationary. Because of looping effects, the targets of the social sciences are on the move” (Hacking, 1999: 108). Due to this lack of stability, social kinds don’t support predictions and explanations in the same way as is the case for natural kinds. If a predicted scenario is undesirable, it can mobilize individuals or groups to engage in behavior to avoid or alleviate its impact, to change one’s voting intention or not vote at all. Similarly, a predicted scenario that is desirable might equally lead to a particular behavior to further support a certain outcome.

As a matter of predicting human behavior, election forecasts are a suitable example of these kinds of claims. Elections are open systems and display a variety of variables that cannot all be modelled. As Northcott (2015) infers from these kinds of considerations regarding the metaphysical possibility of predicting elections, they do not fulfil the metaphysical conditions supposedly necessary for predictive success (see: Northcott, 2015: 1262). Northcott, however, makes the case that methodological considerations rather than metaphysical ones are at the bottom of successful election forecasting. He argues that the crucial element of the successful prediction lies in the ”sophisticate use of case-specific evidence from opinion polling” and not the “pursuit of explanations via general theory or causal mechanisms” (Northcott, 2015: 1260). What is pivotal to successful election forecasts is a certain methodological approach. In the following, I want to follow this line of thought and argue that the success of election forecasts does not lie in the metaphysics of elections, but that it hinges on the development of a stable culture of prediction, involving methodologies, respondents, the media, and others. Therefore, it will be necessary to move one step back to trace the development of a more basic methodological apparatus, namely the sample survey. I will argue that it is precisely due to the difficulties and metaphysical impossibility of predicting election results that a strong and stable culture of prediction is of crucial importance for there to be prediction in the first place.

2.3 Cultures of Prediction

In their seminal work, Heymann et al. (2017) refer to the work of Knorr-Cetina (1999) and state that "like any scientific culture, [cultures of prediction] operate in specific scientific and social contexts and reveal sets of shared knowledge, practices, values, and rules which emerge, stabilize, and shape scientific and public perceptions, conduct, and goals" (Heymann et al., 2017: 6). Following Knorr-Cetina, epistemic cultures are "amalgams of arrangements and mechanisms - bonded through affinity, necessity, and historical coincidence - which, in a given field, make up how we know what we know" (Knorr-Cetina, 1999: 1). Her interest is thereby "not in the construction of knowledge but in the construction of the machineries of knowledge construction" (Knorr-Cetina, 1999: 3). In other words, the question is how the practices and cultures of generating scientific knowledge emerged. When dealing with cultures of prediction, the focus should thus be on the particular arrangements, mechanisms, and practices of coming to know the future.

Drawing on the work of Fine (2007), Heymann et al. (2017) characterized cultures of prediction along five dimensions: (1) the social role of prediction; (2) the character and significance of computational practices; (3) the domestication of uncertainty; (4) the degree of institutionalization and professionalization of predictive expertise; and (5) the cultural impact of predictive practices and claims. Exploring the "constructions of the machineries of knowledge construction" (Knorr-Cetina, 1999: 3) leads our attention to the historical emergence of the instruments and the social practices through which the future is to be brought into the present. Because of its inherent uncertainty, predictions are usually embedded in cultures that serve to stabilize the predictive claims and establish their legitimacy. This allows for the possibility to have shared knowledge about what cannot be known with certainty.

To Fine, the practice of forecasting is "shaped by the contours of group life" (Fine, 2007: 2), rendering it "something akin to art, a personalistic and elusive process of interpretation" (Fine, 2007: 13). Furthermore, "[t]he dark heart of prediction is defining, controlling, and presenting uncertainty as confident knowledge" (Fine, 2007: 103). The question as to what constitutes sufficiently confident knowledge differs across different domains and cannot ultimately be answered: It is "not simply a matter of inventing practices to produce robust and reliable knowledge. It is a matter of conflict, negotiation, and boundary work and is intricately linked to the establishment of social credibility, legitimacy, and authority of scientific claims and policy responses" (Heymann et al., 2017: 27). The legitimacy and credibility of future claims thus have a conflictual nature and are negotiated within scientific communities and the public. Importantly, those cultures of predictions also "represent cultures of power and, hence, transformative forces, which are all the more effective as they are often black-boxed, hidden, and invisible" (Heymann et al., 2017: 7). Taking the perspective of ANT, the negotiation of credibility is successful if cultures of prediction assemble a stable actor-

network. It is about establishing stable actor-networks in which the variety of interests and assumptions harmonize.

Cultures of prediction in the social sciences hinge on what I suggest calling *predictive stability*, referring to the socially shared expectation that predictive claims point toward the right direction. Stability lies in the establishment of networks between sampling instruments, stakeholders, respondents, the public, polling companies and others. Conversely, *predictive instability* refers to the societal shared expectation that predictive claims do not necessarily point toward the right direction. This can be differentiated from the mere capacity to make predictions, which shall be called *predictive capacity*. Predictive capacity alone does not allow any conclusion about the socially shared expectation of getting valid predictions.

2.4 Polling: Two Views

Before expanding on the development of a culture of prediction in election forecasting, I will briefly introduce a differentiation when it comes to opinion polling. Statistics as a discipline is always concerned with estimation, as it aims to estimate features of a population given features of a subsample of it. Election forecasting, however, comes with an additional layer, as it appears to predict an actual event to occur. While this differentiation may not be of relevance to most survey practitioners, as polls and surveys are seen as a method to estimate a characteristic of the population, the public expectation of election polls seems to support such a distinction. Thus, regarding the social attributes toward polling, I will suggest a differentiation between opinion polls without predictive content and opinion polls with predictive content:

- Opinion polls without predictive content: This form of opinion polling is concerned with the creation of knowledge regarding current states of affairs. One might, for instance, try to find out which politician is most liked among the populace.
- Opinion polls with predictive content: This form of opinion polling is concerned with the creation of knowledge regarding future states of affairs. One might, for instance, try to find out whether the majority of a given country will vote for candidate A, rather than for candidate B.

Looking at how, e.g., popular media reports about polling disasters, they are usually concerned with the second class of opinion polls, those with predictive content. The reason for this is that there will be a situation in which the estimation will or will not actually occur. A good track record of a polling company or the polling industry as a whole can thus be understood through the lenses of the concept of *predictive stability,* as predictions would be accompanied by a socially shared expectation that their claims point toward the right direction. In order to make more sense of this, let us consider the development of a culture of prediction in election forecasting.

3 The Emergence of a Culture of Prediction in Election Forecasting

3.1 The Scientificisation of the Oracle

3.1.1 Early Attempts of Election Forecasting

In his study on the origins of election polls, Smith (1990) points our attention to the 1824 US presidential election. This election is generally regarded as a realignment in American politics: While since the 1790s, the US party system was shared by the Federalists and the Democratic-Republicans, the Federalists ceased to be a relevant political power, leaving only the Democratic-Republican Party at the 1824 election. Without having a Federalist opposition, the party split and four of its candidates vied for the presidency. Uncertainty regarding the outcome of this election was particularly high, not only because of the several candidates, but also because they all ran as Democratic-Republicans. This had the effect that past voting behavior could not serve as a guide[6] (cf. Smith, 1990: 23). Not surprising, this confusing situation led politicians, newspapers, and others to attempt to predict the outcome of the election. Smith tells us that, for instance, the number of toasts made to the candidates during the Fourth of July celebration in Pennsylvania was seen as an indication of support toward the respective candidates. Despite drawing on such proxies, people began to conduct, what was later characterized as straw polls during public meetings, such as militia musters or tax gatherings. Those straw polls fulfilled the desire to both know about public opinion and to express one's own opinion. As it seems obvious to a contemporary reader, those straw polls were often highly biased and critiques of their representativity were already raised back then. Those early polls may have fulfilled a societal desire to gain knowledge about the future; the predictive apparatus, was, however, far away from reaching predictive stability.

3.1.2 Learning From Defeat: Assembling Predictive Stability

In subsequent years, a particular procedure evolved out of the variety of approaches that were conducted. Newspapers started to print questionnaires, which readers could fill out and send back. Their address registers were further extended in drawing e.g., on lists of car or phone owners. The Literary Digest plays a particular role here: In 1895, its file already contained more than 350 thousand addresses; by 1932, it had grown to 32 million. To predict the 1928 election result, for instance, the Literary Digest sent out 18 million questionnaires and was able to predict Hoover's victory surprisingly well (cf. Keller, 2001: 33p). Due to their successful track record, the Literary Digest's forecasts were viewed with great confidence and trust. Since the beginning of their polls in 1916, predicting the outcome of presidential elections based on opinion polls was successful five times in total, although the methods were not particularly

[6] Smith (1990) mentions other factors, such as the fact that multiple candidates were running without a party label, and more fundamental change in the political system. The right to vote was extended to all white males and the direct election of electors was introduced.

sophisticated at first. Due to the success of their election forecasts, there was no reason to question their sampling approach, the Digest's numbers were endowed with *predictive stability*. In 1936, however, the highly praised and trusted magazine failed miserably when it wrongly predicted Landon's victory over Roosevelt. We now know that the sample drawn by the Literary Digest was skewed towards wealthy people and not representative of the population.[7] At the same time, George Gallup (1901 - 1984), who was setting up his own polling company, drew a sample of 3,000 people using his own methods and predicted the Digest predictions long before they were published, with an error of only one percentage point. With another purposively drawn sample of about 50,000 people, he correctly predicted Roosevelt's victory, although his prediction of Roosevelt's vote share was also off by quite a bit. The popular and esteemed Literary Digest never recovered from this enormous loss of confidence, which is considered the main cause of the magazine's demise two years later. This episode highlights that methods must be examined not only based on their past track record. As Katz and Cantril (1937) stated back then, "[t]he selective error in the sampling technique of the Literary Digest was logically apparent long before it became empirically important. Merely because a method works fairly well on one or more occasions is no guarantee of its reliability" (Katz and Cantril, 1937: 176).

Twelve years after his brilliant election prediction, however, for the 1948 presidential election, a major polling disaster happened to Gallup himself. Gallup wrongly predicted a victory for Dewey over Truman. Gallup's election prediction enjoyed so much confidence at the time that the Chicago Tribune erroneously headlined "Dewey Defeats Truman" the day after the election. The reason for the misprediction was that the interviewers were allowed to choose whom to interview, given certain quotas. In each of the fixed categories (including gender, age, and economic status), republicans were apparently easier to reach. One of the results of these investigations was a critique of quota sampling (cf. Likert, 1948) and Gallup also drew this lesson from it and subsequently began to use random sampling as the basis for its surveys. Remembering Katz and Cantril's point, reconsidering the sampling technique deemed necessary after errors became empirically, not logically important. There was already definite evidence about the superiority of random sampling, it was already used in official statistics and yet, it was applied in the practice of polling only after *predictive instability* regarding quota sampling became apparent. This event finally led to epistemic closure regarding the best way of how to tap into public opinion. This consensus on the superiority of probabilistic sampling methods became the characteristic element of survey research and the polling industry. Today, opinion polls based on sample surveys are so ubiquitous that it is hard to imagine a world without

[7] The reason for the Digest's incorrect prediction is believed to be coverage bias: respondents were recruited primarily from the telephone directory and a register of car owners, which resulted in more wealthy people being sampled than less wealthy. There was also a low response rate and non-response bias (see for example: Squire, 1988).

them. But also their *predictive stability* has been shattered in the last decade, even though rhetorical closure regarding its theoretical superiority remains intact.

3.1.3 Why so late?

In election forecasts, the voting intention of a subsample of the population serves as a proxy to understand the voting intention of a population and to tap into public opinion. The possibility of making statements about a population based on a sample, i.e., a small part of the population is still relatively young: while sampling was first proposed in the 17th and 18th century by people like John Graunt, William Petty, and Pierre Simon Laplace, it remained largely rejected until it was (again) seriously proposed at the beginning of the 20th century. As I explore in greater detail elsewhere (Griessl, 2022, forthcoming), the 19th century epistemic context was one that saw sampling as speculative and uncertain knowledge, prioritizing full enumeration instead. The central figure in the development of sampling is the Norwegian Anders Nicolai Kiær, who, in 1895, presented his idea of the 'representative method' during the conference of the International Statistical Institute in Bern. His suggestion of a "partial exploration with observations on a large number of scattered localities, distributed over the whole territory so that they form a miniature of that whole" (Kiær in: Kruskal and Mosteller, 1980: 176), was first met with great resistance. In Griessl (2022, forthcoming), I show how this suggestion was followed by a set of rhetorical figures that can be considered as a practice of boundary work (see: Gieryn, 1983; Jasanoff, 1995) by fellow statisticians, who called his approach unserious and dangerous, claimed the terrain of epistemic authority as one that only grants full enumeration the label of science, demarcating it from the apparent unscientific method of sampling. This „communally approved drawing of lines between ‚good' and ‚bad' work (and, not trivially, between good and bad workers) within a single discipline " (Jasanoff, 1995: 53), however, became more and more difficult to maintain. It took until 1925 that Jensen, who reported on the 1925 ISI congress, stated that while in the beginning, the ISI was mainly concerned with the „recognition of the method in principle" (Jensen et al., 1926: 59) there are now hardly any statisticians „who in principle will contest the legitimacy of the representative method" (ebd.). A central figure in this context was Sir Arthur Bowley, who not only defended Kiær's approach, but also further developed it by suggesting randomization as a means to select a sample in contrast to Kiær's purposive selection procedure. At this point, the question within the community was not about whether or not sampling was appropriate, but about which sampling approach to follow. The controversy was about whether randomization or purposive selection was key to solid inference. After the publication of Neyman's (1934) famous paper, randomization eventually became the superior approach to sampling and the controversy reached closure.

3.1.4 Some Notes on Closure

Theories of closure were famously developed by Pinch and Bijker (1984), drawing a picture of technology shaped almost exclusively by social processes. The basic assumption behind this is that stabilized technologies are always the result of long and complex social processes and negotiations. Following this line of thought, both science and technology can be investigated in terms of their socially constructed nature and the social patterns and mechanisms that are effective in the process. This approach became known as SCOT (Social Construction of Technology). Among the forms of closure, Pinch and Bijker (1984) suggest two forms of closure: rhetorical closure, which, in the area of science relates to "some 'crucial' experimental result, 'definitive' proof or 'knockdown' argument which has the effect of closing the debate on some controversial issue" (Pinch and Bijker, 1984: 425). Those results may not convince the scientist from the "Core-Set", but rather the wider community. The second form is closure by redefinition of the problem. In this case, a controversy can be stabilized when the technology in question is being used to solve a different problem. Closure is reached through "redefining the key problem with respect to which the artefact should have the meaning of a solution" (Pinch and Bijker, 1984: 428).

Drawing on those ideas, the consensus reached regarding the superior sampling methods can be called 'rhetorical closure', relating to "some 'crucial' experimental result, 'definitive' proof or 'knockdown' argument which has the effect of closing the debate on some controversial issue" (Pinch and Bijker, 1984: 425). Closure refers to the phase in which a technology or technical artefact acquires a stable identity, in which controversies around how it ought to look like come to an end. In those phases of closure, consensus emerges among scientists and practitioners. Closure, however, does not need to be permanent. The achieved stable identity can turn out to be inadequate for a certain purpose and bring about new developments and reorientations.

When looking at the use of sampling in election forecasts, the story looks slightly different than the history of sampling might suggest. As we have seen in the previous section, the wide adoption of random sampling was not employed after 'rhetorical closure' had been achieved, it happened when the *predictive stability* of other forms of sampling became fragile, when a stable network dissolved. Epistemic closure was reached after the embarrassing miscalls in the course of the 1936 and especially the 1948 elections.

3.2 There is no Election on Public Opinion

Looking at the 1948 election and the controversies it sparked, we can make more sense of the two classes of how polls are perceived, as outlined before: polling without predictive content and polling with predictive content. As already stated, the Gallup poll enjoyed a lot of trust due to its successful track record. Some months before the 1948

election, Gallup stated that public opinion polling methodology became highly successful and reliable.

> "The reliability of methods now employed to gauge public opinion has been demonstrated time and again, not only in the United States but in a dozen different nations. Polls have met successfully the test which any scientific method must meet. They have proved equally reliable when applied in completely different circumstances and by different organizations. […] Modern poll procedures make it possible to conduct a nationwide referendum or plebiscite in a matter of hours, and to report results that would differ by only a few percentage points from the results which would be obtained if the entire voting population of a nation went to the polls. In fact in many situations - particularly those in which a substantial portion of the population fails to take the trouble to vote - the poll results might be even more accurate as a measure of public sentiment than the official returns." (Gallup, 1955: 20–21)

Interestingly, the editors deemed it necessary to add that Gallup made this statement "just a few months before he and his fellow pollsters found themselves under terrific pressure because of their failure to forecast correctly the election of President Truman in November, 1948". In the aftermath of the 1948 election, Gallup made a slightly less confident statement during a symposium on the question of whether public opinion polls should make election forecasts. He stated that polls are always subject to probabilities and also prone to fail, but that pollsters will give their best and continue to improve their methods.

> "With the same certainty that we know we can be right most of the time, we know that we will be wrong some of the time. It has to be that way. We live by the law of probabilities. We will do our best to improve our methods and to do better the next time." (Gallup in: Seymour et al., 1949: 141)

After his statement during the symposium, Gallup was asked a question regarding the purpose of political polling on the background that it wastes a lot of time and energy only in order to see "who is going to get the election when you will know in a few weeks anyway" (in: Seymour et al., 1949: 142). Agreeing with the questioner on the little social value to election forecasting per se, he justified it as a means to test polling methods:

> "It has been my experience that we have had a greater urge, we have done more to perfect our methods, we have actually made greater progress because we knew we were going to have to face an acid test, than we probably would have made if we hadn't been making election forecasts. So I think the only justification of an election forecast is to test polling methods." (Gallup in: Seymour et al., 1949: 143)

What Gallup states here is highly interesting. In the case of, for instance, the famous toothpaste of British men or the public opinion on introducing a speed limit on German motorways, there is nothing out there based on which the truth of such a survey can be measured, except for maybe another survey. In the case of election forecasting, on the other hand, one can say whether or not the polls pointed toward the right direction. This is where the difference between polling with and polling without predictive content becomes relevant. When considering the history of election forecasting and the history of sampling, it becomes apparent that new sampling methods were not widely adopted by pollsters after their superiority has been theoretically shown by survey

methodologists; new sampling methods were rather adopted after shortcomings became evident in the aftermath of election forecasts.
The introduced distinction between polling with predictive content and polling without predictive content comes to the fore here, and we see how the evaluation of the methods of the former can lead to an evaluation of the methods of the latter. Thus, polling with predictive content informs polling without predictive content.

4 Rethinking Closure as Ceremony

4.1 The Formation of a New Network: Enrolling the Non-Respondent

In recent years, maintaining probability (random) based surveys has become more and more difficult and expensive. One of the main reasons for this is that telephone surveys suffer from ever greater non-response, which leads to higher costs and a possible bias in the sample. Figures from the Pew Research Center show that while response rates were still at 36% in 1997, they were at only around 6% in some cases in 2018 and the trend is continuing to decline (Kennedy and Hartig, 2019). This difficulty of capturing public opinion that comes with the dissolution of landline phones and increasing non-response is related to what might be called 'ontological ignorance', a "way of not-knowing that hinges on the particular orderings, the being of the world" (Marquardt, 2016: 3). This is why it has become necessary to increase efforts to reach mobile phones, to use statistical tools to adjust for biases and to tap into public opinion through the internet. The development of the internet brought about new solutions to the problem that there are much more non-respondents than respondents in telephone surveys.
As Bethlehem (2018) shows, conducting surveys online, usually referred to as computer-assisted web interviewing (CAWI), has a history that traces back to the mid-1990s century, when HTML 2.0 became available. Different to HTML 1.0 and E-Mails, this allowed for the transmission of data from the computer of a potential interview partner to the server of the researchers. Whilst E-Mail polls were already experimented with in 1983 (see e.g., Sproull and Kiesler, 1986), the year 1995 and the introduction of HTML 2.0. marks the beginning of web polls and online sampling. Conducting surveys online offers a cheap and fast way to collect large numbers of responses and has thus been very attractive to researchers right from the beginning.
It, however, also evoked strong criticism right from the start, as it turns a basic principle of survey research on its head. The principle is that it must not depend on the individual whether they enter the sample, but that the probability to enter a sample must be the same or known for each person that belongs to the population for which the results are to be generalised. This is the principle that guides polling and survey research at least since the late 1940s when it became apparent that not following this principle can lead to enormous failures. Conducting polls and surveys online does not necessarily mean

that the principle of probability sampling needs to be overthrown. Most online surveys, however, are non-probability based, which means that respondents usually chose to participate themselves, which runs the risk of leading to a highly self-selected sample. This is why adjusting the data in a way that it becomes representative of the target population is a central part of polling as conducted today. In this sense, the shift to non-probability samples not only constitutes a change in the tools through which public opinion is captured; it evokes a different epistemological grid: Whereas in the case of probability sampling, public opinion polling relies on the existence of a list of people from which a sample can be randomly selected (e.g. phone books or randomly generated phone numbers, in the case of online non-probability sampling, no such list exists (there is no list of all E-Mails, nor is it possible to randomly generate E-Mail addresses). This reflects a broader shift in the understanding of what it means to represent public opinion, one from being chosen to enter a sample to choosing to enter a sample. It is through this reordering, through which the non-respondent could be reintegrated into the polling apparatus and new practices of knowledge creation could start to be formed. But as is the case with traditional methods, also the more recent methods are not yet ripe to produce predictive stability.

4.2 No Closure in Sight yet

It is common knowledge that the last decade has seen an upsurge in headline-making mispredictions by pollsters when it comes to high-profile elections and referendums. Most prominent among these are the US elections of 2015 and 2020, as well as the British House of Commons election of 2015 and the Brexit referendum of 2016. These are reminiscent of the miscalls in 1936 and 1948 and may a century from now be treated in the same paradigm-shifting way that Gallup's famous triumphs and failures back then are now treated. Following these mispredictions, many investigations were carried out to explore why such sometimes severe miscalculations occurred. In the case of the 2015 British House of Commons election, for example, Sturgis et al. (2018) concluded that the main reason for the misprediction was a strong bias in the samples. All election forecasts were based on non-probabilistic methods: They concluded that the main reason for the polling miss were "samples which were unrepresentative of the target population's voting intentions" and that "these biases were not mitigated by the statistical adjustments that pollsters applied to the raw data" (Sturgis et al., 2018: 760). In the case of the 2020 US election, however, something different may be observed. Under the title "Revisiting the 'goldstandard' of polling: new methods outperformed traditional ones in 2020", Enns & Rothschild (2020) concluded that non-probabilistic methods performed better than probabilistic ones. They ascertain that the vast majority of election surveys were non-probability samples, but different than in the 2015 British House of Commons election, they could not show that "less expensive, opt-in sample hurt polling accuracy". On the contrary, they show that "[n]on-probability surveys and

surveys combining probability and non-probability methods outperformed probability-based surveys“ (Enns and Rothschild, 2020). If *predictive capacity* was the primary way of achieving closure, probability and non-probability methods seem to be currently roughly on par. The networks in and through which *predictive stability* can be achieved, however, are not yet formed.

4.3 Stabilising Ceremonies

An established network implies closure in the sense that it prevents other actors and relations to enter the network, allowing for the accumulation of scientific knowledge. Considering the history of election forecasting, it can be observed that closure was not achieved through some “‘crucial’ experimental result”, but rather after an approach reached *predictive instability* in the sense of a societal shared expectation that predictive claims do not necessarily point towards the right direction. Historically, those moments led to a reconsideration of used approaches, something that can also be observed in contemporary debates around sampling. Furthermore, one might say that the way how the polling network gets stabilized often runs through particular ceremonial forms of closure. In the immediate aftermath of elections, the question as to how well different polling institutes and approaches fared has become a routinised practice. The evaluation of polls after elections constitutes the precise moments in which links between the media, the public, polling organizations and others are getting stronger or weaker. It is through these processes that the various actors are put in their place and the networks become more and more (de)stabilized.

5 Conclusion

Throughout this essay, I reconstructed the development of a culture of prediction in election forecasting alongside the history of statistical sampling. This presentation was embedded in a broader discussion around cultures of prediction in the social sciences, arguing that it is precisely because of the metaphysical impossibility of predicting future states in open systems like elections that strong and stable cultures form around predictive claims. In this sense, I coined the notion of predictive (in)stability to conceptualize the assembling and disassembling of the heterogenous network behind the making of predictive claims in this field. This notion can become fruitful as a way to think about further aspects surrounding predictive claims in the social sciences. It can help illuminating, for instance, which groups or individuals have the authority to speak about the future, how are those claims maintained despite their intrinsic uncertainty, what practices exist to challenge and weaken predictive claims or who are the groups and individuals who (dis)trust predictive claims. Those kinds of questions surround the field of cultures of prediction and can inform Heymann et al.’s (2017) five dimensions that characterize cultures of predictions, including their social role, the character and significance of computational practices, the domestication of uncertainty, the degree

of institutionalization and professionalization of predictive expertise and the cultural impact of predictive practices and claims.

As has been indicated throughout this essay, predictive claims in election forecasting seem to struggle in assembling stable cultures in recent years. Famous election miscalls, general declining trust in statistics and increasing difficulties to maintain polls and surveys lead to attempts to develop new methods to tap into public opinion and forecast election outcomes. The question of which methods will be able to assemble stable cultures of prediction is open and will depend on many things that need to be put in their place.

6 Bibliography

Bethlehem J (2018) Understanding Public Opinion Polls. CRC Press.

Collin F (2014) Who Are the Agents? Actor Network Theory, Methodological Individualism, and Reduction. In: Rethinking the Individualism-Holism Debate. Cham: Springer International Publishing, pp. 197–217. DOI: 10.1007/978-3-319-05344-8_11.

Diaz-Bone R and Horvath K (2021) Official statistics, big data and civil society. Introducing the approach of “economics of convention” for understanding the rise of new data worlds and their implications. Statistical Journal of the IAOS 37(1): 219–228. DOI: 10.3233/SJI-200733.

Enns PK and Rothschild J (2020) Revisiting the ‘ gold standard ’ of polling : new methods outperformed traditional ones in 2020. medium.com.

Fine GA (2007) Authors of the Storm: Meteorologists and the Culture of Prediction. Chicago: University of Chicago Press.

Gallup G (1955) The Role of the public opinion poll. In: Free Press and Society/Ed. by G. Bird, F. Merwin. New York: Prentice-Hall, p. 20.

Gieryn TF (1983) Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists. American Sociological Review 48(6): 781. DOI: 10.2307/2095325.

Griessl L (2022) Zwischen Konsens und Kontroverse: Grenzarbeit in der Geschichte der Stichprobe (forthcoming).

Guala F (2015) Philosophy of the Social Sciences (ed. P Humphreys). Oxford University Press. DOI: 10.1093/oxfordhb/9780199368815.013.6.

Hacking I (1999) The Social Construction of What?

Heymann M, Gramelsberger G and Mahony M (eds) (2017) Cultures of Prediction in Atmospheric and Climate Science. Routledge. DOI: 10.4324/9781315406282.

Jasanoff S (1995) Science at the Bar: Law, Science, and Technology in America. Cambridge, MA: Harvard University Press.

Jensen A, Saenger K and Bowley AL (1926) Rapport de M. Jensen sur l'Application de la méthode représentative. Bulletin de l'Institut international de statistique 22: 58–61.

Katz D and Cantril H (1937) Public Opinion Polls. Sociometry 1(1/2). [American Sociological Association, Sage Publications, Inc.]: 155–179. DOI: 10.2307/2785264.

Keller F (2001) Archäologie Der Meinungsforschung. Konstanz: UVK Verlagsgesellschaft mbH.

Kennedy C and Hartig H (2019) Response rates in telephone surveys have resumed their decline. Pew Research Center. Available at: https://policycommons.net/artifacts/616911/response-rates-in-telephone-surveys-have-resumed-their-decline/1597639/.

Knorr-Cetina K (1999) Epistemic Cultures : How the Sciences Make Knowledge. Harvard University Press.

Kruskal W and Mosteller F (1980) Representative Sampling, IV: The History of the Concept in Statistics, 1895-1939. International Statistical Review / Revue Internationale de Statistique 48(2): 169. DOI: 10.2307/1403151.

Latour B (1996) On actor-network theory: A few clarifications.: 369–381.

Likert R (1948) Public Opinion Polls. Scientific American 179(6): 7–11.

MacKenzie DA (2006) An Engine, Not a Camera: How Financial Models Shape Markets. Inside technology. Cambridge, Mass: MIT Press.

Marquardt N (2016) Counting the countless: Statistics on homelessness and the spatial ontology of political numbers. Environment and Planning D: Society and Space 34(2). SAGE Publications Ltd: 301–318. DOI: 10.1177/0263775815611421.

Merton RK (1948) The Self-Fulfilling Prophecy. The Antioch Review 8(2). Antioch Review, Inc.: 193–210. DOI: 10.2307/4609267.

Neyman J (1934) On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. Journal of the Royal Statistical Society 97(4): 558–625. DOI: 10.2307/2342192.

Northcott R (2015) Opinion Polling and Election Predictions. Philosophy of Science 82(5): 1260–1271. DOI: 10.1086/683651.

Pinch TJ and Bijker WE (1984) The Social Construction of Facts and Artefacts: or How the Sociology of Science and the Sociology of Technology might

Benefit Each Other. Social Studies of Science 14(3): 399–441. DOI: 10.1177/030631284014003004.

Seymour G, Crossley A, Lazarsfeld PF, et al. (1949) Should Public Opinion Polls Make Election Forecasts?: A Symposium. Journalism Quarterly 26(2). SAGE Publications: 131–144. DOI: 10.1177/107769904902600201.

Smith TW (1990) The First Straw?: A Study of the Origins of Election Polls. Public Opinion Quarterly 54(1): 21. DOI: 10.1086/269181.

Soros G (2013) Fallibility, reflexivity, and the human uncertainty principle. Journal of Economic Methodology 20(4): 309–329. DOI: 10.1080/1350178X.2013.859415.

Sproull L and Kiesler S (1986) Reducing Social Context Cues : Electronic Mail in Organizational Communications Author (s): Lee Sproull and Sara Kiesler Published by : INFORMS Stable URL : http://www.jstor.org/stable/2631506 Accessed : 01-03-2016 17 : 01 UTC Your use of the JSTOR arch. Management Science 32(11): 1492–1512.

Squire P (1988) Why the 1936 Literary Digest Poll Failed. Public Opinion Quarterly 52(1): 125. DOI: 10.1086/269085.

Sturgis P, Kuha J, Baker N, et al. (2018) An assessment of the causes of the errors in the 2015 UK general election opinion polls. Journal of the Royal Statistical Society: Series A (Statistics in Society) 181(3): 757–781. DOI: 10.1111/rssa.12329.