



HAL
open science

How reliable and how valid are expert surveys?

Samy Cohen, Nonna Mayer, Jan Rovny, René Lindstädt

► **To cite this version:**

Samy Cohen, Nonna Mayer, Jan Rovny, René Lindstädt. How reliable and how valid are expert surveys?. 2022, 11 p. hal-04724274

HAL Id: hal-04724274

<https://sciencespo.hal.science/hal-04724274v1>

Submitted on 7 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



Les sciences sociales en question : grandes controverses épistémologiques et méthodologiques

Compte-rendu de la 60^e séance

How reliable and how valid are expert surveys?

20 septembre 2022

Nonna Mayer (Sciences Po, CEE, CNRS) begins by recalling that her friend and colleague Fariba Adelhah is still in prison in Iran and has a thought for her.

Today's session is about expert surveys. Expert surveys have become a popular tool to measure across time and space a wide array of political phenomena, from party positions to the quality of democracy. Expert surveys allow to study a large variety of phenomena. They are quick, they are not expensive, and they allow a general idea on different topics: the position of parties on European integration, the position on the left/right scale, the quality of democracy, and so on. Yet, recently, there was a lot of criticism against these expert surveys. The debate is double: one is about the *reliability* of the measure, and the other is about its *validity*. Are the results accurate? Is it possible to rely on experts' judgments and what exactly do such surveys measure? To

tackle these issues, two specialists are present. Jan Rovny¹ (Sciences Po, CEE & LIEPP) draws from his experience as one of the principal investigators of the Chapel Hill Expert Survey on party positioning to reassess the validity/reliability of such instruments. René Lindstädt² (Head of the School of government, University of Birmingham), discusses the nature of the measurement problems such data raise.

Jan Rovny's presentation

Jan Rovny thanks Nonna Mayer for the general introduction. First, he wants to explain broadly what it is like to run an expert survey. Broadly speaking, an expert survey consists in asking a lot of questions to a lot of people over many countries. It is usually a really large project. Obviously, selecting experts is not easy. Actually, the most important thing is the expert themselves, meaning their ability and skills. Besides, these experts must be willing to spend their time on such surveys. Jan Rovny will start by describing expert surveys in general, before focusing on the "Chapel Hill Expert Survey".

What are the strengths of expert surveys? First, Jan Rovny explains that at the core of the project lies something very intuitive: if one wants to know something, one has to ask experts on the subject, but shouldn't trust just one person, and should ask many people. If many people agree on a point, they are probably close to the truth. Then the degree to which they agree will tell more about the certainty of what is studied. What he and his colleagues are doing is direct quantification. Usually, they conduct evaluations with the same structured scales and they generally prefer not to change. In addition, they try to have as much flexibility as possible by being able to provide information on any topic on which expertise exists by using information on any topic with deep expertise.

¹ Jan Rovny is co-author of "Explaining the salience of anti-elitism and reducing political corruption for political parties in Europe with the 2014 Chapel Hill Expert Survey data", *Research & Politics*, January 2017 (with J. Polk, R. Bakker et al.) and "Measuring party positions in Europe: The Chapel Hill expert survey trend file, 1999-2010", *Party Politics*, 2015, 21(1):143-152 (with R. Bakker, C. de Vries, E. Edwards and al.).

² René Lindstädt (School of government, University of Birmingham) is co-author of "When Experts Disagree: Response Aggregation and Its Consequences in Expert Surveys", *Political Science Research and Methods*, 2020, 8(3):580-588 (with S-O. Proksch et J.B. Slapin).

Obviously, their validity relies on a diversity of sources, but there are many potential shortcomings with expert surveys. Jan Rovny mentions some of them. They may measure different attributes, for example. Besides, experts may measure the same attribute but at a different time or at a different place. Experts may measure rhetoric instead of actual behavior. Also, information asymmetry can happen because experts have different level of information. An obvious bias is that experts don't always know the same thing. Another concern is the problem of retrospective judgment, it may be hard for experts to reconstruct the past.

Jan Rovny then focuses on the Chapel Hill expert survey in particular. He explains that the "Chapel Hill Expert Survey on Party Positioning" is the most complete expert survey on European Political parties that exists, at least based on expert's judgement. The Chapel Hill expert surveys estimate party positioning on ideology and policy issues, and international relations, for national parties in numerous countries across the world. The survey covers six waves: 1999, 2002, 2006, 2010, 2014, and 2019. The number of countries increased from 14 Western European countries in 1999 to 24 current or prospective EU members in 2006 and 32 countries in 2019. Questions on parties' general position on European integration, several EU policies, general left/right, economic left/right, and social left/right are common to all CHES-Europe surveys. More recent surveys also contain questions on non-EU policy issues, such as immigration, redistribution, decentralization, and environmental policy. Jan Rovny also suggests that he and his colleagues are currently trying to innovate with a new form of survey called Speed CHES. These surveys aim to ask a smaller number of questions. They have done the first one during Covid (2020). They are now planning another flash short survey on Ukraine (2022). The later survey was in 2019. There were 32 countries (EU27, UK, Norway, Switzerland, Turkey, Iceland) and 6 candidates' countries (BiH, NM, SR, AL, KO, MN) studied. Regarding this study, the worst response rate was France (and this is actually the reason why Jan Rovny came to France six years ago). The best response rate was Kosovo. Thus, there are countries in which few responses are obtained, and vice versa. The study was answered by 530 experts and covered 335 parties. They tried to focus on the academic field, so the expert were at least PhD candidates and ideally faculty members who have a strong expertise on political competition or political parties. In some cases, they had to open to wider fields, such as journalists who were covering political events. The team who ran the study was

mainly composed by Ryan Bakker, Liesbet Hooghe, Seth Jolly, Gary Marks, Jonathan Polk, Jan Rovny, Marco Steenbergen, and Milada Anna Vachudova.

Reliability is, of course, an aspect that must be questioned. Reliability measures are supposed to give always the same value for the same attribute. If one measures the same thing over and over, one must obtain the same attribute. The way to measure reliability is to consider the dispersion of measurement across experts. If Jan Rovny and his colleagues see a large amount of disagreement amongst experts, it means that either there is a problem with the survey, or there is something specific with the particular object of study. One of the things that has been noticed is that experts judgments tend to have lower standard deviations (from the average judgment) in situation where parties are more differentiated. Conversely experts tend to be less certain in situation where parties have internal divisions. These situations involve more blurring. They would thus produce a more widely ranging estimate of the political parties position. Thus, expert uncertainty can be understood as reflection of positional ambiguity.

Validity cannot be measured directly, considers Jan Rovny, because no one really knows what the truth is. Validity can only be assessed through what he calls “triangulation”. What they try to do is simply to compare their placements of experts with other sources. It is called cross-validating expert surveys. The other sources can be for example partis’ manifesto data. Alternatively, they also can consider how voters place political parties. Jan Rovny then tries to explain the placement differences. For example, experts tend to place right-wing parties more to the right than do individuals, while they tend to place left-wing parties more to the left. Experts thus tend to see political parties as more extreme, more polarized than citizens do. Jan Rovny is not exactly sure about how to explain this difference. He considers that it an interesting finding, nonetheless, and that it should be tested with more precise instruments than what he was able to do.

Another aspect that he finds interesting is the question of cross-national comparability. Is it possible to compare countries? Doesn’t the left-right dimension have a very different meaning in Sweden and in Latvia, for example? This question was asked in the reviews to which they were submitting their paper. Here is another example of a

real reviewer comment they received: “I guess that I would simply want to hear a bit more on why this data is the appropriate source of information, especially as this is based on very subjective opinions by ‘experts’ where varying institutional/political/ideological pressures and contexts are actually present in their answers, but the data is still analyzed as if they were not.” Jan Rovny thinks this is a very fair point. Yes, he suggests, contexts are different. What they have come up with to show it is an experiment called “anchoring vignettes”. The experts are given information about imaginary parties. For example: “*Party A* views the provision of a social safety net as important but believes there is a sharp trade-off between welfare spending and economic competitiveness. It favors limiting government regulation to instances of market failure, and prefers cuts in social spending over new taxes to meet rising social needs”; “*Party B* views the equalization of life chances for all citizens as an important goal of government. It favors active government in regulating domestic and international markets, and supports steeply progressive taxes to fund redistributive social programs”; “*Party C* believes in small government. It favors minimal regulation of domestic and international markets, supports the privatization of many government operations, and opposes high taxes to fund redistributive social programs.” Obviously, these three vignettes serve to give a general view on the political parties. They are deliberately broad and imprecise. Then they would ask the following question: “On a 0-10 scale with 0 being extreme left and 10 being extreme right, where would you place this party?”. Here, $B < A < C$ was the expected ordering, meaning that B should be seen more to the left, A central and C more to the right. So, did the expert get it right? Yes, since it turned out that $B < A < C$ (the expected order) had a proportion of 72.9%. Jan Rovny then shows that there are strong variations the by country in the placement of the 3 vignettes parties. The experiment gives a sense of how the political differs amongst countries.

An interesting question is: who are the experts? The experts were sometimes a little reluctant to answer questions about themselves. Jan Rovny explains that they were sometimes a little shy about asking more questions to the experts about themselves after having already asked them so many political questions. They tried to keep it to a minimum (their age, their gender, the time of their dissertation defense, etc.). While analyzing the answers, it appears that there is a gender problem, because there are twice as many male experts as female experts. Another thing that must be noticed is

that the modal year of birth is 1980. They have increasingly younger experts. Unsurprisingly, the expert pool is dominantly from academia, so they have an over-representation of people who identify as economically left-wing, and there is also a majority of experts who identify themselves as socially liberal. Still, they do have an interesting number of people who identify as culturally conservative, in Denmark in particular. Jan Rovny wouldn't have expected that and has no explanation in mind. Less surprisingly, they also have right-wing experts in Latvia, Bulgaria and in other central eastern European countries. All in all, broadly speaking, the extreme majority of experts are social liberals in all countries.

Jan Rovny concludes on different aspects related to the reliability and validity of expert surveys. Expert surveys are just one way of assessing party placements. The approach is intuitive: if one wants to know something, one should ask the people who are experts in the field. There should be as many people questioned as possible. If they all agree, then the result tends to be as valid as possible. The degree of reliability is therefore related to the degree of validity. Yet, it is true that expert surveys suffer from some systematic bias. But the good thing is that since these biases are systematic, they can be measured and analyzed. Finally, he considers that expert surveys are a useful complement to other sources of information.

René Lindstädt's discussion of Jan Rovny's presentation

René Lindstädt thanks Jan Rovny for his presentation and also for the work he and his colleagues have done on this subject. He says he is not here to criticize Jan Rovny's work, which he finds admirable, but simply to question how people are using these surveys. Why are we using experts survey for? Why should we use expert surveys rather than other types of technique? What is the real purpose of these expert surveys? He mentions that these types of studies are also used in other fields, namely the medical field. The whole point of medical surveys would be medical analysis. It is to question whether the patient's life is endangered or not. In social sciences, the questions that are posed are slightly different and more complicated. Maybe the researchers are expecting too much with expert surveys. The Comparative Manifesto data, in particular, has known many misuses. Essentially, manifestos are what parties are writing about themselves to inform voters. The question the manifesto project asks

is: what parties are aspiring to look like in the eyes of the voters? That is a huge difference with the sort of question expert surveys ask, which is more: what kind of policy political parties actually enact? This is why Jan Rovny and his cohorts has made a huge impact, compared to the Manifesto project.

Also, he disagrees with the idea that these studies are inexpensive. René Lindstädt raises the question of the right length of surveys. A balance must be found between surveys that are too long (that would become too costly) and too short (where the experts might not be so expert). Obviously choosing experts very carefully for their quality of expertise takes time, and costs money. René Lindstädt returns to the analogy with medical surveys. He explains that medical surveys are sometimes done with a smaller number of experts. It allows the experts to meet with each other. In the case of large online surveys, the experts do not know each other, they do not meet. But again, even if the possibility of meeting each other has a lot of advantages, it takes a lot of time and it also costs a lot of money. On the other hand, perhaps it is better to have fewer experts in terms of numbers, but people who are truly global specialists in their field, and truly committed. But how do you know who is really an expert? It's not that easy.

Finally, René Lindstädt also wonders about the difference between the experts contacted who agree to answer and those refusing. He wonders if there is a way to know if those who respond and those who refuse to respond are fundamentally different. René Lindstädt concludes by saying that his remarks are in no way critical, since he admires the work done by Jan Rovny and his colleagues. His remarks aim to stimulate further discussion.

Jan Rovny's response to René Lindstädt

Jan Rovny thanks René Lindstädt for his questions and remarks. The key concern, he suggests, is the creation of the left-right scale. Some things are better known than others (the left of the political spectrum, for example, compared to the right). Jan Rovny wondered whether the new type of surveys, Speed-CHES, might not overcome some of the shortcomings pointed out by René Lindstädt. They contain fewer questions. For

example, in the case of the Covid survey, there were only five questions. This reduces the risk of misunderstanding.

To measure the degree of uncertainty of the experts, Jan Rovny explains that at one point they had envisioned a kind of “radio button” that the experts could turn according to their degree of uncertainty. But the survey software (Qualtrics in this case) they use does not have this type of measurement tool.

To select experts, they are mainly interested in their publications. They ask who they are, what they have written, where, with whom, etc. Sometimes they are not very selective when they need more experts, so they open up to more experts or less qualified ones. Afterwards, they are also interested in the content of the experts’ answers. If they realize that an expert is likely to have answered without reading (by putting “5” to each question, for example), they remove this expert from the study. Jan Rovny explains that more and more young researchers in Central and Eastern Europe are trained in this type of survey, so they become very qualified. This is a good thing, but one of the risks is to create an “epistemic community” that is a little too closed in on itself.

Nonna Mayer’s question to Jan Rovny

Nonna Mayer wants to know the proportion of experts contacted who respond versus those who refuse to respond.

Jan Rovny response to Nonna Mayer

Jan Rovny explains that this varies greatly from country to country. In France, only 20 to 30% of the experts contacted accept to answer. Unfortunately, he and his colleagues don’t really have time to do more detailed research on the profile of non-respondents. Sometimes, experts don’t respond simply because they have passed away; sometimes they have changed institutions and their email is not updated, etc. Sometimes they receive hostile emails from people who tell them “to get lost”. That being said, they are receiving fewer and fewer mean emails. This is probably due to the fact that their work is more and more known.

Julien Argoud's question to Jan Rovny

Julien Argoud (Sciences Po, CERI) questions Jan Rovny about the classification of political parties. Political parties can be classified in terms of their views on democratic freedom. Libertarian or post-materialistic parties favor expanded personal freedom. Julien Argoud wonders if the labeling of political parties as “traditionalist” versus “authoritarian” has an impact on the expert responses. Often, a political party will accept more easily the label “traditionalist” than the label “authoritarian”, because “authoritarian” is more negatively connoted. For example, if you ask Poutine or Bolsonaro if they are “traditionalist”, they will probably answer “yes”. But if you ask them if they are “authoritarian”, they will probably say “no”. Does it have an impact?

Omer Faruk Metin's question to Jan Rovny

Omer Faruk Metin (Sciences Po, CEVIPOF) wonders if Jan Rovny has considered using country differentiation criteria. The left/right political division may not be the same in different European countries.

Jan Rovny's response to Julien Argoud and Omer Faruk Metin

Jan Rovny agrees that the political division between countries is complex and multidimensional. They spent many hours questioning those questions. For example, it is possible to discuss the differences between constitutional liberalism and cultural liberalism. Moreover, the differences probably tend to increase in the past few years. Today, it is really interesting to study the differentiation between both. Abortion is an interesting example. In France, the far right is divided on this specific topic. Vladimir Poutine is not opposed to the right to abortion, but he is still dismantling democracy by force. One solution might be to adapt the vocabulary of questions in the surveys as the political vocabulary evolves over time. But at the same time, they want to be able to make comparisons over time, so the terms used must remain the same. All in all, this is a fundamental tension that he and his colleagues systematically confront.

Nonna Mayer's questions to Jan Rovny

Nowadays, there are expert surveys on "populism". This makes things even more complicated. Then, Nonna Mayer wonders whether a "party position" really exists. What would be such a thing? What are the cues? Is it what the leaders say? What the voters think? Sometimes, within the same party, there are strong cleavages. In France particularly, as mentioned before, very strong internal party debates exist. Then, she adds, the very notion of "change" can be criticized. If there is an epistemic community forming, it may be quite difficult to capture the "changes".

Jan Rovny's answer to Nonna Mayer

Regarding the "populist" concept, Jan Rovny explains that he and his colleagues basically ask "anti-elite / aliens" item. Obviously, there are probably more dimensions to populism. But they can't add seven more items either... They are forced to take shortcuts. Then, Jan Rovny suggests that party position cannot be more than an abstraction. A party position is necessarily hard to define. The unclear positions within the political parties will probably be reflected in the standard divisions. Other questions could be added to the surveys to better capture these internal disagreements between political parties, but again, this would double the length of the survey. Jan Rovny come back to the notion of "changes". He agrees that experts can be biased. Yet, Jan Rovny concludes that in his view there is no point in multiplying the number of experts indefinitely. As René Lindstädt pointed out, they don't study the experts; they study the political parties. The number of experts can be as small as possible without affecting the validity of the surveys. It is all about their actual expertise.

Emilien Houard-Vial's question to Jan Rovny

Emilien Houard-Vial (Sciences Po, CEE) wonders if Jan Rovny has ever thought of sending forms to specific experts of certain parties (not necessarily from the same country). He understood that it would not necessarily be easier.

Jan Rovny's answer to Emilien Houard-Vial

Jan Rovny replies that they build their surveys per countries, so it would mean that they would have to design specific surveys per parties and per experts. It would be too time-consuming.

Nonna Mayer warmly thanks Jan Rovny and René Lindstädt for their clear and fascinating presentations, and all participants for their questions.

This summary was written by Justine Brisson and checked by the speakers.