*Science in Context* **15**(3), 427–445 (2002). Copyright © Cambridge University Press
DOI: 10.1017/S0269889702000558 Printed in the United Kingdom

# Sampling and Democracy: Representativeness in the First United States Surveys

*Emmanuel Didier*

*Max Planck Institut für Wissenschaftsgeschichte. E-mail: em.didier@free.fr*

## Argument

How did statistical representativeness first appear in the United States? To answer this question we turn to the archives of the Department of Agriculture. Since the 1860s, agricultural statisticians have been working to come up with representative groups of farmers able to answer questions about crop production. The archives describe the methods and tools used to devise representative samples. But the explicit justifications of *why* a given sample is particularly representative are nearly always missing. Where can we find *explanations* for why one sampling method is representative and not another? Where can we find the *why* when we have only the *how*? This paper argues that we can turn to the *theorists of democracy* for answers, because selecting a sample and selecting a group of representatives have a lot in common.

Historians of statistics attribute the invention of "representative surveys" to the Norwegian Anders Kiaer who presented his ideas from 1895 to 1903 at meetings of the International Statistics Institute. These origins are unquestionable, for it was indeed his speeches that had the strongest influence on the theory of surveys as we know it today.

The fact remains, however, that other types of surveys were already being used elsewhere at the time, even though they had a lesser impact on recognized theory. In particular in the United States the "Division" (a section of the Department of Agriculture)[1] conducted partial surveys beginning in 1863 for the purpose of measuring the country's agricultural production. These surveys were based on a fairly simple principle of estimation. The variable in question was the total production of different cereals in terms of volume. To calculate it, the Division asked informants to evaluate two intermediate variables per cereal: the area cultivated and the yield. Area

[1] We have chosen to refer to this administration as the "Division" because its name constantly changed during the period in question. Although it is generally known as the "Bureau of Agricultural Economics," this name was not used before the 1930s. Independently of the name, the archives of this administrative entity have all been filed in the U.S. National Archives, Record Group n° 83.

https://doi.org/10.1017/S0269889702000558 Published online by Cambridge University Press

was expressed in acres (and thus statisticians called this variable "crop acreage"), and yield was termed "condition of growth" (informants were asked what the plants' conditions of growth had been during the period surveyed). All this information was requested in the form of a percentage of variation compared to a norm, established on the basis of the ten-year census. The Division then simply had to translate these percentages of variation into an absolute value and then multiply the area by the yield to obtain an estimate of production. The Division's main problem was not estimation, however, but the selection of respondents since taking an annual census was unfeasible. To whom could questions be put so that partial results might be valid for the whole? How did "representativeness" itself blossom outside the scope of Kiaers' speculations?

To address these questions, we will trace the history of the respondent: how he was conceived and what operations were performed by the Division to employ him. We shall forget all our representations concerning today's polls and surveys, which tell us a lot about the respondent (that he is only interested in money, that he tends to lie, etc.), to see how he was conceived and questioned in the first thirty years of the twentieth century. By looking at the accounts of the first surveyors, we will describe this person insofar as he can be representative of others.

It will be helpful to turn to the political sphere. Statistical representativeness poses problems very similar to those of representation in "representative democracy." Democratic systems, like surveys, seek solutions so that a part (the representation or the sample) applies to the whole (the general will or the total collection). Since democratic modes of representation have been defined far more systematically and explicitly than the first surveys, they might relate to statistical models of selection.

In this paper, we will present three main selection methods used in these surveys. We will see that the methods used may seem strange today. Yet our aim is not to smile at how things were done in the past but rather to understand why statisticians used particular selection methods. This takes us up to the early 1930s, before statisticians studied the question of randomness. To conclude, we will examine the link between democratic theories and statisticians.

## I. Voluntary Crop Reporters: the Spokespersons

The first method of selecting respondents dates back to the beginning of the century. Those selected were named "voluntary crop reporters" by the Division. Selection was not done on a one-off basis as it is today; it was carried out with a view to building lasting relations. The Division had a mailing list of the crop reporters it had selected, and every time a survey was launched it sent a questionnaire to each person on the list, which was extremely stable. It also sent them *Crops and Markets*, the monthly bulletin in which survey results were published (see fig. 1). The Division made an

https://doi.org/10.1017/S0269889702000558 Published online by Cambridge University Press

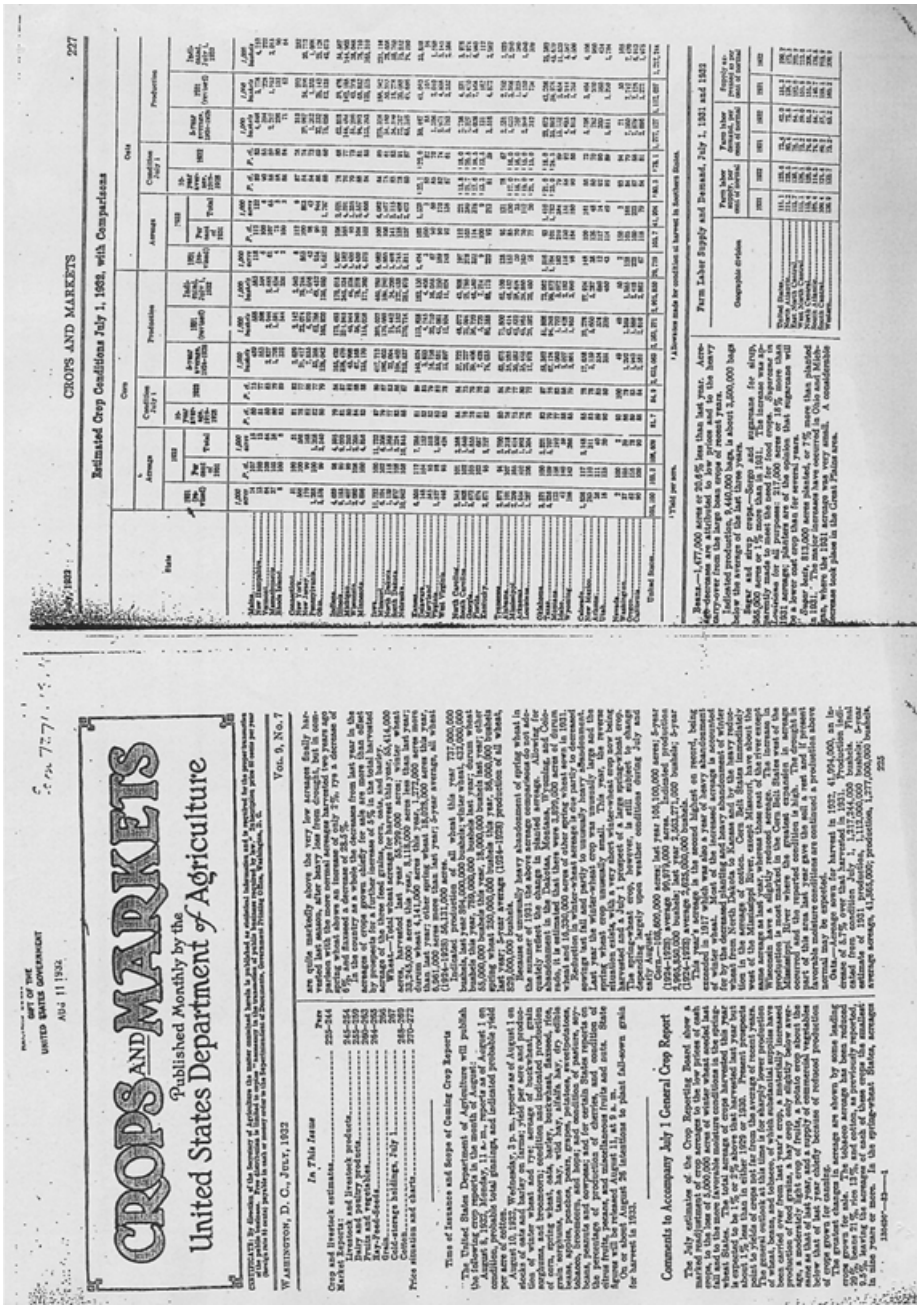


**Fig. 1.** *Crops and Markets* shows the Division's activity, some figures, and their commentary. Note the three columns "acreage," "condition," and "production," the three main variables of the agricultural statisticians. After a general comment, the Division says a word about each crop.

https://doi.org/10.1017/S0269889702000558 Published online by Cambridge University Press

effort to maintain close and lasting contact with its crop reporters and even sent Christmas greetings.[2]

Reporters were chosen individually on the basis of two criteria. First, they had to be volunteers; they had to agree to the fairly heavy constraint of sending back about a dozen questionnaires each year. This criterion automatically excluded a large number of potential respondents, for rural people were reluctant to comply with such demands: "the building up of a large mailing list was rather difficult and progress was relatively slow."[3]

Second, they had to be able to compile and report real information on agriculture. Figures were constructed from "intelligent and reliable" answers and excluded the misleading ones of "the ignorant and those fearful that the truth may increase taxation." Thus, crop reporters were chosen for "their known intelligence and judgment" (Taylor and Taylor 1952, 188). The quality of agricultural statistics was based precisely on the superior abilities of the crop reporters to recognize the truth, so that they were able to conduct detailed surveys. "Their figures [were] based upon personal knowledge and inquiry."[4] Crop reporters were therefore almost as much interviewers as they were respondents. Extremely active, they inspected the crops and used their knowledge and experience to translate their observations into figures. The crop reporter *knew* for others.

The reporters provided information about a large area. "The reporters should embrace a neighborhood, and not merely [the] reporter's own crops, for their condition may depend upon his own good or bad husbandry" (Taylor and Taylor 1952, 177). Even though it was a particular person talking, he was to talk about more than his own specific case. Since the reporters were considered intelligent and reliable, they were judged capable of distinguishing their own practices from those of others.

To ensure representation of the entire territory, it was systematically divided up along the administrative lines of agricultural townships, the aim being to have at least one crop reporter (i.e., a representative) per township. "This list was started in 1896 and was made up principally of farmers, the object being to have each and every agricultural township represented by at least one reporter" (Church 1943, 7). This was more or less achieved, since in 1926, for example, this list consisted of 40,000 names, far more than the number of agricultural townships (USDA 1933, 4). The

[2] R.G. 83, Entry 85: Schedules. This entry contains the notes of the section responsible for posting the documents. We salute Miss Skinner who had to put the documents into envelopes for she had an impressive amount of work, for example, 14,000 questionnaires to send in July 1929!

[3] Church, V., *Personal recollection of the Government Crop reporting service*, photocopy, March 1943, p. 120. This text, an internal document of the National Agricultural Statistical Service (N.A.S.S., today's descendant of the Division, responsible for the production of agricultural statistics) and unpublished, is the autobiography of an agricultural statistician (that is, a civil servant responsible for figures for one of the 48 states).

[4] R.G. 83, Entry 74: Advisory committee on grain crop reports. "Grain crop reports" unpublished article with neither author nor date.

https://doi.org/10.1017/S0269889702000558 Published online by Cambridge University Press

selection mechanism here was thus to find at least one voluntary representative for each element comprising the whole.

The reason this selection method could be used to aggregate partial figures into a single figure that applied to the whole strangely resembles the reasons put forward by the Federalists – one of the two parties that drew up the American Constitution – to explain how representatives can combine individual wills in a democracy. For the Federalists there existed a "real interest of the country" which could not be confused with that of ordinary private individuals who might sacrifice it for "short-lived and partial considerations" (Madison [1787] 1961). The interests of the whole country transcended the sum of interests attached to the different conditions in which citizens lived; there was no reason for these "partial" interests to correspond to the interests of the entire country. Similarly, the crop reporters had to see the real value of production in the entire neighborhood, without their judgment being influenced by their own results.

In the Federalists' view, this real interest could be perceived only by chosen individuals who were distinguished and superior. James Wilson, for example, could see nothing wrong "with the fact of appealing to men who were particularly noteworthy for their virtue and talents" to represent the nation (Wilson [1888] 1966). "Whether the scale was that of wisdom, virtue, talent or purely and simply fortune, all [the Federalists] hoped that the representatives would be somewhere above the electorate" (Manin 1995, 158). Similarly, the crop reporters were allegedly the most intelligent members of the rural population.

The consequence of this conception was that representatives were no longer anything like those they represented. Wilson dared to use the term "aristocracy" (despite the oratorical risks this involved in a debate on the constitution of a democracy) to refer to representatives. Yet everyone shared the same transcendent interest. The crop reporters similarly formed a sort of elite described as "public spirited," for they alone, thanks to their own inspection of the neighborhood and to their public spirit, were able to discover the truth of total production, valid for all.

The first agricultural surveys thus relied on an informant who seemed to be inspired by his role as representative as conceived in the federalist theory of representative democracy. This link between statistical surveys and a political philosophy is based not only on formal proximity but also on the history of surveys in the United States. Throughout the nineteenth century crop reporters had been elected like representatives of townships. The first figures on U.S. agriculture date back to 1862 and were produced not by the state but by a journalist, Mr. Orange Judd, managing editor of the *The Agriculturist* who suggested: "let the readers of the *Agriculturist* in every town counsel together, and select some man who may be relied upon for good judgement and general ability to estimate [the condition] with some degree of accuracy in regard to the leading crops, wheat, corn, etc." (Taylor and Taylor 1952, 176). In 1863 the federal government undertook to produce these statistics and seems to have continued the practice of questioning representatives, even

https://doi.org/10.1017/S0269889702000558 Published online by Cambridge University Press

if the different sources do not agree on *which* representatives were chosen. For an historian of agricultural statistics, it was members of Congress who submitted to the Division the names of local representatives chosen as future crop reporters (Vogel 1995). In the memoirs of a former practitioner, "monthly questionnaires sent to township supervisors provided the basic information for the published state reports. There were about 1,200 townships each having a supervisor who was elected by popular vote annually in April" (Church 1943, 1). When the government took over publication of the figures it also adopted the idea that the best sources of information were elected representatives. Then the Division expanded the group of crop reporters to include people who were volunteers but not necessarily elected representatives. Yet personal characteristics peculiar to the political sphere (superiority in discernment and public spirit) continued to influence the conception of crop reporters.

This method functioned until 1921, when an error was noticed. The government had initiated a campaign to reduce production. According to the Division, all the farmers *said* they would comply with the Government initiative, and the crop reporters believed them and reported the reduction, but the farmers *did* nothing of the kind. Therefore, final estimations happened to be far below actual production. The Division consequently deduced that the crop reporters had been deceived by false information supplied by their neighbors and tried to improve its method by making it more "objective." This incident had far-reaching consequences.

## II. Individual Farm Inquiry: Objectivity of Informants

After the error of 1921, a solution was proposed. They decided to replace the crop reporter with a "crop meter," a machine similar to a car's odometer, equipped with several buttons. It worked in the following way: an official would, for example, drive in his car along a wheat field and the crop meter would measure the length. When he moved on to a cornfield he would press the "corn" button and the machine would measure the size of the new field. Initially the machine was showered with praise because it was objective, despite some problems (in particular, it measured only one of the two dimensions of the field). But Nat Murray, the Division's chief statistician, vigorously opposed doing away with the crop reporters' services. He pointed out that they were "representative and public spirited men" and that it would be a mistake to remove "faithful reporters who have always been and must continue to be [the Division's] principal reliance for securing data" (Murray to Yeager, 15 January 1922, RG 83, entry 75). We have to insist here on the fact that the reporters are still seen as *willing* persons. It is not *they* who have been misleading, but *other* farmers who grudged the first who kept being the Division's allies. Hence, the problem was not so much to control the reporters more efficiently, but to make sure that the information they collected was not biased. In other words, it was the way of selecting the reporters that had to be reformed.

https://doi.org/10.1017/S0269889702000558 Published online by Cambridge University Press

The improvement consisted of asking only farmers for information. Before 1921, as long as crop reporters were thoroughly familiar with the area, they were considered capable of informing the Division, irrespective of their profession. They were often farmers, since they lived in farming areas, but they were not exclusively farmers. For example, we read about a retired army officer or a postal employee (Church 1943, 8). But after 1921 it was decided that only farmers could "give actual facts as to the acreage":

> Instead of asking the reporters to give their opinion as to the change in acreage from year to year, a large number of farmers in every locality are now asked from time to time to give actual facts as to the acreage in various crops and numbers of livestock on their own farm. (USDA 1925, 14)

This development forced a halt to relations established with the few who did not own land, but it had the big advantage of "saving" almost all the crop reporters since most of them were farmers, and their duties were nearly unchanged: they still received the questionnaires, the monthly paper of the Division, etc. and they still had to send back information. However, by changing their qualifications, the Division was able to describe the statements of voluntary crop reporters as the objective statements of farmers. How was this possible?

> In the early days of the crop reporting service "judgment" inquiries concerning acreage changes were the principal indications used concerning acreage changes. In these inquiries a crop correspondent was asked to express his judgment of changes in acreage of crops in his locality by reporting the current acreage as percentages of the corresponding acreages of the previous year. Analysis of the results from this type of schedule indicates that it is difficult for a farmer or anyone else to report with sufficient precision as to the acreage changes that have occurred within the locality for which he is reporting. . . . Analysis of the results obtained from the record of actual acreage of crops on sample farms from year to year indicates that this method is much superior to the use of judgment inquiries. Consequently the inquiries requiring the judgment of the correspondents concerning acreage have been practically discontinued. They are replaced by sample farm acreage data – records of the actual acreages of crops on farms of crop reporters. The change to this basis was a definite forward step toward increasing the accuracy of the reports. (USDA 1933, 16–17)[5]

Use of percentages was maintained before and after 1921 but the area the crop reporter had to study changed. Before he was in his "locality"; afterward he was only on his farm. Apart from the fact that the latter was smaller than the former, the nature of these two spaces was different. A farm was totally under the control of its farmer, whereas the neighborhood seemed to encompass the crop reporter.

[5] It was probably someone called Sarle who wrote this text. He is almost always the author of more general sections on the Division.

https://doi.org/10.1017/S0269889702000558 Published online by Cambridge University Press

When the question concerned the locality, the instructions written on the questionnaires were as follows: "Report ONLY on such crops named as are grown in your locality. Report for the part of the country around you which comes under your personal observation" (USDA, Nov. 1923, RG 83, entry 74 *November crop schedule*). In these conditions, the crop reporter was included in his locality and "your" referred to that which surrounded him. By contrast, the "individual farmer inquiry" specified: "can you fill in the schedule below for your farm or plantation?" (USDA 1933, 69). Here the respondent owned the farm, for after the change, "your" meant "what you possess." This is clear in a citation above, where farmers were asked to provide data on their "own farm" (USDA 1925, 14).

On the questionnaire that we reproduce here (which is a livestock schedule, but the difference is in our case of no importance), note that the farmer should indicate the situation "on farm" for each one of the three main categories of questions (see fig. 2). Again, in the explanatory text, it is said that the correspondent should express the changes that occurred "on [his] farm." Why did the fact of owning the farm enable a farmer to give the "real surface area of crops"? Why did the Division use on the same page the expression "objective method" to describe the survey on farms? A physicist does not own his electron, but nonetheless he knows it pretty objectively. Why then, when the crop reporter possessed what he was talking about, was he considered more objective?

An answer to this question can be found in the liberal tradition initiated by Locke and subsequently represented by Schumpeter. For example, when Schumpeter criticizes the Enlightenment thinkers' rationalist anthropology, he describes the conditions in which citizens act rationally and objectively, as follows:

> And so it is with most of the decisions of the daily life that lie within the little field which the individual citizen's mind encompasses with a full sense of its reality. Roughly, it consists of the things that directly concern himself, his family, his business dealings, his hobbies, his friends and enemies, his township or ward, his class, church, trade union or any other social group of which he is an active member – the things under his personal observation, the things which are familiar to him independently of what his newspaper tells him, which he can directly influence or manage and for which he develops the kind of responsibility that is induced by a direct relation to the favourable or unfavourable effects of a course of action. (Schumpeter [1942] 1976, 258)[6]

We note how Schumpeter emphasizes possession and observation. For him, the citizen is rational with regard to the "little field" which he "can directly influence or manage." More precisely the philosopher points out that the further one moves from this field, the more one observes "a decrease in the ability to discern facts," until finally one witnesses "a quasi-total disappearance of the sense of reality" (ibid., 360).

[6] I thank Laurent Thévenot for pointing out this comment to me.

https://doi.org/10.1017/S0269889702000558 Published online by Cambridge University Press

https://doi.org/10.1017/S0269889702000558 Published online by Cambridge University Press



**Fig. 2.** Monthly Live Stock Schedule by the U.S. Department of Agriculture. Note the third line of the explanations: "By using a large number of such reports, each considered as a sample."

For Schumpeter, the more the citizen possesses something that he can directly influence and control, the better he knows it.

This is the reasoning used by the Division's statisticians. Initially, they set the correspondents in their townships, one of the elements of their field. But the 1921 error proved that it might escape the correspondents' sphere of action. It happened to be too vast. To enhance the crop reporters' objectivity, statisticians chose to set them in a space over which they had more means to act, the greatest means possible, their own farm. Thus, it seems that for the statisticians, truth lay in the bond between each individual and his "property," as Locke calls this bond, and to ensure truth, they had to ensure the bond.

But this solution to the problem of the correspondents' objectivity posed a new problem: how could multiple answers, each more objective than before, be combined when there was no longer complete representation of the territory? How to sum up the collection of reports, when the sum would not be equal to the whole U.S.? The solution has been that answers could be generalized because farmers were *examples* of those who remained silent.

The farmers' role as examples can probably be illustrated by the appearance of a new word in this system of statistics: "sample." Statistics manuals in the 1920s, including in the U.S., did of course use this word unambiguously in the sense of a group of individuals,[7] but then this meaning was not the same as the Division's, where each respondent was itself a "sample." The "sample" was a person rather than a group, for example, in the formula noted on the back of a questionnaire from 1920 addressed to farmers, since it specified that everybody was "considered as a sample" (RG 83, Entry 85; see fig. 2). We note that the word derives from the French *exemple* and that its initial definition was "a fact, incident, story or suppositions case which serves to illustrate, confirm or render credible some proposition" (OED 1973). In the 1920s this definition was perhaps forgotten because it dated from the sixteenth century, yet there is still something very close to example in the above usage. A "sample" was not yet a group of persons, for each person constituted a whole sample. The 1920s sample can therefore be understood as an example of what was happening around the person concerned and about which statisticians were perfectly sure.

But even so, the Division encountered a difficulty: the lack of respondents, evident in the following citation from *Omnibus*: "Koenig of Pennsylvania reports very favourable returns. . . . Of the schedules sent to farmers on the pig survey card, he received a 20 per cent return."[8] What is now known as the rate of non-response was therefore extremely high, since Koening was satisfied with a rate of 80 per cent non-

[7] One example among many statistics manuals written by Americans is King 1924.

[8] *Omnibus* was a small newspaper produced by one of the agricultural statisticians. He copied it and sent it to his colleagues. There was therefore nothing particularly official about it. It essentially contained little comments on noteworthy facts concerning this occupation, poems that a colleague had written, jokes, etc. These small newspapers were kindly shown to me by Richard Allen, current director of the NASS, who keeps them carefully in his office. Koenig was an agricultural statistician.

https://doi.org/10.1017/S0269889702000558 Published online by Cambridge University Press

response! This lack of information was a problem since too few examples made generalization hazardous. An improvement in the selection method aimed to solve this problem of a scarcity of respondents.

## III. Detours by Mail Carriers: Large Numbers and Representativeness

Improvement in the method did not call directly into question farmers' objectivity regarding their own farm. It consisted of selecting them differently, by relying on mail carriers to choose the people who would receive a questionnaire.

> In an effort to obtain a . . . much larger sample of individual farm acreages of crops, the cooperation of the Post Office Department was obtained in 1924 for a survey to be made by the rural mail carriers covering a definite number of farms on each rural mail route. . . . This survey does obtain a very large sample. (USDA 1933, 17)[9]

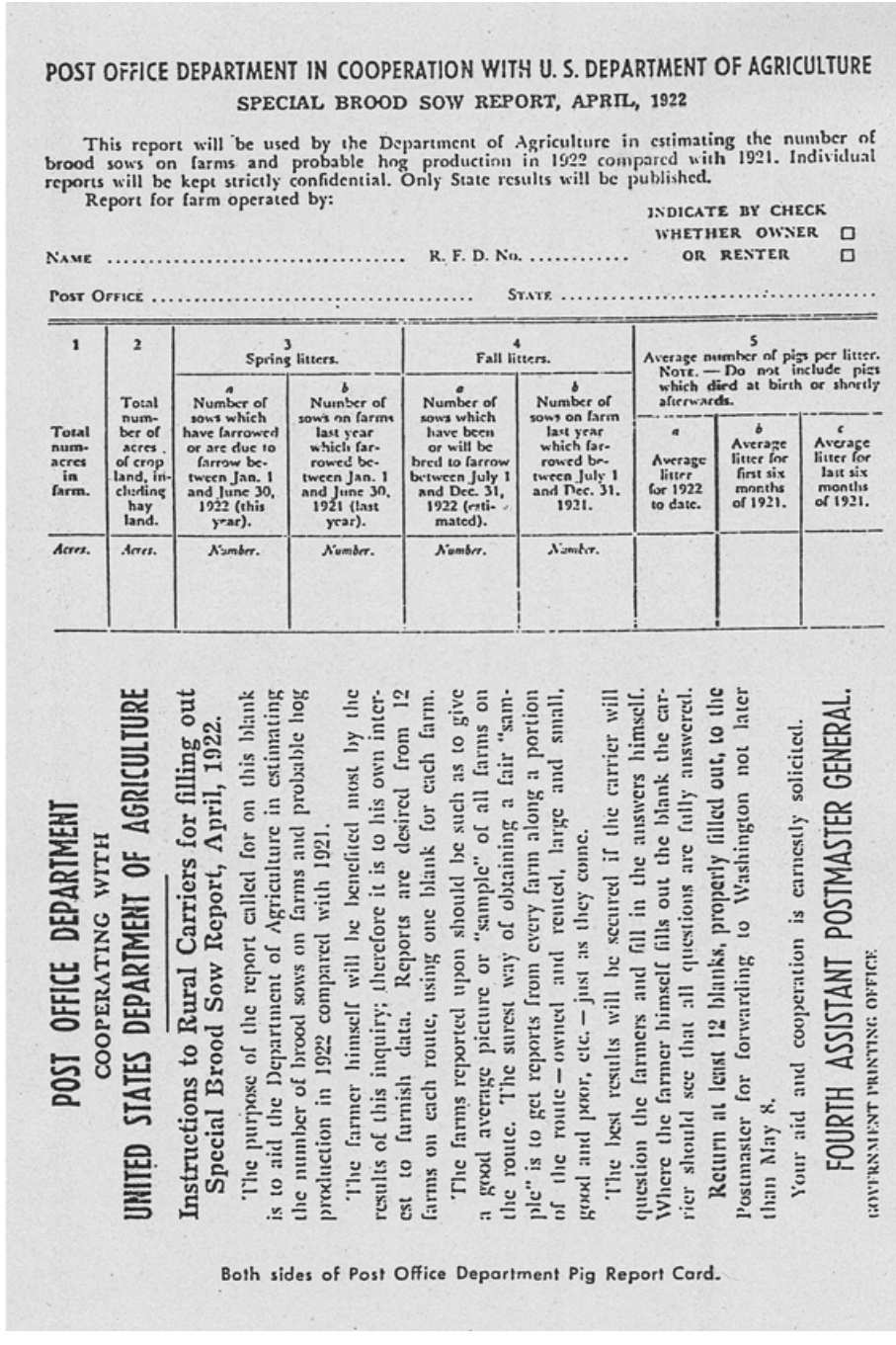
To reach more farmers the Division turned to the Post Office because mail carriers could distribute questionnaires in mailboxes. There would be far more questionnaires distributed than the number of names on the list of crop reporters, since instructions recommended distributing ten per postal route (USDA 1933, 68), which is much more than the number of names on the Reporters' list. But at the same time, this instruction shows also that the aim was not to have the most answers possible. For example, Henry Taylor, the then-Division Head, wrote, "careful studies have been made to determine the number of farms that are necessary to give an accurate picture of each area" (USDA 1925, 15). The aim was simply to have the "right number" of useful questionnaires.

The fact that this distribution experiment happened to be successful showed that when the number of people who receive questionnaires is multiplied sufficiently, the number of answers obtained is also multiplied. Thus loyalty, the lasting agreement between the farmer and the statistical system, which previously ensured that questionnaires were returned, was no longer necessary. It was enough for crop reporters to answer the questions when they were asked. It was then that the selection became a rapid interaction in which it was enough to enter into brief contact with an unknown person. This job was readily entrusted to mail carriers: "The best results will be secured if the carrier will question the farmers and fill in the answers himself" is written on the example here reproduced (see fig. 3).

But how were mail carriers to distribute the questionnaires? Could they just get rid of them any way at all? No. To aggregate the data they had to follow certain distribution instructions (written on the questionnaire itself), as follows:

[9] Official cooperation began in 1924, but experiments in cooperating began as early as 1921.

https://doi.org/10.1017/S0269889702000558 Published online by Cambridge University Press



POST OFFICE DEPARTMENT IN COOPERATION WITH U. S. DEPARTMENT OF AGRICULTURE

SPECIAL BROOD SOW REPORT, APRIL, 1922

This report will be used by the Department of Agriculture in estimating the number of brood sows on farms and probable hog production in 1922 compared with 1921. Individual reports will be kept strictly confidential. Only State results will be published.

Report for farm operated by:

INDICATE BY CHECK WHETHER OWNER ☐ OR RENTER ☐

NAME ..... R. F. D. No. .....

POST OFFICE ..... STATE .....

| 1 | 2 | 3 Spring litters. | | 4 Fall litters. | | 5 Average number of pigs per litter. NOTE. — Do not include pigs which died at birth or shortly afterwards. | | |
|---|---|---|---|---|---|---|---|---|
| | | *a* | *b* | *a* | *b* | *a* | *b* | *c* |
| Total number acres in farm. | Total number of acres of crop land, including hay land. | Number of sows which have farrowed or are due to farrow between Jan. 1 and June 30, 1922 (this year). | Number of sows on farms last year which farrowed between Jan. 1 and June 30, 1921 (last year). | Number of sows which have been or will be bred to farrow between July 1 and Dec. 31, 1922 (estimated). | Number of sows on farm last year which farrowed between July 1 and Dec. 31, 1921. | Average litter for 1922 to date. | Average litter for first six months of 1921. | Average litter for last six months of 1921. |
| *Acres.* | *Acres.* | *Number.* | *Number.* | *Number.* | *Number.* | | | |

POST OFFICE DEPARTMENT
COOPERATING WITH
UNITED STATES DEPARTMENT OF AGRICULTURE

Instructions to Rural Carriers for filling out Special Brood Sow Report, April, 1922.

The purpose of the report called for on this blank is to aid the Department of Agriculture in estimating the number of brood sows on farms and probable hog production in 1922 compared with 1921.

The farmer himself will be benefited most by the results of this inquiry; therefore it is to his own interest to furnish data. Reports are desired from 12 farms on each route, using one blank for each farm.

The farms reported upon should be such as to give a good average picture or "sample" of all farms on the route. The surest way of obtaining a fair "sample" is to get reports from every farm along a portion of the route — owned and rented, large and small, good and poor, etc. — just as they come.

The best results will be secured if the carrier will question the farmers and fill in the answers himself. Where the farmer himself fills out the blank the carrier should see that all questions are fully answered.

Return at least 12 blanks, properly filled out, to the Postmaster for forwarding to Washington not later than May 8.

Your aid and cooperation is earnestly solicited.

FOURTH ASSISTANT POSTMASTER GENERAL.

GOVERNMENT PRINTING OFFICE

Both sides of Post Office Department Pig Report Card.

**Fig. 3.** Special Brood Sow Report, April 1922, published by the Post Office Department in cooperation with the U.S. Department of Agriculture.

https://doi.org/10.1017/S0269889702000558 Published online by Cambridge University Press

> The farms reported upon should be such as to give a good average picture or "sample" of all farms on the route. The surest way of obtaining a fair "sample" would be to get reports from every farm along a portion of the route – owned and rented, large and small, good and poor, etc. Just as they come – care being taken to select a portion that will be fairly representative of the entire route. ("Special hog report," December 1, 1923, RG 83, entry 74)

To understand this mode of selection, we first note that postal routes corresponded to partitions of U.S. territory (in reality there were areas where mail was not distributed, but the Division made an approximation). The problem therefore was to obtain good data from each one and then simply add them together. The originality of these instructions lies in the idea that relevant information is to be obtained from "every farm along a portion of the route." They are original for two reasons. First, they mark a departure from the past: the statisticians neither rely on a superior individual nor on good examples, as they used to do before, but on a large number of cases. Second, they differ from all the other experiments carried out elsewhere in the world: we have seen in other articles in this volume that in Norway and Russia, where sampling was also carried out from the beginning of the twentieth century, statisticians did not select clustered groups but one individual every n-th (they did what we call nowadays "systematic selection"). Thus, the idea of selecting small groups seems to appear in the U.S. context; which makes it worth studying.

Moreover, the feature consisting in selecting a small area will constantly continue to influence American statistics. For example, the Division played a crucial role in inventing a method now known as "area sampling," which consists of selecting randomly small zones of territory instead of individuals (Vogel 1995). Hence, we have to account for this American idiosyncrasy: why did they select "every [unit] along a portion of route"?

The most explicit justification of these instructions is cited above: the portion allowed the Division to calculate a "good average image" in the route. Actually, as the survey was exhaustive on this small area, the average calculated on it would have meaning. The relation between average and exhaustiveness has been established since Quételet.[10] For the average to relate to something real in a group, it has to be calculated from exhaustive data in that group. For the Belgian astronomer, the "average man" who characterized a nation had meaning only if his characteristics

[10] Note that Quételet's "average" is related to the political philosophy of Victor Cousin. Laurent Thévenot has shown the extent to which Quételet in his work mixes "political forms of 'representation' of the 'people' or the 'crowd,' a political philosophy of 'superior power' of the 'great man' and the 'misery' and 'smallness' of 'individuality,' a statistical law of the 'common type' and a metaphysics of 'the general' and 'the particular'" (Thévenot 1994). This philosophy involves the concept of 'esprit': "the great man [who for Cousin is a person but for Quételet is the 'homme moyen'] is the harmony of particularity and generality; he is great only at that price, on the condition that he represents the general 'esprit' of his people" (Cousin quoted in Quételet 1835).

https://doi.org/10.1017/S0269889702000558 Published online by Cambridge University Press

were calculated on the basis of a population census. One of the reasons why the average must be calculated from an exhaustive group is that minor differences cancel one another out. That is why mail carriers also had to be careful to include "large and small, good and poor" farms in their surveys. This instruction made it possible to calculate meaningful averages on the portion of the route that had been selected. But how could this information then be generalized to the whole route? How could an average valid for the *portion* be valid for the *entire* route? What are the characteristics of this group and of its relation to the whole that make it fair to generalize?

The answer lies in the concept of "representativeness" that appeared in the Division at the time of this survey and that we must therefore clarify. We note, first, that representativeness goes with the word "sample" but only when the latter denotes a group and no longer an individual. Initially the sample only hesitatingly came to designate a group; we see the hesitation in the quotes used with the word in mail carriers' instructions. But as the idea of a "representative portion" was established, "sample" came to denote a group of persons, exclusively. For example, at the end of Taylor's sentence cited above, he claims to be seeking "the number of farms that are necessary . . . , in order that a truly representative sample may be obtained" (USDA 1925, 15). Representativeness became meaningful to statisticians when sample was understood to be a group of people.

But why, in America, did this small number of persons have to be clustered in a portion of rural route? The answer lies in the wish of the agricultural statisticians to get reports from neighbors, from people who lived close to one another and entered in direct interactions. This is related to the reason given to explain the 1921 mistake: it was the conversations, the face-to-face relations between farmers (including the correspondent) that allegedly caused the error. So, to get a good representation of a route, statisticians wanted of course to gather objective information about the agricultural practices of the respondents, but they also wanted to grasp something of the dense and rich set of behavior, conversations, relations that influenced the correspondents' *declarations*. Interviewing neighbors was a way to get information not only about farmers' production, but also about the influences (some real and some misleading) that each farmer exerted on the others through face-to-face relations, it was a way to grasp something of the error committed by interviewees.

The problem remained that these two sets of "information" (facts and misleading influences) were mixed up in the answers of the sample. The statisticians still had to separate one from the other. Therefore, they sought other representations of the fields than the reports, and tried to estimate – on the basis of this supplementary data – the differences in the sample between declaration and production. For example, the statistician in charge of Wisconsin wrote in 1922:

> In the South there was some tendency to reduce the area of tame pasture and put it into crops. Our preliminary acreage as estimated [by the sample] at present shows an increase

https://doi.org/10.1017/S0269889702000558 Published online by Cambridge University Press

> in the vicinity of 4 per cent. This, I believe is somewhat high. Our final revisions in December will probably show an increase of between 2 per cent and 3 per cent. ("Comments on Crop Reports, Becker, Wisconsin," 1 October 1922, RG 83, entry 83)

The statistician takes note of the farmers' declarations since his reasoning is based on the average of their reports. But he also corrects it, he allows for a bias (here based on his own observations and convictions but it could have been on some other type of information, for example data provided by the Weather Bureau or by the Census Bureau).

And we can see that it is this difference, this error, that he generalized to the whole population. The selection of a group of neighbors gave some sense of reality to the error made by respondents (it came from their direct interactions), and it could therefore be generalized legitimately. The precise type of influence that characterized interactions between individuals in a portion of the route was efficient all along the route. A farmer could be in any situation on a route, he could rent his farm or own it, he could grow rice or grow barley and – very important from the statistical point of view – he could be in or out of the sample, in any case he would not escape this special type of influence that his little group of neighbours exerted on him and that he himself exerted on them. He would, more or less, err as much as they if he were to give an account of his production. The declaration errors committed by all the neighbors in a portion of the route was allegedly stable enough to be generalized from the sample to the whole route.

Where are the origins of the practice of selecting small groups? First, we have to admit that the archives of the Division have not given any *explicit* answer to this question. So we also searched mail carriers' records, in the hope that the chiefs would have explained to the carriers what is a "representative sample" and where it comes from. But we did not find anything of interest to us (only few claims concerning freely distributed mail, in general, and nothing about the schedules of the Division themselves)[11]; nor did we find anything in the historiography of that institution (see e.g., Cullinan 1968).

The only thing that we may say is that a common feature of American statisticians and American social scientists is that they give special importance to small groups of individuals. Allan Silver, sociologist at Columbia University, remarks that American sociologists have always shown a great deal of interest in "small groups," to which they have given a "curious importance" (i.e., an explanatory value) compared to European sociologists (Silver 1990, 64). He notes the following American examples: "Charles

[11] We went through every issue from 1924 and 1930 of *The Postal Record*, a monthly dedicated to the mail carriers.

https://doi.org/10.1017/S0269889702000558 Published online by Cambridge University Press

Horton Cooley, E. A. Ross, George Herbert Mead, Mary Parker Follet, and later Robert Park and W. I. Thomas [who endowed] the small, face-to-face group with strategic significance"; however, his European counter-example is Durkheim who never studied anything other than "impersonal dynamics" without the intermediary of small groups.

The origins of this social science feature, according to Silver, is to be found in a certain colonial spirit, when little communities of pilgrims were independent and autonomous and at the same time constitutive of a bigger set: the whole new world. "Small groups of interacting persons [were] both constituted by the larger society and, in turn, constitute[d] it. It follows that the small group [became] essential to understand the society as a whole" (Silver 1990, 7). Silver explains that small groups became central at the level of analysis because they first were the key social forms that linked the local with the general at the level of government.[12]

Statisticians who had put neighborhood at the heart of their reflection had, due to the failure in 1921, come to the conclusion that although crop reporters were "public spirited," they could not be made into a sort of rural "aristocracy" because they had been misled by their neighbors. Their answers depended too much on the interactions in which their neighborhood surveys led them. To address this problem, the statisticians got interested in the multiple face-to-face relations of the farmers. To know the production of a postal route, they relied on the answers of a whole group of neighbors because their interactions were supposedly the same, identical throughout the whole route and thus could be generalized. This idea of identity seems to be linked to the congregational theory of equality in the church.

## Conclusion: Representativeness and Democratic Topics

We have observed the evolution of selection methods in the first partial U.S. surveys. This evolution is interesting because it enables us to see how the concept of a sample

[12] More precisely, Silver traces the importance of small groups in the social sciences in America back to the Congregationalist ecclesiology because of its conception of the equivalence of the local and the global in the church. Congregationalists could write, for example: "The whole is not itself except as it is made up of free and autonomous parts with direct access to Christ, but the parts are not themselves except as they belong to the whole which is also informed by Christ" (Walker 1960). Their ideal of equality between individuals conceived as autonomous persons, who chose to enter into face-to-face relationships in the local churches made possible the global Church as they conceived it and vice versa. Moreover, this conception of equality between the parts and the whole also influenced American democracy. "Congregationalists have long claimed that their theology and church government are seminal not only for religious but for political democracy" (Silver 1990, 9). And their special influence resides precisely in their theory of linking equality between humans, the type of interactions into which they enter, and the "compacts among freely choosing persons" (as they described their local churches) which they form. It seems that the Congregationalist theory of equality was in the United States a very efficient intellectual shortcut to link the local and the global.

https://doi.org/10.1017/S0269889702000558 Published online by Cambridge University Press

was forged and how that of a representative sample emerged in a tradition totally unaware of debates at the International Statistics Institute.

Moreover, this study shows us that models of representative democracies can be used to understand how selection methods made it possible to generalize partial data to the whole. We will conclude by examining the concrete relations between selection procedures and democratic theories.

First, we have to keep in mind that there still remain many differences between these two worlds. Principally, the status of the action is not the same in the two instances. In democracy, citizens act primarily because they are the ones who elect their representatives. In surveys, we have seen that this practice was soon given up. It was the statistics office that assumed the prerogative of selection. Even the importance of the wish to respond was gradually reduced as much as possible. In democracy, the active participation of citizens is crucial; in statistics, on the contrary, the people are supposedly observed, thus passive. These differences perhaps explain why the relationship between democracy and surveys is not direct. In any case, it is clear that these people in their surveys never made explicit reference to the political philosophers whom we cite as sources.

Second, Silver proposes that religious culture is an "influential deep structure," acting on the social scientists, though they were unaware of it, as if they were unconsciously guided by their culture. This argument does not hold because statisticians mobilized several schemes one after the other. They were looking for a good method of selection, one that would give as precise results as possible, and *tried* several of them, in a process of trial and error. If a method were not good enough, they then simply changed it at the same time as its link – whatever it is – to representative democracy. One can hardly do this with a "deep structure." Democratic theories were not "deep structures" that acted upon the statisticians' minds because the latter changed their minds much too quickly for such a "deep structure."

But, despite those incompatibilities, representative democracy has the advantage of addressing problems very similar to those with which statistical generalization is confronted. Democracy first has to clarify the common good to which that form of government is to provide access (we have seen three examples: general interest, property, and communion). Similarly, surveys must give a status to the general truth they wish to obtain (total U.S. agricultural production is not entirely the same if defined as transcendent and independent of individual practices, as dependent on the concrete actions of actors on their farm or, lastly, as the result of individuals' face-to-face relations). Democracy must also describe relations between representatives and their electorate – it must decide whether the representative is the best of all or, on the contrary, whether, as in local churches, those exercising authority are equal to those who do not exercise it. Surveys must say how the data obtained on some individuals also apply to others (do those questioned know better than the others? Are they examples? Or is it a sub-group that is representative of the whole group?). We thus

https://doi.org/10.1017/S0269889702000558 Published online by Cambridge University Press

see that representative democracies as well as surveys must articulate a body of representatives and all the individuals so that the globality incorporated by some applies to all the others.

That is why it seems to us that democracy can be considered a *resource* for surveys, as Boltanski and Thévenot suggest (Boltanski and Thévenot 1991).[13] We have chosen to present theories of democracy that were forged well before the period under study. As these theories were widespread, they were available to the actors. Statisticians could pick up the main characteristics of the solution built by representative democracy – particularly the substitution of the whole by the part – and *translate* them into their own realm. Hence, the philosophical solutions to the problem of representation were resources for statisticians; they could use them, as they would use bricks and mortar, to solve their own problem of generalization.

To conclude, we wish to stress the fact that we have not described such resources for the purpose of questioning the figures produced. These modes of aggregating individuals are not wrong, simply by virtue of having been discovered by philosophers. On the contrary, these models are coherent and explicit. Before condemning them or deriding such practices, we recall that changes of method were aimed at improving the *precision* of the figures. In the case of representativeness, science and politics seem not to have been opposed, but on the contrary they seem to have assisted one another. We should also remember that one of the foundations of Greek democracy was random selection (see Glotz 1968; Hansen 1993; and Griffith 1966) – still the basis of the most scientific theory of surveys.

## References

Boltanski, Luc and Laurent Thévenot. 1991. *De la justification*. Paris : Gallimard.

Church, V. H. 1943. *Personal Recollection of the Government Crop Reporting Service*. Washington D.C.: National Agricultural Statistical Service.

Cullinan, Gerald. 1968. *The Post Office Department*. New York: Praeger.

Elliot, J., ed. 1888. *The Debates in the Several State Conventions on the Adoption of the Federal Constitution as Recommended by the General Convention at Philadelphia*, 5 vol. New York: B. Franklin.

Glotz, G. 1968. *La cité grecque*. Paris : Albin Michel.

Griffith, G. T. 1966. "Isegoria in the Assembly of Athens." In *Ancient Society and Institutions*, edited by E. Badian, 115–138. Oxford: Basil Blackwell.

Hansen, M. H. 1993. *La démocratie athénienne*. Paris: Les belles letters.

King, W. I. 1924. *The Elements of Statistical Method*. New York: Macmillan.

Madison, James. [1787] 1961. "Federalist 57." In *The Federalist Papers [1787–1788]*, edited by C. Rossiter, 350–351. New York: New American Library.

Manin, B. 1995. *Principes du gouvernement représentatif*. Paris: Flammarion.

*OED*. 1973. *Shorter Oxford English Dictionary*, 2 vol. Oxford: Clarendon Press.

Quételet, A. 1835. *Sur l'homme et le développement de ses facultés. Essai de physique sociale*. Paris: Bachelier.

[13] Readers familiar with this book will already have realized just how much the reasoning presented here has been inspired by it.

https://doi.org/10.1017/S0269889702000558 Published online by Cambridge University Press

Schumpeter, J. A. [1942] 1976. *Capitalism, Socialism and Democracy.* New York/London: Harper/Allen and Unwin.
Silver, Allen. 1990. "The Curious Importance of Small Groups in American Sociology." In *Sociology in America*, edited by J. Gans, 61–72. Berkeley: Sage.
Taylor, Henry C. and Anne Dewees Taylor. 1952. *The Story of Agricultural Economics in the US, 1840–1932*. Ames: Iowa State College Press.
Thévenot, Laurent. 1994. "Statistiques et politiques: la normalité du collectif." *Politix* 25:5–20.
U.S.D.A. 1923. *November crop schedule*. RG 83, entry 74.
U.S.D.A. 1925. *Annual Report of the Department of Agriculture*, Washington D.C., Government Printing Office.
U.S.D.A. 1933. *The Crop and Livestock Reporting Service in the United States*. Washington, D.C.: Government Printing Office.
Vogel, F. A. 1995. "The Evolution and Development of Agricultural Statistics at the United States Department of Agriculture." *Journal of Official Statistics* 11(2):161–180.
Walker, Williston. 1960. *The Creeds and Platforms of Congregationalism*. Boston: Pilgrim Press.
Wilson, James. [1888] 1966. "Speech on 4 December 1787." In *The Debates in the Several State Conventions on the Adoption of the Federal Constitution as Recommended by the General Convention at Philadelphia*, edited by J. Elliot, vol. II, 473–474. New York: B. Franklin.

https://doi.org/10.1017/S0269889702000558 Published online by Cambridge University Press

https://doi.org/10.1017/S0269889702000558 Published online by Cambridge University Press