CHAPTER EIGHT

HOW TO MAKE THINGS WHICH HOLD TOGETHER: SOCIAL SCIENCE, STATISTICS AND THE STATE

ALAIN DESROSIÈRES

The various parts of this collective book examine the way in which over the past two centuries the social sciences have acquired more-or-less stable forms, and have contributed to structuring specific discourses about society. Everyone agrees on the importance of the *consolidation* of these sciences – a consolidation that operates in two dimensions, *institutional* and *cognitive*. To consolidate something means to give it the ability to *endure*, to *be transmitted* from hand to hand and to *resist* possible deformations. According to Durkheim, "social facts" can only be "treated as things" to the extent that they possess these attributes, which thus render them comparable to any other scientific object.

Carrying out such a program for the social sciences themselves necessarily poses the question of the link between the two dimensions – institutional and cognitive – of this consolidation. It is clear that the various currents of the sociology of science do not agree on the nature of this link. Some content themselves with analyzing the institutional conditions for scientific progress, while according science itself (sometimes explicitly but more often *de facto*) a great degree of cognitive autonomy. These writers are not, then, principally interested in the actual procedures and techniques of the consolidation of knowledge. On the other hand, others seek to study simultaneously and *in the same way* both the scientific and the social practices leading to *things that hold together*, to *facts*. These facts have at the same time *been constructed* (the constructivist viewpoint) and yet once constructed have sufficient *existence* that none can deny them (the realist viewpoint).

This "strong program" has the advantage of reintegrating all scientific practices, whatever their nature, into sociological research (Latour, 1987).

P. Wagner, B. Wittrock, and R. Whitley (eds.), Discourses on Society: Volume XV, 1990, 195–218.

Further it permits us, by way of a very productive methodological decision rather than by philosophical fiat, to get out of the false opposition between 'constructivism' and 'realism'. Indeed, by deciding to take any social fact as *at the same time* constructed *and* real, one discovers a way of at a stroke transcending the two apparently opposed positions constituting positivist scientism and denunciatory relativism. By taking *all* scientific procedures, technical and social, seriously, we can take science just as seriously as society.

Statistical objectifications have played a key role among the techniques that have contributed to the institutional and cognitive consolidation of the social sciences. Using these, social scientists have forged tools enabling them to transcend individual or conjunctural *contingencies* and to construct *more general things* that characterize for example the social group (for the sociologist) or the long term (for the historian or the economist). However, there are several ways in which statisticians can set about this task. They can draw certain pre-constructed equivalences from society. Equally, they can construct the equivalences themselves. Thus statistics is particularly well suited to testing the hypotheses of the program in the sociology of science referred to above. In this paper, in exploring some research possibilities along these lines, we will summarize various work done elsewhere (1).

Any attempt to give form to the chaos consisting of the multitude of recordings of individual events entails a use of sources and ways of coding that are always open to two challenges from the scientist: are they available, are they reliable? In the objectivist perspective, the tools of generalization are supposed to be already there, and the only questions to be asked concern the gathering of information and its, ultimately automatic, technical treatment.

However, the Durkheimian tradition, referred to above, has left as its legacy not only this imperative of objectification ("we need to treat social facts as things"). It also has a second aspect, which relates to the social and constructed nature of apparently natural classifications ("how do social facts *become* things?"). This question was first of all asked about so-called primitive societies (the "elementary forms of religious life"), whose distance rendered more visible the fact that their classifications are not natural. Then, since the 1960s, it more and more often oriented the research of those who sought to construct and use statistics about developed societies. This implied a relationship between social science and statistics very different from that which researchers as consumers of

predeveloped coding (by themselves or by others) were accustomed to. This time they sought to *use history* to reconstitute the genesis of cognitive tools formerly considered "natural": socioprofessional categories, unemployment indicators, calculations of means or of correlations, drawing representative samples.

Towards a Social History of Statistical Coding

The focus of this work on the history of the instruments of description and of knowledge changed over time. At first, during the 1970s, it was a matter of relativizing (if not denouncing) techniques judged to be ahistorical. Next, in the canonical perspective of the "critique of sources", the emphasis was on associating historical reconstitution with information about the procedures whereby this was achieved, and thus about the genesis of the sources. But little by little these studies came to reveal the extent to which the setting up of systems of statistical recording goes hand in hand with the *construction of the State*. They led to the idea that the objects being studied were just as pertinent subjects of historical analysis as the numbers to which they were supposed to lead. Thus the succession of professional nomenclatures used by the different censuses is just as interesting as any eventual long-term comparison – which moreover it would be impossible to implement, given these changes in taxonomy. Or it is at the least interesting on other grounds, in that it encourages us to look at the evolution of the social links which make particular instances be treated as *equivalent*. We need to find an answer to this question before we can enquire *how many* cases there are in any given equivalence class.

From this point of view, classifications appear to be *conventions*, as is shown by comparisons between different countries, or between different historical periods in statistical descriptions of the social world. In order for international organizations concerned with statistical comparisons or for historians seeking to construct long series to be able to work, differences between nomenclatures are *problems* that must be *eliminated*, so that a grid can be constructed that is valid for all countries or at all times. The alternative comparative or historical attitude would, on the other hand, jealously preserve information about indigenous or contemporary taxonomies, instead of dissolving them in some ahistorical construction.

This change of viewpoint marks a clear rupture with the classical way in which quantitative social sciences use the numbers that are supposed to

express things that exist independently of the conventions establishing them. But this perspective is also very different from the relativist critique of statistics, which stigmatizes "quantofrenia". This critique has been formulated often since the 1960s, for example by ethnomethodologists (Cicourel, 1964). So doing, they throw the baby out with the bathwater. Here again, we can draw inspiration from the Durkheimian tradition, which proscribes dissolving into a total relativism whereby, since nothing is ever equivalent to anything else, the social sciences are reduced to particular, incomparable, case studies. According to this tradition, social facts can be treated as things *only* if they are *hardened* in one way or another (Thévenot, 1984) – as are institutions, laws, religion, customs. This is the reason why Durkheim began his study of "society" by looking at hard facts and was wary of the subjective and of unstable facts. This suggests the second feature common to research on the history of social coding: we have to study the extent to which the "things" described by statistics are solid and "hold together" (Latour, 1987). They are solid to the extent that they are linked to hard social facts: institutions, laws, customs etc. (Héran, 1984). So doing, one defines a position as distant from the relativist viewpoint defended by the ethnomethodologists as from that of the militant positivists, for whom there are "objective facts" that you just need to count up in order to silence idle ideological polemic.

Coding, Equivalence and Equity

Over the past twenty years, various studies have been conducted in more or less the spirit just outlined. They have dealt with the nomenclatures of consumption (Boltanski, 1970), branches of industry (Guibert, Laganier, Volle 1971), socioprofessional categories (Desrosières, 1977; Desrosières and Thévenot, 1988), studies of social mobility (Thévenot, 1976), school statistics (Briand, Chapoulie and Peretz, 1979), national accounting (Fourquet, 1980), industrial statistics (Volle, 1980) and unemployment (Salais, Baverez and Reynaud, 1986). Further, two volumes on the history of statistics have been published by the French Bureau of Statistics (INSEE, 1977; INSEE, 1987).

One significant consequence of these various pieces of research has been to highlight ideas of *coding* and *equivalence*. A coding is a conventional decision to construct an equivalence class between diverse objects, the "class" being judged more "general" than any particular object. A precondition for this is the assumption that these objects can be

compared. This is by no means apparent. Recent work on the history of probability theory (Coumet, 1970; Hacking, 1975; Daston, 1987b) can be read as posing the question: under what conditions, social and cognitive, can chance events or repeated sensations be compared? The oldest tradition dealing with this taxonomic problem no doubt goes back to *jurisprudence* (Serverin, 1985): to judge and to code both come down to classifying a *case* in a legal category, or, in legal terms, to *qualifying* it. Thus the tight link between the political and cognitive dimensions of coding becomes apparent: the legal category refers at once to the king and to knowledge. Coumet (1970), looking at how "the theory of chance was no chance development", speaks of "complicity between the judge and the geometer'.

Thus there is a link, sometimes overlooked, between *equity* and *equivalence.* It is overlooked because the social sciences came into the world in the nineteenth century through the act of cutting the umbilical cord tying them to seventeenth- and eighteenth-century political philosophy (Boltanski and Thévenot, 1987) and by progressively severing the links between the *prescriptive* and *descriptive* aspects of these disciplines – as Lorraine Daston (1987b) has shown with respect to the "end of the classical era of probability theory". At first the comparability and the equivalence between objects were less questions of knowledge than of *justice* – for example with respect to the law governing market exchange. When Adam Smith founded the discipline of economics, a precondition for the existence of a central market was not only the existence and the uniqueness of a system of *prices*, in general mentioned, but also, more often forgotten, the existence and uniqueness of a system of *goods*, subject to a common definition (Eymard-Duvernay, 1986; Boltanski and Thévenot, 1987). Before the birth of political economy as a science, the question of the definition and equivalence of goods were posed in cases where *conflict* arose, and judges had to decide if an item were of satisfactory *quality.*

Thus a social history of the creation of equivalence turns the spotlight on the figure of the judge, and behind him, the figure of the highest judge, the *king*. If he has to arbitrate conflicts between his subjects, it is in order to be able to *sum* their forces, to concentrate them on him (Latour, 1984), by raising armies or taxes. For that he has to *record* births, marriages and deaths (Edict of Villers-Cotteret of 1539) and to take a regular *census* of the population. English seventeenth-century *political arithmetic* (Graunt, Petty) drew the consequence of the administrative coding that constitutes

the demographic registers and proceeded to the first regular tallies. Counts of deaths enabled them to construct *tables of mortality by age*, and these then furnished the basis for *life-insurance* rates: the distribution of chance risks between people, or between periods of the life of a single person, was the occasion of many *equivalence conventions*, and, as a result, of totals in *numbers* (Daston, 1987a).

However, the term statistics itself did not emerge from this English context of quantification, but from eighteenth-century Germany, with Conring and Achenwall's "cameral statistics", which was a science of the description of the State in its most varied aspects. The history of the successive retranslations of the word "statistics" in itself furnishes a summary of the act of separating the (political) management of people from the (scientific) management of things – this separation leading to the autonomy of the various fields of knowledge. The word's meaning has changed so profoundly between the eighteenth century and the present that it calls to mind the axe whose head and handle have frequently been changed, but which is said to still be the same axe. We can summarize these changes into three phases. In the eighteenth century, statistics was an administrative activity for describing the *State*, in *literary* terms or eventually numerically. In the nineteenth century, it was only the *numerical* part of the description of the *State*. In the twentieth century, it refers to *mathematical* techniques for *numerical* analysis of data of *whatever type* – it can be applied equally to the State, to biology or to physics.

Addition, Action, Causality: The Case of Medicine

The meaning it has ended up with thus seems to bear little relation to its initial meaning. However, it is possible to find a common thread in this history. This is the link between *addition* (rendering equivalent) and *coalition* (action). It is possible to show that, from the German mini-state of the eighteenth century up to modern science, clusters are justified if they render action possible, if they create *things* which can act and which can be acted upon (a prince, a nation, a social class, an animal species, a microbe, a physical particle, a sickness, an unemployment rate). In each case it is necessary to transcend the contingency of particular cases and circumstances and to make *things which hold together*, which display the qualities of generality and permanence.

The only way of understanding the recurrent opposition in politics, in history and in science between on the one hand contingency, singularity

and circumstance and on the other hand generality, law, regularity and constancy is to ask: "for what purpose?" The question is not: "Are these objects *really* equivalent?" but: "Who decides to treat them as equivalent and to what end?" The debate is therefore endless. Some say: "I act as if these objects were equivalent, and I calculate means, I construct a time series of GNPs, etc.". Others object: "No, you are counting together things which *in fact* are not the same". They redivide the big entity into little ones, which are said to be *non-comparable*. This debate, and this "critique of the creation of equivalence classes" can be found in almost identical terms in many disciplines.

In the nineteenth century, discussion about the use of statistics, and in particular about the legitimacy of the use of *means* by Quetelet provide a good illustration of this relationship between ways of *thinking* the social world and of *acting* on it (Desrosières, 1988a). For example, we could look at the debate about the "numerical method" in medicine, between the 1830s and the 1860s (Murphy, 1981). Here we find not *two* but *three* distinct positions. Doctor *Louis* suggested the use of this method for comparing the efficiency of treatments, for example the use of bleeding or purging for treating typhoid fever. This met with two types of criticism. The first, originating in the old vitalist tradition of the eighteenth century, maintained, with *D'Amador*, that each patient was a *unique case*, not comparable with any other. He said that only a particular, personal and prolonged interview between doctor and patient enabled the former to come to an understanding of the case and how to treat it.

The second objection, in appearance close to the first in that it challenged statistical totalization, came from *Claude Bernard*. This latter, a symbol of modern science, was by no means a vitalist. But, he said, doctors cannot treat "by averages". In order to eliminate a disease completely, they have to find its *direct* causes. This hostility against statistics was linked to a determinist conception of microcausality, in terms of which probability and statistics were synonymous with *approximation* and lack of rigor. We could show that the three protagonists of this debate were right each after their own fashion, to the extent that their cognitive tools were coherent with the action that they took. These were the *hygienic movement* and preventive, *collective* social medicine for Louis (Lécuyer, 1982), *family* doctor in close, daily attendance for Amador and experimental medicine involved with the *technicization* of the clinic for Claude Bernard. Each of these positions still exists, and has its own coherence.

For its part, sociology, like Doctor Louis, embarked on a quest for

macrocausality. The foundation of such a probabilistic use of statistics in social science was Quetelet's importing an interest in *normal* distributions and the calculation of *means* from astronomy and the natural sciences into social science. This was decisive for developing a response to the question: "How can we make things which hold together?" In fact, in the natural sciences, the mean appeared in the theory of *measurement errors* (the height of a star, the diameter of the earth). This theory was used to calculate the mean of imperfect, distributed measurements according to a normal distribution, in order to estimate the best approximation of the "real thing". (Accordingly, this was called the *objective mean.*)

Then, by making an analogy between the normal shape of the graph of the distribution of measurement errors around a true value and that of the heights of conscripts around a mean (called a *subjective* mean), Quetelet postulated the existence of an *average man*, more *real* and more *general* than any particular individual.

Further, although the height of individuals varies greatly, its mean is remarkable *constant* from one year to the next. This provides support for the idea that there is a reality beyond that of contingent manifestations, which are themselves analogous to chance observations of an object external to the observer. Finally, from physical statistics of height or of weight one can pass to *moral* statistics: of marriages, crimes, suicides. Although for an individual these decisions are a matter of free will, and are thus unpredictable and chance events, for society as a whole, on the other hand, the *rates* of marriage, crime or suicide are remarkably constant. The *terrifying regularity of crime rates* contributed to founding the idea of a *thing which held together*, in this case the human propensity to kill.

Debates about the Use of Statistics in Social Science

This network of connections between astronomical measurements, the heights of a large number of men and marriage or suicide rates lent great force to the idea of macrosocial causality, and formed the basis of Durkheim's arguments in his *Suicide* for the existence of *social facts* radically distinct from individual psychological facts. It is true that Durkheim later modified his point of view, judging that the average man was a sorry case, with no morality: he doesn't want to pay his taxes or to

go to war (today, one would add: he is in favor of the death penalty). In this way the civic, moral man became differentiated from the average man referred to by statistics. This contradiction in Durkheim's own work is indicative of the tension between the two ways of constructing totality (Desrosières, 1988a).

But this Durkheimian challenging of the average man only appeared at the end of the century. Further, it was not in itself enough to sap the force of arguments based on macrosocial regularities. Thus the reasoning of astronomer and statistician Quetelet became very widely known throughout intellectual circles in Europe, thanks to the use made of it by English historian Henry Buckle. His monumental *History of Civilization in England* (1857) aimed to bring out underlying social tendencies, independent of individual wills, by using Quetelet's methods to transcend the sound and fury of daily contingencies. He contributed greatly to spreading Quetelet's ideas in sectors of opinion that seemed a priori impermeable to statistics or demography, by furnishing a reassuring mental tool-kit for those who feared the political upheavals resulting from the French Revolution or from the industrialization of England. But he was very severely criticized by those, particularly in Germany, who saw in the creation of equivalence classes that forms the basis of statistics a product of the abstract universalism of the Franco-English Enlightenment. These critics insisted on the incommensurability of situations defined by *local* cultures and traditions. This criticism was, for example, made by economists belonging to the German "historical school" of the end of the century, and was at the root of methodological debates that Max Weber later participated in.

From the preceding discussion, we can see that Quetelet's model could be integrated into a *holistic* model (this is what Durkheim did initially), but could also be criticized in the name of that very holism (this is what Durkheim did later, as did the German historical school). It is certain that by his insistence on the *average* features of social groups, and by the fact that he only looked at the *distribution* of the individual features analyzed, Quetelet opened himself up to this double reading, and one could find a basis for the two opposed criticisms in his work.

And so in the 1890s modern mathematical statistics was born, with the work of English eugenicists Galton and Pearson (MacKenzie, 1981). They at once inherited the formalism of the normal law, and completely altered its use, since henceforth the accent would be on *individual* features, their *distribution, correlation* and *classification* (Thévenot,

1987) – this perspective being largely absent from Quetelet's work. This opened the door to completely new ways of *making things hold together.* Variables that are correlated, or of which some are fully *explained* by others using a *regression* model, created things that were incomparably more solid than those resulting simply from a Gaussian distribution (which, moreover, little by little lost its power to fascinate). The word "explain" then took on a new meaning than that which historians were accustomed to. Henceforth these latter, however reluctant they were to use these new techniques, would have to vacillate the two meanings of this word, which could only be understood in context.

Further, we should note that Karl Pearson himself, the inventor of these techniques, only spoke of *co-occurrences* and denied any claim to causality, which he considered a metaphysical concept. Indeed, outside of his statistical work, he had published a significant work in the philosophy of science in 1891. *The Grammar of Science* does not contain any mathematical theory, but is a treatise, influenced by Ernst Mach's criticism of empiricism, on the impossibility of uncovering the ultimate nature of things. His argumentation is more subtle than Quetelet's crude realism. Like the eighteenth-century sensualists before him, he said that we cannot know anything about *real things*, we can only know impressions on the brain, which constitute *perceptual routines.* From this point of view, the regularities of correlations or regressions appear as the only possible "laws" capable of describing and measuring these "routines", which therefore constitute the *things* of the new sciences.

Pearson's book, forgotten today, was published in three editions, and had a marked influence on the founders of quantitative social science in America at the beginning of the twentieth century, such as Giddings, Ogburn and Lazarsfeld. It was published in French by Alcan in 1911, in a translation by Lucien March, director of the French Bureau of General Statistics – the forerunner of the INSEE, the current Bureau of Statistics – even though it was a purely philosophical and non-technical work. But it is striking that his very modern, relativist approach never extends to a consideration of the *socially constructed* nature of the things described and of the classification categories used – be it even in terms of *language*, as Cassirer, for example, would do a few years later. We may wonder what would have become of empirical social science if at the turn of the century Durkheim, Pearson and then Cassirer could have become acquainted with one another. And this not only because, as Selvin (1976) has maintained, Durkheim did not perceive the advent of mathematical

statistics, but also because Pearson did not perceive the importance of language and of society for the construction of reality.

Within France, Durkheim's influence on the spread of the use of statistics in social science (apart from through *Suicide*, which was little read and cited by statisticians themselves) came through two of his disciples, Simiand and Halbwachs. The former's doctoral thesis concerned salaries and contained some significant statistical analyses (1908). The latter's thesis, on the living conditions of working-class families, was largely based on an examination of budgets, such as Engel and Le Play had carried out. Further, his complementary thesis was about "Quetelet and the Average Man". However, this academic activity, which used very novel methods, made little impression on official and administrative statistics – save during one period, the First World War.

Periods of war, which involve an extreme concentration and planning of the economy and of the management of people (soldiers and workers) are favorable occasions for meetings and alliances inconceivable in other times. This was the case in the cabinet of the socialist Minister of Munitions Albert Thomas, who gathered together various high-ranking scientists: mathematicians, statisticians, economists and Durkheimian sociologists (Desrosières, 1985). Thus developed an intensive collaboration between scientists and political and administrative personnel, one which disappeared after the beginning of the 1920s (Kuisel, 1981). Simiand and probability theorist Emile Borel together approached the President of the Republic Millerand, to try to convince him to create a large-scale Bureau of Statistics, thereby continuing the planning work carried out in Albert Thomas' cabinet.

However, this approach was unsuccessful, and it was only twenty-five years later in 1945 (in the second postwar period) that the INSEE (Bureau of Statistics), the Planning Commission and several scientific and administrative research institutes drawing largely on statistical methods were created. At the same time, the American Census Board began to implement three major innovations permitting the development of large-scale statistical/administrative institutes: sampling surveys, national accounting and computer science. The standardized procedure implicit in these three techniques would make a major contribution to fashioning "hard things" such as unemployment rates, the gross national product and the balance of payments (Duncan and Shelton, 1978; Anderson, 1988).

"Cadres", "Professionals", "Angestellte"

The project of producing a social history of statistical coding can be illustrated by two examples, which are treated in more detail elsewhere: *socioprofessional classifications* and *surveys* of representative samples.

The social categories described and measured by statisticians, demographers and sociologists are constructed *things*. Nothing shows this better than a comparison between the taxonomies used for example in France, Great Britain, the USA and Germany. These taxonomies are very different, and it is difficult to translate them from one language into the other. Thus where the French speak of *cadres*, in Great Britain and in the USA they talk about *professionals* and *managers* (which are distinct categories), and in Germany about *Angestellte* (private-sector employees) and *Beamte* (public servants). These groups are not the same, but they do nevertheless intersect at various points. An examination of their current definitions hardly helps us get a hold on the logics underlying such diverse representations. Only a social history of statistical taxonomies allows us to discern these logics. Here we will look only at the region of social space corresponding to the French *cadres*. We will draw on research involving the four countries: Boltanski (1987) for France, Szreter (1984) for England, Freidson (1988) for the United States and Kocka (1981) for Germany.

French *cadres* are the result of intensive mobilization and representational efforts concentrated on a kernel of engineers carried out at the end of the 1930s, following the 1936 General Strike and the Matignon settlements. German *Angestellte* are the outcome of similar work, carried out fifty years earlier, at the time when Bismarck was establishing the first great social security laws and when *non-manual* workers were seeking to distinguish themselves from the workers (*Arbeiter*), who were highly organized both politically and into trade unions by the social democratic movement (Kocka, 1981). This is why they include not only cadres but also employees (of the private sector alone, since public servants are, for their part, covered in German bureaucracy by their own particular statutes).

Anglo-American *professionals* are defined by a level of higher *education*, and their pre-eminence in English and American nomenclatures is inherited from the biologically based meritocratic theories of the English eugenicists of the beginning of this century (Szreter, 1984). These theories are based as much on an implicit denuncia-

tion of the old aristocracy as of the view of bourgeois business people that wealth is a sign of superiority. In the United States (Freidson, 1988), the central point was the moral, civic and even religious justification of the "professions" (in the English sense of the word) and the affirmation of their utility and importance for the entire nation. This importance was justified in terms of their technical competence, without there being any direct intervention from the State to guarantee it (this is the major difference with the French case). Parsons is the theorist of this conception of *professions*.

This summary comparison indicates the extent to which the social construction of equivalences is not the same from one country to another, and also that, within a given country or historical period, it is extremely difficult to *see* that construction. Thus in France before the 1930s, no one spoke in terms of *cadres* (in his *Psychology of Social Class*, published in 1938, Halbwachs never used the word), whereas today this group is completely natural, to the point that no one comments on the strangeness of the idea of *counting* it.

Sewell (1980) carried out an analogous enquiry on the group of *ouvriers* in France. He situated the moment when this group began to be thought of as such precisely in the period 1832–1834 (after the insurrections at Lyon and Paris). Before then, no one thought of counting together on the one hand the *compagnons* (craftsmen) of traditional professions and on the other hand the *manoeuvres* (laborers), workhands and vagabonds who could be found right at the bottom of nomenclatures – for example, in the 1800 prefects' statistics (Bourguet, 1988). Thus, when Marx began to write, the working class was still a *taxonomic neologism*.

Political Representation and Statistical Representativity

The few cases looked at above have demonstrated that the visibility of a group is a result of social or political mobilization, and of particular historical circumstances: the 1830 insurrections for the French *workers*, the establishment of Bismarck's social laws in 1880 for the German *Angestellte*, the debate about the inheritance of intelligence in 1900 for the English *professionals* and the organization of collective negotiations between employers and employees in 1936 for the French *cadres*. This work of representation and of construction of a common language is itself linked to the construction of the State, as demonstrated by the comparison between these four countries, wherein this construction has taken such

different forms. Further, this social *reification* (the making of a thing) is also closely analogous to the work of taxonomic statisticians who create categories to which they give *titles* which represent them. We will now look more closely at the link between these diverse form of representations (political, statistical and cognitive), taking current French socioprofessional nomenclature as our empirical base. This nomenclature is caught between two quite distinct logics (Desrosières and Thévenot, 1988).

Classifications used in French censuses in the nineteenth century show the persistence of the vocabulary of guilds and crafts, despite their abolition in 1791 (Sewell, 1980). The *corps* constituted elementary social groups, and the distinction between master and craftsman did not have the same meaning as that between employer and employee. The crafts formed the first stratum of the classification, and we can still find many traces of them in the current one. Things changed towards the end of the nineteenth century, when the labor legislation emerged. This clearly defined the statute of salaried employee (*salarié*) – and, thereby, that of the unemployed – inconceivable before the statute of salaried people (Salais, Baverez and Reynaud, 1986). During this period, administrative and political measures were taken which enabled the definition of the current categories: active population, unemployment – and as a consequence productivity, and all the other measurements which macroeconomics is based on. A clear distinction between salaried employees and employers still exists in French classifications, and this distinguishes them from their English or American homologues, and from the international nomenclature of the International Labour Office.

The third stage of this history is marked by the diffusion during the 1930s and 1940s of *collective agreements* underwritten by the State and negotiated by the unions of salaried employees and employers. At this time, a new vocabulary, originating in the metallurgical industry of the 1920s, paved the way for a coding of workers' qualifications: skilled workers (P1, P2, P3), semi-skilled workers (OS1, OS2), unskilled workers and terms relating to the training period (also called *Parodi categories*). These latter served as a *partial* substitute for the old vocabulary of trades. At the same time, the group of *cadres* made its appearance, and elections into the new *comités d'entreprise* (co-determination bodies) were made in three categories: cadres; employees, technicians, supervisors (ETAM); and workers. Thus from the beginning of the 1950s any employee knew which category he or she belonged to,

since this had important consequences in terms of social security, retirement and choice of profession. This indeed constitutes a *hard fact*, a thing that holds together and is the same for all. This does not, of course, imply that there are no battles at the borders of these groups. One of the difficulties of making comparisons between countries is that it is very difficult to know in each case the *relative stability* of the borders – a question rarely posed by statisticians.

Current French classification thus displays a complex admixture of these diverse historical strata, and for this reason it is sometimes denounced as a non-scientific hotchpotch, with no "single criterion" to structure it. However, like genealogical layers, these strata bear witness to the past, and at the same time they reflect the *current* situation, since all these categories and frontiers are still simultaneously operative. The reworking of the nomenclature in 1982 presented an opportunity for a systematic investigation of these historical components (comparable to the exhumation of medieval Paris during the construction of a new metro line). It ended up preserving them to the extent that they were still active, and not replacing them by a *unique criterion* as some wished.

The Body and the Institution: Two Types of Regularity

French socioprofessional nomenclature comports about thirty categories, and since the 1950s there has been a lot of research done using them. When in the 1960s French statistician Jean-Paul Benzecri's techniques for analyzing correspondences were developed, they were systematically applied to tables crossing social categories with diverse variables. This enabled the production of very stable *multi-dimensional* representations of social space. Thus, unlike Anglo-American sociology and demography, which use a classification system whose unidimensionality is written in from the start (Szreter, 1984), French social scientists employ a more complex tool for the accumulation of knowledge. The second dimension of this space, which, to put it very schematically, distinguishes cultivated people from rich ones, proves to be highly pertinent for interpreting for example voting patterns, religious practices and opinions on moral questions, etc. (Bourdieu, 1984).

The relative stability of this configuration with respect to whatever variable is being studied (consumption, leisure, votes, marriage, opinion, etc.) can make a very favorable impression, since it offers, for example, the hope of *predicting*, and thus, for once, of producing a thing which

holds together. However, this sense of satisfaction is very similar to that which enchanted Quetelet and his contemporaries: there are many macrosocial regularities, knowledge of which gives one a strong feeling of power. Quetelet discerned the divine plan therein, and, in his inspired moments, Benzecri rediscovered Quetelet's voice and sung the permanence and divine order of nature, which revealed balanced and stable symmetries in clouds of points displayed by computer programs. If one does not accept God (at least in this context), then one can debate the foundations of these regularities. Here again, the Durkheimian tradition provides some help. Where Durkheim insisted on *institutionalized* social facts, like law or religion, his nephew Mauss turned his attention to *bodies* and to the incorporation of habits and of schemas organizing practice learnt unconsciously by the child in its family. This is the same as Bourdieu's *habitus* (1977) – the word's etymology goes back to the body. From this perspective, the body and its permanence are the ultimate explanations of social regularities (this again is close to Quetelet's position). The other way of interpreting these regularities is to link them to institutions that are socially and historically standardized, and thus capable of *holding things together*. Taken together, these alternatives provide a conjugation of permanence inscribed on the one hand in bodies and on the other in historical laws and institutions. Doubtless such a conjugation affords us a first explanation of statistical regularities.

It should be possible to show that there is a correspondence between these two modes of transmission of permanence (by the body in the family and by codified institutions) and the two ways of *constructing categories* described by cognitive psychology and also to be found in the definition of socioprofessional categories. The first refers to *typical cases*, which other cases *resemble*, and the second to a *criterion* (Boltanski and Thévenot, 1983). This opposition also helps us to understand what among the methods of enquiry used in social science distinguishes on the one hand *monographs* that relate cases adjudged *typical* and whose generalization is based on the idea of their exemplary nature and on the other hand studies that report the findings of *surveys* of representative samples, and whose generalization is based on probabilistic reasoning (Desrosières, 1991).

The Part for the Whole: How to Generalize?

The social history of the tools of enquiry and its implicit modes of generalization enables us to distinguish between several ways of considering the management of the social world – and in particular the management of poverty, to which the empirical social sciences have been linked since their origin. Surveys based on representative samples only appeared at the end of the nineteenth century with the Norwegian Kiaer, and even then in a form more intuitive than formalized. The first calculations of the confidence interval by Bowley in England date from 1906, and a detailed formalization of stratification techniques was only developed in 1934, by Neyman.

However, enquiries based on a small number of people were carried out very early in the nineteenth century, by engineers (Le Play, Cheysson) whose mathematical culture was more than sufficient to use the probability theory necessary for taking a sample. Laplace had already used this theory in the eighteenth century to estimate the population of France (Bru, 1988), but this pioneer work was not to bear fruit for a century.

The fact that methods of probabilistic surveys did not come into widespread use until scarcely more than fifty years ago shows that the discovery and the utilization of a technical innovation involve conditions that are inseparably social and cognitive. Before inventing the solution to the problem, one had to invent the problem itself – that is to say the constraint of *representativity*, in the sense henceforth given this word by statisticians. Now this concern, expressed in terms of a relationship of similarity, for certain well-defined elements, between the part and the whole, comes a long time after the development of censuses (Quetelet in about 1830) and of monographs (Le Play, about the same time). The history of social surveys gives the impression that we have passed directly from a period where the question of representativity was never considered to another where it is never discussed because it is so obvious.

Kiaer's new "representative method" was discussed by the International Institute of Statistics between 1895 and 1903. This debate did not focus on the constraint of representativity as such, but on the questions of deciding on the one hand if it was legitimate to "replace the whole by the part" (comparison with the census) and on the other hand if, so doing, one produced better results than Le Play-style monographs, which were then still highly valued. However, this "better" was not decided in terms of

representativity as the condition of a *precise measure*, but on the possibility of describing a *diversified* population.

It seems that between 1895 and 1935 the norms presiding over legitimate descriptions of the social world were completely changed – at least with respect to the possibility of *generalizing* observations of a part of it over society as a whole. Why? We can find an initial response to this question in a comparison of the social context in which this technique appeared both in Norway and in England at the turn of the century with that surrounding the monographs of the earlier period.

All these enquiries involved looking at *poverty* and at the problems of the working class resulting from industrialization and urbanization. In the nineteenth century, these problems were treated at the town or parish level. Enquiries were *local*, as were proffered solutions to the problems of poverty, founded as they were on direct relationships between rich and poor. Le Play's monographs, which resulted from direct contacts between surveyors and families surveyed, were in tune with this type of solution. Their generalization rested on the concept of *exemplarity*, and the few cases described (a handful) were taken to be *typical*. Society was thought of as a whole, in a holistic perspective: knowledge about one case provided knowledge about them all. The typical case was taken to be representative of the whole if one had personal contact with that case.

However, at the end of the nineteenth century, purely local and municipal solutions to the problems of poverty proved insufficient. To the extent that the causes of hardship were national (the economic crisis of the period 1875–1890), they required national measures underwritten by laws and no longer based on local charity. These laws were at the root of the welfare state. English historian E.P. Hennock (1987) has described the passage from Booth and Rowntree's local enquiries in the 1880s to Bowley's national survey after 1900. He has clearly brought out how the fact that purely local solutions to poverty appeared henceforth impossible led the Poor Law Commission of the House of Commons to propose a *national* enquiry. Bowley organized this latter on the basis of a *representative sample*, and formulated the notion of the *confidence interval* for the occasion. Similarly, a few years earlier (in 1895) the Norwegian Kiaer had an inkling of the notion of representativity when, participating with his country's parliament in the preparation of new social security laws, he was asked to carry out a similar survey. In these different cases, we see the continuity between *management* and *description* of the social world, both being based on standardization and codification.

It is because people had already been coded according to a criterion and thus supposed to be *equivalent* that the probabilistic model of the *urn* from which one draws black or white balls could henceforth be evoked and applied. In the former model of knowledge, based on a holistic view of society, people were not comparable and it was not possible to imagine placing them in an urn. A significant indication of this is the fact that Le Play, who amassed a great number of *budgets* of working-class families, never had the idea of *adding them up*, in order to take averages and *compare* them in order to deduce general laws. At the same time, the German Engel carried out this type of calculation on similar types of data and deduced his famous "Engel's law" linking consumption and income. A few years later, Halbwachs himself made similar comparisons, in his thesis on "the working class and living standards".

Social Science and Statistical Forms

In the techniques of *objectification*, the social sciences and history found a powerful tool for backing up the basic insights of the Durkheimian school and the Annalists (Marc Bloch, Lucien Febvre and Fernand Braudel). The *social group* for the one and *long duration* for the other displayed *regularities*, in opposition to *particular instances* or conjunctural *events*, which were contingent and unpredictable. This use has not failed to invoke criticism and inspire debate – for example from those who look to history to find arguments for denouncing objectifying social science, which they take to attack human freedom and to be blind to crises and ruptures. This tension comes out particularly clearly in the case of social history. The latter is itself the product of a mingling of the two traditions evoked above, by way of the great names of Simiand and Labrousse, who both did a lot to introduce statistical methods into a discipline of history that no longer wanted to concern itself solely with great men, treaties and events.

In the critical mood that held sway in the 1960s and 1970s, one axis of the denunciation of statistics was the idea that, as the very etymology of the word shows, statistical activity is in league with the *State*, and that it therefore reflects the point of view of those in *power* in that it is totalizing and reductive of diversity. Looked at in this way, quantification often presents a completely ambiguous face to contemporary social science: on the one hand it constitutes an "obligatory passage point" (in Latour's (1984) sense) and is of extreme social import, but on the other hand it is

implicitly considered to overlook the essential, to be impoverished, simplificatory and to explain nothing.

In this paper, I have tried to show that the only way of getting out of the quasi-magical juxtaposition of these two necessarily opposed modes of thought is to take the act of coding and the construction of equivalence classes seriously. In order to do so, the *social history of cognitive forms* becomes a pressing task. By treating statistical forms as *things which hold together*, we have indicated all the *work* which was needed in order to make them solid (Thévenot, 1984). So doing, we have not, as has sometimes been the case, sought to denounce the falsity of these things by referring to the historical context of their birth. Rather we have tried to draw attention to the fact that these cognitive schemas are linked to political categories in a much more basic way than certain single-minded criticisms of the 1970s suggest, to the extent that one can reconstitute a multiplicity of coherent cognitive and political schemas (Boltanski and Thévenot, 1987).

Thus the question of the relationship between politics and statistics cannot be reduced to an evocation of the eventual manipulation of facts by "those in power", nor to modern forms of resistance against control and computer databases as seen for example in West Germany. An example of coherence between political and statistical forms could be developed by working from the concept of *representation* in its three senses: cognitive, political and statistical. In the works on social classification that we have referred to, the authors have tried to look at these three dimensions together: (1) What spontaneous images of class are used in daily life?; (2) By what political processes do social groups acquire their historical solidarity?; and finally (3) What does an average represent? How is the representativity of a sample validated? How can we generalize?

In recent years there has been a lot of work on the history of the mathematical techniques of statistics (Benzecri, 1982; Stigler, 1986; Kruger *et al.*, 1987), and the social context of these products has also been studied (MacKenzie 1981; Porter, 1986; Hacking, 1990). However, few have looked at the way in which statistics, taken as a set of operations of *recording* and *coding* (and not only as a mathematical tool) informs social science not only in terms of its numerical data but also and above all in terms of the specific schemas which transform devices into things which hold together. For example, the passage from the problematic of the mean in Quetelet to Galton and Pearson's problematic, which

combines distribution, correlation and linear regression, constituted a significant advance along this path, by allowing a number of elementary things (variables) to hold together in *probabilistic models*. Procedures of the construction of classes (taxonomies) and enquiries based on representative samples for their part provided the wherewithal for constructing things that social science can no longer do without.

In each of the four constructions just referred to (mean, correlation, social categories and sampling), equivalence conventions necessary to the development of cognitive schemas were linked, in very diverse ways, to State procedures or to attempts at mass mobilization. These schemas thus came into being as ways of holding things together. These things, in turn, allow us inextricably to think about the social world and to act on it.

Note

1. This text is a synthetic presentation of various works referred to in the references. For the concept of *things which hold together* it owes a lot to Luc Boltanski (1987) for his study of the *crystallization of the group of managers*, to Laurent Thévenot (1984) for his analysis of *investments of form* and to Michel Callon (1986) and Bruno Latour (1987) for their research on the *construction of scientific facts*. Further, the work by Boltanski and Thévenot (1987) on *economies of scale* greatly contributed to my understanding of the multiplicity of ways that there are of making things equivalent and of generalizing, and thus of *representing* things – this is particularly important for statistical work.

References

Anderson, M.J. (1988): *The American Census: a social history*, New Haven and London: Yale University Press.

Benzecri, J.P. (1982): *Histoire et préhistoire de l'analyse des données*, Paris: Dunod.

Boltanski, L. (1970): 'Taxinomies populaires, taxinomies savantes: les objets de consommation et leur classement' in *Revue Française de Sociologie*, **XI**, 34–44.

Boltanski, L. (1987): *The Making of a Class: "cadres" in French society*, Cambridge: Cambridge University Press.

Boltanski, L., and Thévenot, L. (1983): 'Finding one's way in social space: a study based on games' in *Social Sciences Information*, **22** (4–5), 631–680.

Boltanski, L., and Thévenot, L. (1987): *Les économies de la grandeur*, Les Cahiers du Centre d'Etudes de l'Emploi, Paris: Presses Universitaires de France.

Bourdieu, P. (1977): *Outline of a Theory of Practice*, Cambridge: Cambridge University Press.

Bourdieu, P. (1984): *Distinction: a social critique of the judgement of taste*, Cambridge, MA: Harvard University Press.

Bourguet, M.N. (1988): *Déchiffrer la France: la statistique départementale à l'époque napoléonnienne*, Paris: Editions des archives contemporaines.

Briand, J.P., Chapoulie, J.M., and Peretz, H. (1979): 'Les statistiques scolaires comme représentation et comme activité' in *Revue Française de Sociologie*, **XX**, 669–702.

Bru, B. (1988): 'Estimations Laplaciennes. Un exemple: la recherche de la population d'un grand Empire, 1785–1812' in J. Mairesse (ed.), *Estimation et sondages: cinq contributions à l'histoire de la statistique*, Paris: Economica.

Bulmer, M., Bales, K., and Sklar, K.K. (forthcoming 1991): *The Social Survey in Historical Perspective*, Cambridge: Cambridge University Press.

Callon, M. (1986): 'Eléments pour une sociologie de la traduction. La domestication des coquilles Saint-Jacques et les marins-pécheurs dans la baie de Saint-Brieuc' in *L'année sociologique*, **36**, 169–208.

Cicourel, A. (1964): *Method and Measurement in Sociology*, New York: Free Press of Glencoe.

Coumet, E. (1970): 'La théorie du hasard est elle née par hasard?' in *Annales. Economie Société Civilisation*, **9** (mai-juin), 574–598.

Daston, L.J. (1987a): 'The domestication of risk: mathematical probability and insurance, 1650–1830' in L. Kruger *et al.*, *The Probabilistic Revolution: Volume 1, Ideas in History*, Cambridge MA: MIT Press.

Daston, L.J. (1987b): 'Rational individuals versus laws of society: from probability to statistics' in L. Kruger *et al.*, 1987, *The Probabilistic Revolution: Volume 1, Ideas in History*, Cambridge MA: MIT Press.

Desrosières, A. (1977): 'Eléments pour l'histoire des nomenclatures socioprofessionnelles' in *Pour une histoire de la statistique, tome 1, contributions*, Paris: INSEE, 155–231 (republished in 1987, Paris: INSEE, Economica).

Desrosières, A. (1985): 'Histoires des formes. Statistiques et sciences sociales avant 1940' in *Revue Française de Sociologie*, **2**, 277–310.

Desrosières, A. (1988): 'Masses, individus, moyennes: la statistique sociale au XIXème siècle' in *Hermes* (ex *Cahiers Sciences Technologie Société*), **2**, 41–66.

Desrosières, A. (1991): 'The part in relation to the whole: how to generalize? The prehistory of representative sampling' in Bulmer, M., Bales, K., and Sklar, K.K., *The Social Survey in Historical Perspective*, Cambridge: Cambridge University Press.

Desrosières, A., and Thévenot, L. (1988): *Les catégories socioprofessionnelles*, Paris: La Découverte.

Duncan, J.W., and Shelton, W.C. (1978): *Revolution in United States Government Statistics, 1926–1976*, Washington: U.S. Department of Commerce.

Eymard-Duvernay, F. (1986): 'La qualification des produits' in Salais, R., and Thévenot, L. (eds.), *Le travail, marchés, règles, conventions*, Paris: Economica.

Fourquet, F. (1980): *Les comptes de la puissance – Histoire de la comptabilité nationale et du Plan*, Paris: Encres (Recherches).

Freidson, E. (1988): *Professional Powers: a study of the institutionalization of formal knowledge*, Chicago and London: University of Chicago Press.

Guibert, B., Laganier, J., and Volle, M. (1971): 'Essai sur les nomenclatures industrielles' in *Economie et statistique*, **20**, 23–26.

Hacking, I. (1975): *The Emergence of Probability*, Cambridge: Cambridge University Press.

Hacking, I. (1990): *The Taming of Chance*, Cambridge: Cambridge University Press.
Hennock, E.P. (1987): 'The measurement of poverty: from the metropolis to the nation, 1880–1920' in *Economic History Review*, 2nd series, **XL**, 2, 208–227.
Héran, F. (1984): 'L'assise statistique de la sociologie' in *Economie et Statistique*, **168** (juillet-août), 23–26.
INSEE (1977): *Pour une histoire de la statistique, tome 1, contributions*, Paris: INSEE.
INSEE (1987): *Pour une histoire de la statistique, tome 2, matériaux*, Paris: INSEE, Economica.
Kocka, J. (1981): 'Class formation, interest articulation and public policy: the origins of the German white-collar class in the late nineteenth and early twentieth centuries' in Berger, S. (ed.), *Organizing Interests in Western Europe*, Cambridge: Cambridge University Press.
Kruger, L. (1987): 'The slow rise of probabilism: philosophical arguments in the nineteenth century' in L. Kruger *et al.*, *The Probabilistic Revolution: Volume 1, Ideas in History*, Cambridge MA: MIT Press, 59–89.
Kruger, L., Daston, L., and Heidelberger, H. (eds.) (1987): *The Probabilistic Revolution: Volume 1, Ideas in History*, Cambridge MA: MIT Press.
Kuisel, R.F. (1981): *Capitalism and State in France*, Cambridge: Cambridge University Press.
Latour, B. (1987): *Science in Action: how to follow scientists and engineers through society*, Cambridge, MA: Harvard University Press.
Lécuyer, B.P. (1982): 'Statistiques administratives et statistique morale au XIXème siècle' in *Actes de la journée "Sociologie et statistique"*, Paris: INSEE, Société Française de Sociologie, 155–165.
MacKenzie, D. (1981): *Statistics in Britain, 1865–1930; the social construction of scientific knowledge*, Edinburgh: Edinburgh University Press.
Mairesse, J. (ed.) (1988): *Estimation et sondages: cinq contributions à l'histoire de la statistique*, Paris: Economica.
Murphy, T.D. (1981): 'Medical knowledge and statistical methods in early nineteenth century France' in *Medical History*, **25**, 301–319.
Pollak, M. (1979): 'Paul F. Lazarsfeld, fondateur d'une multinationale scientifique' in *Actes de la recherche en sciences sociales*, **25** (janvier), 45–59.
Porter, T.M. (1986): *The Rise of Statistical Thinking, 1820–1900*, Princeton: Princeton University Press.
Salais, R., Baverez, N., and Reynaud, B. (1986): *L'invention du chômage*, Paris: Presses Universitaires de France.
Selvin, H. (1976): 'Durkheim, Booth and Yule: the non-diffusion of an intellectual innovation' in *Archives Européennes de Sociologie*, **XVII**, 39–51.
Serverin, E. (1985): *De la jurisprudence en droit privé; théorie d'une pratique*, Lyon: Presses Universitaires de Lyon.
Sewell, W.H. (1980): *Work and Revolution in France: the language of labour from the Old Regime to 1848*, Cambridge: Cambridge University Press.
Stigler, S.M. (1986): *The History of Statistics: the measurement of uncertainty before 1900*, Cambridge and London: The Belknap Press of Harvard University Press.
Szreter, S. (1984): 'The genesis of the Registrar General's classification of occupations' in *British Journal of Sociology*, **XXXV** (4), 522–546.

Thévenot, L. (1976): 'Les enquêtes Formation. Qualification professionnelle et leurs ancêtres français' in INSEE (1987), *Pour une histoire de la statistique, tome 2, matériaux*, Paris: INSEE, Economica, 117–165.

Thévenot, L. (1984): 'Rules and implements: investments in forms' in *Social Science Information*, **23** (1), 1–45.

Thévenot, L. (1987): 'Forme statistique et lien politique: éléments pour une généalogie des statistiques sociales' (note INSEE Unité de Recherche, no. 112/930 of 9 April, 1987 – forthcoming).

Turner, S. (1988): 'Giddings as the Father of American Sociology', paper presented to the Conference of the I.S.A. Research Committee for the History of Sociology, Madrid, May 1988.

Volle, M. (1982): *Histoire de la statistique industrielle*, Paris: Economica.