

Andrea Bonaccorsi *Editor*

The Evaluation of Research in Social Sciences and Humanities

Lessons from the Italian Experience



Springer

The Evaluation of Research in Social Sciences and Humanities

Andrea Bonaccorsi
Editor

The Evaluation of Research in Social Sciences and Humanities

Lessons from the Italian Experience

 Springer

Editor
Andrea Bonaccorsi
DESTEC
University of Pisa
Pisa, Italy

IRVAPP-FBK
Trento, Italy

ISBN 978-3-319-68553-3 ISBN 978-3-319-68554-0 (eBook)
<https://doi.org/10.1007/978-3-319-68554-0>

Library of Congress Control Number: 2017959886

© Springer International Publishing AG 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

What Does the Italian Experience in Research Evaluation Tell to the International Debate on SSH?

The widespread diffusion of research evaluation practices is part of the new higher education and research landscape at all levels. While scholars active in those disciplines that are conventionally labelled science, technology, engineering and mathematics (STEM) find themselves relatively at ease with this reality, their colleagues in social sciences and humanities (SSH) raise a number of theoretical and practical issues.

Research evaluation is an activity that has two, not just one, sources of legitimation. On the one hand, the parliament and the government in many countries have created dedicated structures to carry out research evaluation and actively make use of their results. This is a legal and institutional form of legitimation. It comes from the legitimate democratic authority of modern states. On the other hand, however, scientific and academic communities create and manage a different source of legitimation, based on scientific recognition and expertise.

To borrow the influential theory of power in organizations developed in 1959 by French and Raven, there are five forms of power: legitimate, reward, coercive, expert and referent. The first three come from organizational sources, the latter two from personal sources. The evaluation may use legitimate power from the parliament and the government and may also have reward and, to a certain extent, coercive power if its results are used for the allocation of results, but its legitimation in the academic community is mostly based on expert and referent power—in other words, whether evaluation is perceived as based on sound principles of knowledge and whether it is respected and esteemed.

For this reason, it is mandatory for those involved in research evaluation to open a never-ending dialogue with the scientific communities, in order to gain legitimation from a bottom-up and trust-based process, not only from the (inevitably top down) institutional procedures. Communities in SSH raise serious problems that deserve respectful listening, close scrutiny, a lot of discussion and conversation, some cre-

activity in finding solutions and persistence. The best way is to establish a dialogue on those issues that are close to the heart of the people: how to do better research.

This book addresses the difficult and controversial issue of evaluation of research in SSH with the support of experiences and data from the recent activities carried out in Italy.

Why is the case of Italy relevant for the debate on this issue?

First, this country has experienced perhaps the single largest and most comprehensive assessment process in large European countries, in a few years. After a long legislative process, started in 2006, the National Agency for the Evaluation of Universities and Research Institutes (ANVUR) was created in 2011. There was large recognition that the Italian system had been lagging behind the best European experiences in the field. The agency started immediately to implement several initiatives, on which the chapters in this book give large details. Here, it is important to recall the main features of these initiatives, in order to offer the overall framework.

- A research assessment exercise (VQR 2004–2010, or Valutazione della Qualità della Ricerca) was launched in 2011 and published in July 2013. *All* researchers submitted to the evaluation: researchers affiliated to universities had to submit three products, while those affiliated to public research organizations (PROs) were requested to submit six products. Research products were evaluated by 14 panels (GEV, or Gruppi di Esperti della Valutazione), composed by 450 members overall. Panels in STEM disciplines adopted bibliometric methods, with minor role for peer review, while panels in SSH exclusively used peer review (with the exception of economics and statistics). Overall, more than 14,000 referees were mobilized and over 180,000 products were evaluated. The results have been used by the ministry to allocate a share of the performance-based funding to universities.
- A second research assessment exercise has been launched in 2015 and completed by the end of 2016. Several improvements were introduced after the first VQR. As in the 2004–2010 exercise, individual researchers have received their own personal scores in a confidential way, while the scores of all disciplines (with at least four researchers per department) have been published in a highly disaggregated way.
- A major change in the legislation for academic recruitment and promotion was put in force in 2012. A National Scientific Habilitation for Associate and Full Professors (ASN, Abilitazione Scientifica Nazionale) system was created, with some similarities to the ones in place in Spain and France. The legislation introduced a system of quantitative indicators to be satisfied not only by candidates but also by full professors wishing to enter the national habilitation committees. This was a major departure from academic traditions. ANVUR was requested by the ministry to calculate the national distributions of indicators on scientific publications and to publish the median value as a threshold to be satisfied. In SSH sectors, given the absence or lack of validity of bibliometric data, three indicators

were selected: number of books, number of book chapters and articles and number of articles in A-rated journals.

- As a result of the implementation of ASN, ANVUR was requested to classify journals, in order to separate scientific from non-scientific journals, and to identify a set of A-class scientific journals. ANVUR asked learned societies to give an opinion and appointed an expert group. The starting point was offered by the self-administered database in which all Italian researchers report their publications. It turned out that there were more than 60,000 titles of journals. Out of these titles, the agency classified around 24,000 as scientific journals and 3000 as A-class journals. The lists were published in 2012 and revised periodically. By end 2017, a general revision is expected.
- In the context of quality assurance principles applied to teaching activities and to departments, a data collection exercise was undertaken, in which departments filled a database on their research activities, including metadata on publications. The resulting database (SUA-RD, Scheda Unica Annuale della Ricerca Dipartimentale, or University-level sheet on departmental research) has not yet been published at the time of this book. However, several authors in this book asked access to the SUA-RD documentation in order to compare data. This source will be important in particular in SSH fields, in which no external or bibliometric source is considered valid.
- Finally, in 2016, after a preparation process, the evaluation of the third mission of universities and PROs was undertaken, on the basis of a data collection process initiated experimentally in 2014. The process is based on the informed peer review methodology. In parallel to the VQR 2011–2014, the results have been made public in 2017.

From this list one can figure out how many issues and controversies have been addressed in a short time frame. In just a few years, universities and PROs, rectors and administrators, academicians and early researchers have been involved in completely new processes, arguments and metrics. They had to learn about citations, h-index and median values. A lot of controversies have been generated and are still alive. A huge data collection is now available, most of which in the public domain (although, due to resource limitations, mainly in Italian language). Several papers have been published in the last few years on this experience.

A second reason of interest is that SSH disciplines have a strong tradition in Italy, due to historical reasons. Entire branches of humanities, for example, from philology to history of art or archaeology, have a long tradition in this country, which originated from the Middle Ages and the Renaissance. Communities in SSH are lively, create discussions and take action.

Third, in the recognition that research on evaluation is a crucial element of the quality of evaluation, ANVUR started a major initiative aimed at funding, with small-scale grants, original research projects. This allowed scholars in a variety of fields to engage into theoretical or empirical studies that might contribute to the improvement of evaluation practices. A Call for Proposals was issued, and, following a competitive *ex ante* selection procedure, a series of projects were funded.

Some of the papers that originated from the call are included in this book and acknowledged directly in the text. The authors cover a large range of expertise, from computer science and computational linguistics to library sciences, from sociology to law, in addition to many branches of humanities. This is a remarkable enrichment of the community of those that study research evaluation in a systematic way.

The book is opened by a chapter in which **Andrea Bonaccorsi** calls for an epistemic approach to evaluation. By epistemic approach it means the systematic analysis of the ways in which communities in SSH produce *valid* knowledge. Evaluation must rest on the solid foundations of the criteria by which epistemic communities validate and value the knowledge they produce. There is a remarkable theoretical gap here.

Despite the work of sociologists such as Collins, Abbott, Ben-David, Becher and, more recently, Michèle Lamont, we lack a full-scale theory of the epistemic practices of scholars in SSH. By and large, most philosophy and sociology of science in the twentieth century have been dealing with hard sciences, not with SSH. We do have regional theories, particularly in social sciences (e.g. methodology of social sciences, philosophy of economics), but the linkage between epistemic issues and research evaluation is missing. The chapter offers a preliminary sketch of such a linkage and calls for further research.

Part I of the book deals with the controversial issue of criteria of research quality in SSH. Do scholars in SSH define research quality in a consistent way? Do they converge on common definitions?

The part suggests the idea that research quality criteria *do exist* in the practice of scientific communities in SSH, but they require a long process of conversation and elicitation to be formulated in linguistic terms, made reflexive and explicit and put in the public space. Scholars put these criteria in practice innumerable ways in evaluating papers and books submitted for publication, in making recruitment and promotion decisions and in the daily activity of reading others' works. These criteria are typically held in an implicit, non-declarative or tacit form. They form a crucial part of the *personal knowledge* of scholars, following the concept introduced by Michael Polanyi. The evaluation asks scientific communities to articulate these criteria, in order to make them declarative and subject to intersubjective discussion. Claims about what is valid and what is not valid in the knowledge produced in a community, and about the degrees of quality of valid knowledge, require rational justification. In this perspective, the evaluation is a kind of democratic process in which decisions are put under the light of public conversation and hence require justification.

The chapter by **Chiara Faggiolani** and **Giovanni Solimine** examines the recent Italian experience of research evaluation in the VQR 2004–2010, comparing it with the international experience. It is confirmed that the pattern of publication of scholars in SSH is structurally different from the one in STEM. The central role of books creates a difficult issue, because the only way to evaluate books is by reading them from cover to cover. They examine some of the most common solutions adopted in various countries, namely, the weighting of coefficients for monographs by type and size, the rating of publishers and series, the use of reviews in evaluation and Library Catalogue

Analysis (LCA). They do not recommend the uniform weighting of books, given the extreme heterogeneity of their type and content. Also, while publisher rating and the use of reviews might contribute to approximate elements of quality, the authors find that the evidence of validity and reliability of these methods are still not sufficiently demonstrated. Finally, they show caution with respect to LCA.

Andrea Capaccioni and **Giovanna Spina** examine international experiences in the drafting of guidelines for peer review in SSH. These initiatives are important in a variety of ways. First, they call for a continuous activity of self-reflection of communities on the way in which research quality criteria should be formulated linguistically. It is one thing to hold a notion of criteria, another thing to formulate it in verbal terms in such a way that other people, sharing the same notion at conceptual level, apply it in a consistent way. Second, peer review is a learning process for both those that carry it out and those who must accept it. The guidelines help to put in place solutions that improve the learning generated in the process, for example, in terms of training of referees involved in research assessment exercise or calibration of subjective judgements. Finally, there is a delicate issue of conflict of interest. It does not only refer to the problem of direct and personal conflict given by personal relations but to the more subtle issue of whether referees evaluate products for which they have little expertise. This problem can be managed, to a certain extent, by constructing databases of referees based on their past performance in peer review, an activity that the Italian Agency has implemented. However, there is room for self-constraint based on forms of ethical codes. Referees should not accept the evaluation of works for which they have no scientific legitimation. In the SSH field, this might be a controversial issue, given the extreme specialization and articulation of disciplines in subfields.

The chapter by **Andrea Bonaccorsi** reports about an initiative taken by the Italian Agency to open a dialogue with scientific communities in SSH about the improvement of the peer review process. A number of expert groups were formed, which were given full autonomy and freedom to formulate suggestions, in several SSH areas, from architecture to humanities and arts, from social and political sciences to law, to business studies. The outputs of these groups were transmitted to the evaluation groups under the new research assessment exercise (VQR 2011–2014) and were used to modify the evaluation sheets to be used in the peer review and to enlarge the range of admissible products. As an example, there was a large debate on the meaning of “internationalization” as a research quality criterion after the first research assessment exercise (VQR 2004–2010). One major concern was that referees could have interpreted this broad notion in a mechanistic way, by giving automatically higher scores to articles or books in English. The expert groups elaborated on the meaning of this notion for their respective disciplines and suggested solutions, most of which have been implemented in the new wording of evaluation. This is an example of the way in which engaging into conversation with the SSH communities may improve the validity of the research assessment. At the same time, it shows that communities should be challenged to formulate their own criteria in a declarative way, so that they can be implemented by other experts (in this case, the members of the evaluation groups at the national level).

The chapter by **Ginevra Peruginelli** and **Sebastiano Faro** enters into the troubled waters of research evaluation in legal studies, or law. This is an open issue at the international level, and several initiatives are under way to study the problems and formulate suggestions. In this field the adoption of national language is mandated by the object of research, which is mainly the positive law at the national level (with some exceptions, such as philosophy of law and history of law or, obviously, the various branches of international legal studies).

At the same time in these fields, there is a remarkable trend to incorporate judicial decisions produced by international courts, particularly at the European level, into the decision of national judges, using the original English formulation. This means that the epistemic process of production of valid knowledge in law will endogenously become more international in the near future. A second problem that has been hotly debated in the Italian context has been the legitimization of the agency to classify scientific journals. The initial position of learned societies was negative. A legal action was taken in 2012 by a prominent scientific society, based on the argument that only scientific communities, and learned societies on behalf of these communities, had the legitimization to produce research quality criteria. The administrative judge rejected the claim, arguing that there is no legally defined role for learned societies, while an agency created by the law, acting on the principles of transparency stated in the law, may directly refer to scientific communities in order to define research quality criteria. This may happen in a variety of ways, by appointing experts, consulting documents, examining data and combining different sources of judgement. As a matter of fact, the classification of journals has been carried out also in law. The chapter suggests that, after the initial conflict, there is now a reflexive and constructive attitude by the academic legal community.

Part II of the volume deals with the role of books and monographs in the evaluation of SSH research and the appropriate methods.

Geoffrey Williams, **Antonella Basso**, **Ioana Galleron** and **Tiziana Lippiello** first examine the treatment of books in research assessment in several European countries, characterized by largely different methodological approaches: qualitative systems (UK and Netherlands), quantitative systems (Czech Republic and Poland) and database systems (Norway and Flanders). In addition, they mention the Spanish experience of rating publishers. They submitted a questionnaire to a sample of Italian SSH scholars, using a version of an instrument first developed in France, and received 578 usable answers. Among the results, two are in my view particularly interesting. The first is that the respondents mention, among the communication channels they use for their research work beyond traditional academic channels, the Internet as the first option. This opens interesting directions for further research. The second finding is that the choice of communication channels by scholars seems to follow a well-defined strategy. On the one hand, junior researchers consider research articles much more than in the past due to their speed of publication but also because journals offer, differently from books, a formal peer review process. It does not seem that researchers choose journal articles because they are easier to publish, as some critics often argue, but quite the contrary, because they submit their results to a rigorous process of validation. On the other hand, respondents are clear

in answering that, in their departments, books are evaluated more than articles. My reading of their results is that a transition is in place in the way in which SSH scholars communicate their research results. This transition does not follow the path illustrated by the alarmistic wisdom according to which research evaluation will eliminate books and monographs, because researchers turn their attention exclusively to easy-to-write, easy-to-publish journal articles. Rather, a more complex and sophisticated strategy is emerging, mixing journal articles on refereed journals (particularly in the very early stages of career) and books (in a more mature stage). This academic strategy seems to be associated with the enlargement of communication channels towards the Internet.

Carla Basili and **Luca Lanzillo** offer a survey of guidelines for the evaluation of book and monographs, combining sources from international bodies (ISO standards), national experiences (Spain, Australia and New Zealand) and university-level manuals (University of Turin and University of Bologna).

The analysis shows a rich variety of objects covered under the notion of book and monograph, almost all of which tend to be included in research assessment exercises at the national level and university level. The authors observe a trend of inclusiveness of new categories over time. These are, almost universally, subject to peer review. This creates a tension, however, with the need to include books and monographs in simple counts of publications, based on administrative data and/or self-reported data. Existing classification systems, in fact, have a hard time distinguishing, to make just an example, between those edited books that are the result of complex, multi-year, interdisciplinary research projects and those edited books that are just a collection of disparate contributions. The tension between full reading books in peer review and just counting books is inevitable. The authors confirm the recommendation to make use of peer review.

The need to introduce methods for the evaluation of books without the need to read them from cover to cover was at the origin of the proposal of Library Catalogue Analysis, the object of the chapter by **Maria Teresa Biagetti**, **Antonella Iacono** and **Antonella Trombone**. The authors examine in great detail the internal working of digital catalogues in academic and nonacademic libraries and the evolution of standards and digital technologies. The evaluation question here is simple: can we rely on metrics based on the diffusion of books in qualified libraries at the international and national level, in order to build up a proxy of their quality? Are more diffused books, or books present in a qualified subset of libraries, also better books? The chapter examines several apparently technical issues that, however, are crucial for addressing this issue, such as the distinction between acquisition and donation of books, the commercial policies of publishers in the acquisition process or the distinction between the original edition, new edition and republishing of previous editions. According to the authors, existing digital catalogues fail to offer reliable data for evaluation purposes.

They also put under scrutiny new development in the digital technology applied to libraries, namely, Advanced Discovery Tools and Linked Data Technologies. The former are commercial tools offered to researchers for the fast and comprehensive search of metadata on publications of interest, using advanced data mining tech-

niques. The authors carry out a simulation, comparing the output of some Advanced Discovery Tool with first-hand information from library catalogues, and conclude that there are large gaps in information that prevent its use in the evaluation procedure. Linked Data Technologies are the answer to the challenge of opening library information systems to the new environment created by the web and open data.

Ginevra Peruginelli, Sebastiano Faro and Tommaso Agnoloni return to the field of legal studies with an original survey on Italian scholars. They obtained an unusually high rate of return to an online questionnaire (26% of the population, or 1241 usable responses), a result that indicates the interest (and controversies) raised by research evaluation in the legal disciplines. 809 respondents, or 76.4% of the total, are members of editorial boards of journals, and 796, or 75.5% of total, declare they are engaged in *ex ante* evaluation of journal articles.

With respect to journals, an interesting finding is that a minority of respondents believe peer review is carried out systematically (23%). The bulk of respondents (55%) believe it is carried out only in most cases, while 22% believe it is not carried out at all. This finding seems to be in contrast with the large diffusion (76.4% of the total) of editorial roles in the legal community. This discrepancy can be explained by the organization of the *ex ante* process in most Italian journals in law. Peer review is not systematically managed with the double-blind, or single-blind, approach, by recruiting a large pool of anonymous referees and establishing formal procedures. Rather, it is directly managed by the editor or the editorial board. The presence of almost all scholars in at least one of the journals, demonstrated by the data, is a way to protect oneself, and one's students and collaborators, against the uncertainties of a fully competitive process. The chapter calls for more transparency in the description of the peer review procedures followed by legal journals.

A second interesting finding is that, once asked which research quality criteria they mention with respect to books, there is a consistent ranking across legal disciplines in which the top criteria include the inclusion in a particular series with special characteristics and the publisher. In other words, respondents admit that they use information on the book series and the publisher as an indicator of quality. Yet, they vigorously reject the notion of rating or ranking series and publishers and probably, as it is demonstrated in other chapters, rightly so.

Part III deals with one common methodology used in the evaluation of SSH research, namely, the rating of journals. As it is well known, it is a largely adopted methodology but also one subject to some controversy, as the experiences of Australia and France show.

In her chapter, **Domenica Fioredistella Iezzi** investigates the notion of "publication strategy", by examining the mix of types of publications submitted to the research assessment exercise (VQR 2004–2010) and the resulting average scores. It turns out that large heterogeneity can be detected. To summarize the argument, there are two broad areas (Area 10, or antiquities, philology, literary studies and art history, and Area 11, or history, philosophy, pedagogy and psychology), in which there is little difference in the average score across the main categories of products (journal article, monograph, edited book, book chapter and proceedings). The scores are in the range 0.63–0.72 in Area 10 and 0.55–0.64 in Area 11. In these fields, it

seems that scholars may submit whatever product they have produced and receive similar scores. Also, monographs receive the highest score (0.72 in Area 10 and 0.64 in Area 11). Furthermore, in these areas proceedings are highly valued (0.66 and 0.59 average score, respectively). The remaining three fields share a common negative evaluation of proceedings: their score is as low as 0.38 in social and political sciences, 0.26 in law and a mere 0.07 in economics and business. In these fields, paper presentation to conferences is not considered equivalent to articles or book chapters. They differ, however, under other respects. In the field of law, chapters, monographs and journal articles are valued high (range 0.56–0.63), while edited books are penalized (0.33). It seems that the editorial work of collecting chapters is not considered scientifically valid. Edited books are more valued in political and social sciences (0.51), while they are extremely penalized in economics and statistics (0.04). The area of economics and statistics is also interesting because the scores assigned to monographs, all evaluated in peer review, are extremely low (0.11 on average). In this area, the only product that is highly valued is the journal article (0.51). From these data, it is clear that there is a transition in the pattern of scientific communication and that scholars should adopt a well-defined strategy. The chapter also investigates the role of language and finds that, particularly in humanities and arts (Area 10), there is a large role for products in languages other than the national one. Interestingly, these products are not in English but in (many) other non-national languages.

The paper by **Andrea Bonaccorsi, Tindaro Cicero, Antonio Ferrara** and **Marco Malgarini** addresses the contested issue of journal rating. In the SSH field, there is almost a universal agreement on the impossibility to use indexed journals as the basis upon which to build indicators, since they represent a tiny fraction of journals used. Therefore, all journal rating exercises make use of expert-based assessments or combine quantitative indicators with expert judgements. After an extensive review of international experiences and of the academic literature, the chapter addresses the issue of whether the classification of journals (or the container) is a good predictor of the evaluation of individual articles published in journals (or the content). The Italian context offers an unprecedented opportunity to examine this issue on a large scale, due to the presence of two (relatively) independent evaluation processes, one carried out on individual articles (VQR 2004–2010) by thousands of anonymous referees and another carried out on journals, by a small-size expert group, which used also evaluations from learned disciplinary societies. While there might be debate about the degree of independence between the two measures (in particular, a small subset of journals was rated within the VQR exercise and communicated to referees, following the notion of *informed peer review*), it is true that there is no evidence that referees did not read the articles (the content) and followed the suggested classification of the journal (the container). In this case the variables would clearly be non-independent. Therefore, a classical multiple-regression approach seems to be justified. The authors find strong support for the validity of journal rating in predicting the average quality (score) of individual articles, although there remains (and should remain) variability around the predicted values.

Part IV examines a range of frontier problems in the evaluation of SSH research: new indicators based on Google Scholar (GS), library holdings and the issue of impact and/or third mission of SSH scholars.

There is a large literature comparing indicators from Google Scholars (GS) to traditional indicators based on Scopus and Web of Science. These studies invariably deal with STEM disciplines. The chapters in this book are highly original because they deal with Google Scholar indicators in SSH and compare these data with information derived independently from research assessment exercises.

Alfio Ferrara, Stefano Montanelli and **Stefano Verzillo** have developed a novel methodology for the large-scale disambiguation of GS data. They utilize the large database created within the National Scientific Habilitation procedure, opened in 2012. In this procedure candidates submitted their lists of publications and published them on the official ministry website. Since candidates declared which scientific discipline they were applying for, it is possible to establish a systematic linkage between the sets of words appearing on the metadata of their publication lists (e.g. words in titles) and their disciplines. The authors have applied state-of-the-art computational linguistics tools to train the algorithms for searching GS data, obtaining excellent results in terms of recall and precision. The authors conclude that GS can indeed be used within evaluation activities. It is important to underline that this approach does *not* require that authors build up their own profile on GS: the algorithms were able to identify precisely whatever author was in the list of Italian scholars *without* using profiles. I believe this methodology might be extended to other disciplines and other countries and open a new research direction for the use of GS as a complement to, or alternative to, other bibliometric sources.

In a companion chapter, **Ferruccio Biolcati-Rinaldi, Francesco Molteni** and **Silvia Salini** make use of Ferrara et al.'s methodology, and of the related dataset, to investigate the publication behaviour of scholars in social sciences. They compare the data from GS to data on publications from a newly created archive of publications of Italian scholars, called SUA-RD, which is yet unpublished. This archive comes from self-declared publications collected at the departmental level by ANVUR. They confirm that GS data can be used for evaluation purposes, although self-declared publications are a much larger number. At the same time, they show that many authors do not receive any citation, not only from indexed journals (which might be a result of the adoption of national language) but also in GS. Since Google Scholar reflects the largest available collection of documents, it is noteworthy that part of the academic community shows such a small impact.

Maria Teresa Biagetti, Antonella Iacono and **Antonella Trombone** address the issue of whether the diffusion of books in libraries can be used as a valid and reliable information for evaluation purposes. They run a simulation exercise, by comparing the data obtained by consulting digital catalogues of a sample of high-status libraries. They find large variability across catalogues, a result that casts doubts about the possibility to normalize this kind of information. Summing up, they do not recommend data on the diffusion of books in library holdings as a viable alternative to peer review or a useful complement to it.

These three chapters give us interesting recommendations for the use of new tools in evaluation: green light to Google Scholar, but only conditional on the adoption of a rigorous disambiguation system, and red light to Library Catalogue Analysis, until large improvements are introduced in the treatment of some critical issues.

The last two chapters of this part deal with a novel dimension of research evaluation in SSH, that is, the impact of research or the third mission.

Luca Lanzillo offers an articulated discussion about the concept of social impact assessment, in the light of the orientations of the European Commission and the experience of the REF in the UK. He defends the view that scholars in SSH do have a large impact with their research, although the definition and measurement of this impact require much research and implementation work.

Brigida Blasi, Sandra Romagnosi and Andrea Bonaccorsi report about the first effort to evaluate the third mission of universities and PROs undertaken by ANVUR, after a preparation process. In particular they examine the differences between STEM and SSH scholars in the activities labelled under the Public Engagement heading, a large and comprehensive category created in order to take into account the variety of interaction modes between research and society. At the time of closing this volume, the data on third mission are still under evaluation. The chapter offers a very preliminary snapshot of data, in the form of a frequency distribution of activities carried out by scholars. It turns out that scholars in SSH are no less active in public engagement than their colleagues in STEM.

Taken together, these last chapters call for a rigorous discussion of the notion of impact of SSH research. On the one hand, as the authors remark clearly, SSH researchers are very active at the interface with society and carry out a variety of initiatives. They are less visible because their products are more intangible and immaterial than those produced by STEM research (e.g. patents, spinoff companies, science and technology parks, etc.). They are also less visible simply because the notion of third mission has been initially considered equivalent to the valorization and commercialization of research, which is only part of it. There is a need to place the impact and third mission of SSH higher in the agenda, for both SSH researchers and policymakers and the public opinion.

On the other hand, we should avoid the trap of imitating STEM research in trying to demonstrate the economic, or short-term, impact of SSH research. This approach would never deliver the kind of immediate, tangible, measurable results that are often the request of policymakers. Rather, it is important to build up a full-scale and compelling argument, in which the epistemic dimensions are associated to ethical and political arguments, in order to show the constitutive role of SSH research for the cohesion and advancement of contemporary societies. In studying languages and texts, in preserving the material, iconic and symbolic memory of civilizations, in supporting the creation and interpretation of the rules by which people regulate their rights and duties, produce and exchange goods and services and make collective decisions, SSH research builds up and maintains “the cement of society”, to use the celebrated expression of Jon Elster. On the basis of such an argument, it will be possible to spell out the relevant dimensions of evaluation for SSH communities.

The book is closed by a contribution by a non-Italian scholar, a leading expert in the evaluation of research in SSH, **Alesia Zuccala**, from the Royal School of Library and Information Science of the University of Copenhagen. This is an important contribution to the book, which would otherwise be subject to self-referentiality. The author offers first a fascinating theoretical discussion of the role of language and of national and cultural specificities in the SSH fields. She then surveys the chapters of the book in an effort to examine the contributions they give to the international debate on research evaluation, either at the conceptual level (epistemic issues, classifications, research quality criteria, impact and third mission) or the methodological and technical level (peer review, LCA, journal rating, publisher rating, book reviews, Google Scholar). The position taken by the author is that the Italian experience stands in front of the elder and more developed Anglo-Saxon approach: “The plural evaluation culture that I describe in this chapter – a European-Anglo culture at this point – possesses greater opportunities for uniting when there is a subtle respect for differences, rather than uniting due to forced ideals”. Plurality of evaluation cultures is, perhaps, the name of the game for the future of research evaluation.

I thank the authors of the chapters for long discussions on the content and on the overall project. ANVUR not only provided the financial support to some of the projects whose results are reported in chapters but also authorized the publication of this book and, most importantly, ensured complete intellectual freedom to all researchers involved.

At the end of the day, the legitimation of research evaluation will be achieved when people will recognize that it is an integral part of the academic profession. We evaluate and we are evaluated. We see and we are seen, all the time. And since working in the academy is perhaps the most rewarding profession in the world, one might also expect that evaluating oneself and the others is a source of professional satisfaction, while being evaluated (yes, sometimes negatively) is part of life.

In a poem that all Italian children learn at school, *The solitary bird*, the great poet Giacomo Leopardi offers a vivid picture of the joyful company of young people in the village: “Dressed for the festival/ young people here/ leave the houses, fill the streets, /to see and be seen, with happy hearts” (tr. A.S. Kline). *To see and be seen*, this is part of the happy hearts. But the poet himself does not share the joy, he stays solitary: “I go out, alone,/ into the distant country,/ postpone all delight and joy/ to some other day”.

The solitary attitude of the poet is certainly more noble and more inspired than the common attitude of the young company in the village. To see and to be seen is a social activity, and like all social activities, it can be considered by someone with a certain detachment. Authoritative scholars in SSH are perhaps right in pointing to the inevitable lack of depth of evaluation, when compared with the uniqueness, richness and erudition of the knowledge produced.

This book is an invitation for all who stay “in the distant country” to join, at least in part, the company of those who accept that “to see and be seen” is part of life.

Acknowledgment

The book has been realized under the auspices and with the financial support of the Italian Agency for the Evaluation of Universities and Research institutes (ANVUR).

Contents

Towards an Epistemic Approach to Evaluation in SSH	1
Andrea Bonaccorsi	
Part I Research Quality Criteria in SSH	
Mapping the Role of the Book in Evaluation at the Individual and Department Level in Italian SSH. A Multisource Analysis	33
Chiara Faggiolani and Giovanni Solimine	
Guidelines for Peer Review. A Survey of International Practices	55
Andrea Capaccioni and Giovanna Spina	
Peer Review in Social Sciences and Humanities. Addressing the Interpretation of Quality Criteria	71
Andrea Bonaccorsi	
Research Quality Evaluation: The Case of Legal Studies	103
Ginevra Peruginelli and Sebastiano Faro	
Part II The Role of Books and Monographs in SSH Research and Their Evaluation	
More, Less or Better: The Problem of Evaluating Books in SSH Research	133
Geoffrey Williams, Antonella Basso, Ioana Galleron, and Tiziana Lippiello	
Research Quality Criteria in the Evaluation of Books	159
Carla Basili and Luca Lanzillo	
Quality Evaluation of Online Library Catalogues, Advanced Discovery Tools and Linked Data Technologies	185
Maria Teresa Biagetti, Antonella Iacono, and Antonella Trombone	

A Survey on Legal Research Monograph Evaluation in Italy	211
Ginevra Peruginelli, Sebastiano Faro, and Tommaso Agnoloni	
Part III Journal Classification and Rating	
Publication Strategies in SSH: Empirical Evidence and Policy Implications	233
Domenica Fioredistella Iezzi	
Journal Ratings as Predictors of Article Quality in Arts, Humanities, and Social Sciences: An Analysis Based on the Italian Research Evaluation Exercise	253
Andrea Bonaccorsi, Antonio Ferrara, and Marco Malgarini	
Part IV Exploring New Indicators for the Multilevel Assessment of Research	
Google Scholar as a Citation Database for Non-bibliometric Areas: The EVA Project Results	271
Alfio Ferrara, Stefano Montanelli, and Stefano Verzillo	
Assessing the Reliability and Validity of Google Scholar Indicators. The Case of Social Sciences in Italy	295
Ferruccio Biolcati-Rinaldi, Francesco Molteni, and Silvia Salini	
Is the Diffusion of Books in Library Holdings a Reliable Indicator in Research Assessment?	321
Maria Teresa Biagetti, Antonella Iacono, and Antonella Trombone	
The Social Impact Assessment in Social Sciences and Humanities: Methodological Issues from the Italian Experience	345
Luca Lanzillo	
Do SSH Researchers Have a Third Mission (And Should They Have)?	361
Brigida Blasi, Sandra Romagnosi, and Andrea Bonaccorsi	
Part V What Do We Learn from the Italian Experience?	
Language, Culture and Traversing the Scholarly Evaluation Landscape	395
Alesia Zuccala	
Index	413

Towards an Epistemic Approach to Evaluation in SSH

Andrea Bonaccorsi

1 Introduction

The reaction of scientific communities against evaluation of research is almost entirely concentrated in SSH disciplines. STEM disciplines seem to have accepted that research evaluation, as it is often stated, “is here to stay”. Some exceptions are sometimes raised in Mathematics or Clinical medicine, in particular against some of the practices in bibliometrics (e.g., Impact Factor), but overall these arguments constitute a minority opinion.

Why is this the case? I suggest that the reasons cannot be purely sociological, or related to the way in which scientific communities organise their work and communicate their result. Nor can it be political or ideological: the evidence that political or ideological opinions significantly differ by discipline is scattered and not robust. Something different, or deeper, must be at work. Since scholars are motivated more by the intrinsic logic of their scientific work than by external incentives (although incentives matter a lot), the explanation must be found at the epistemic level.

By “epistemic” I mean the way in which scientific communities produce valid knowledge, or the procedures, criteria, practices by which they recognise inter-subjectively the value and validity of the knowledge produced by others, and by this way submit themselves to the same rules. In this perspective, the inter-subjective dynamics of communication and validation are not separated from the internal dynamics of knowledge, or the intrinsic persuasiveness of the knowledge exchanged (Ziman 1978, 2000).

In this sense epistemic is not the same as epistemological, since the latter requires a second-level abstract and professional reflection on the rules of scientific work.

A. Bonaccorsi (✉)
DESTEC, University of Pisa, Pisa, Italy

IRVAPP-FBK, Trento, Italy
e-mail: a.bonaccorsi@gmail.com

Not all scientists are also philosophers of science (indeed, only very few), but all good scientists have a solid mastery of a series of rules that are used to discriminate knowledge claims according to their purported validity.

Nor is it the same as sociological, since at this level the main interest is the way in which communities build up their agreement (or disagreement), irrespective of the content of knowledge. Sociological studies of science are mainly interested in the way in which socially defined actors, like scientists, set the boundaries of scientific vs non-scientific knowledge (Gieryn 1983, 1995, 1999; Taylor 1996), define scientific disciplines (Lenoir 1997; Abbott 2001), reach agreement or disagreement about claims (Knorr Cetina 1999), use material infrastructure and laboratory facilities to build up shared meanings (Latour and Woolgar 1979), create the conditions for repeatability of experiments (Collins 1975, 1985, 1999) or balance scientific power relations (Frickel and Moore 2006). For programmatic reasons, sociological studies do not deal directly with the epistemic content of knowledge, as separated (or separable) from the social interactions associated with it (Barnes and Edge 1982; Mulkay 1991; Barnes et al. 1996; Yearley 2005).

While I will use materials from epistemology as well as sociology of science, the main focus will be on the epistemic level, as elaborated by authors such as Ziman (1978, 2000).

In this chapter I address the following questions:

- What are the epistemic differences between STEM and SSH that may explain the differences in the orientation towards research evaluation?
- Are there epistemic differences across disciplines in SSH that may explain intra-SSH differences in the orientation towards research evaluation?
- Are there research quality criteria on which communities in SSH may converge? Or, is it possible to address epistemic differences with procedural fairness?

2 Epistemic Differences Between STEM and SSH

An influential stream of literature, inspired by logical positivism, argued that the difference between STEM and SSH is very simple: the former are scientific disciplines, the latter are not (Steinmetz 2005).

By scientific discipline meant a body of knowledge that could, at least in principle, formulate causal propositions. The formulation of causal relations requires a number of conditions that are found in natural sciences, but not fully in social sciences, even less in disciplines that deal with language. In natural sciences it is possible to assume the invariance of the object, so that controlled experiments can be carried out.

This view has dominated the scientific literature for decades after the Second World War. It is still maintained by some authors.

However, it is no longer assumed as the dominant theory, particularly after the developments of philosophy of science and social studies of science in the 1960s. The impact of Kuhn (1962) has been crucial here: the reason why scientists may formulate causal propositions is not that they control *each* of them in isolation,

but because these propositions are consistent with an overall paradigm, whose foundations do not have the same level of controllability. In addition, scientists produce a large variety of propositions, not only causal ones, referring to their experimental apparatus, the concrete rules of operation in the laboratory setting, or the practices of exchanging results.

Post-positivistic accounts of modern science admit a larger range of propositions as scientifically valid. This opens the way for asking to what extent disciplines in SSH may be defined scientific as well.

It is possible to summarise this issue separately for Social Sciences and Humanities.

In Social Sciences the issue of scientific validity of propositions has a long history, starting with the foundations of classical political economy and sociology in the eighteenth and nineteenth centuries. In the thought of classical authors such as Weber and Durkheim, knowledge produced in Social Sciences may well be defined as scientific, but not in the sense of producing invariant causal propositions (explanation) but rather propositions that make the behaviours of social actors intelligible by referring to their motivations (interpretation). Social Sciences are no less scientific than natural sciences, to the extent to which they submit their propositions to the same kind of rigorous control, but not through the use of experiments (which cannot be done by definition) but by establishing some level of stability of the relation between reasons for action (motivation) and observed action.

To what extent these disciplines can be defined “scientific” and what are the differences with respect to STEM? In the following I reject the notion that SSH disciplines are not scientific and investigate rather which epistemic differences can be identified.

First, researchers in STEM aim at discoveries, while researchers in SSH have only occasional discoveries (a new archaeological site, document, text, manuscript...) but most often aim at new interpretations of existing texts. The focus on discoveries means that scientists are in competition amongst themselves. Science is competitive because researchers fight to be the first to publish discoveries and receive the credit.

Second, research in STEM is cumulative, because scientists build upon the contributions of others, either in the past or from current competition. Science is a collective undertaking, not an individual enterprise. Science is both competitive and collaborative at the same time. There is a sharp difference here with respect to SSH, in which cumulateness is much lower (Walliser 2009a, b).

Third, because of competition for discoveries and cumulateness, the appropriate communication channel is the journal article (Lindsay 1978; Bazerman 1988; Cronin 1984, 2005). The scientific journal is serial or periodical, it offers researchers all over the world the opportunity to be updated regularly on discoveries, the format of the article is suitable for communicating discoveries, and the peer review system is efficient in solving issues of information asymmetries on the attribution of priority. The scientific journal system follows the competitive structure of science (Dasgupta and David 1994). Over time, the competitive dynamics generate a hierarchical system based on a cumulative process of reputation building: scientific journals that have published important discoveries are read more frequently; consequently authors

compete to be published in them; the increase in the number of submissions makes it possible to raise the rejection rate, making the quality of journals even higher and attracting more readers, and so on. It is this structure of scientific activity that makes it possible to build quantitative measures of research quality. In particular, once the role of citations is clarified in an unambiguous way, and the set of journals for which scientists compete is sufficiently large, then the very competitive dynamics generate a system in which the underlying quality is reflected in the relative measures of citations applied to papers, authors, institutions, while the impact factor of journals is considered a reliable measure of their average quality. Therefore there is a strong connection between the nature of scientific activity in discovery-driven fields, the overall system of academic publishing, and the reliability of quantitative measures of research quality based on citations.

On the contrary, the journal article is not the suitable medium for SSH, because new interpretations require long explanations best suited for the book format (Baldi 1998; Brooks 1985, 1986).

Fourth, because of competition for priority, cumulativeness, and the workings of the scientific journal system, citations are an essential element of scientific communication. In STEM citations have unambiguous meaning of credit assigned to authors that made the previous discoveries. As it has been originally discussed by Merton, Garfield and De Solla Price, and more recently formalised by Dasgupta and David, citations to the previous literature are a necessary ingredient of scientific publishing (Bornmann and Daniel 2008). This necessity is neither ethical nor practical, it is functional. By functional is meant, according to Merton's sociological approach, that individuals are forced to use a citation system that complies with the collective rules of the scientific community, irrespective of the individual willingness.

In order to be credited for a discovery, the authors must demonstrate their contribution is new with respect to the state of the art. In the absence of citations, it would be on the shoulders of the readers to check carefully whether there is anything new, clearly a very inefficient solution. Thus the overall system of scientific journals is based on referees who directly check the credibility of the authors' statements, acting as agents on behalf of the scientific community. In doing so, they force authors to list all relevant citations. Furthermore, due to the cumulativeness of scientific discoveries there is no need to cite authors from the distant past, but only the papers published in the last few years, which include all the relevant knowledge. This is a striking feature of scientific papers: only a few scientific authorities of the past are cited, not because they are ignored, but because their contribution is embedded in the citations of more recent authors.

In SSH, on the contrary, researchers quote authors from a distant past, very often classical authors, and produce works that are not cumulative but complementary, segmented or even alternative to each other. While the segmentation into scientific fields and sub-fields is largely agreed in STEM, and is usually not the outcome of individual decisions, in SSH part of the activity of most creative authors is the definition of new fields or new segmentations. The existence of progress, i.e., that some works are not worth being cited because their contribution has been subsumed

into others' contributions, is usually recognized very late, often after the authors cease their activity or die. Consequently, there is a need for a different theory of citations in SSH. Citations serve different purposes and should be classified accordingly.

Finally, there is a different role of paradigmatic pluralism. In STEM there is most often a dominant paradigm, sometimes with one or a few minority positions. Due to the cumulative nature of science and limited paradigmatic diversity, competition is open. On the one hand, within disciplinary boundaries all researchers compete fiercely for discoveries, without internal segmentations that may protect against competitors. On the other hand, since peer review is (generally speaking) a blind process, the past reputation and academic status of authors are not relevant to the probability of being published. This means that incumbents, or people with a recognised academic status, do not enjoy monopolistic positions in the long run. New entrants like junior researchers and authors with unorthodox views are easily recognised. Under these conditions, it is not possible for a single author or group of authors to monopolise the citations or to manipulate the reputational indicators.

This is not the case in SSH, in which paradigmatic pluralism is not the exception but the rule. On the one hand, there are internal segmentations that are not due to disciplinary differences but rather to paradigmatic options (rooted in the choice of object, methodologies and techniques) but also to value-laden positions (academic schools and traditions, ideological positions, political affiliations and attitudes, cultural orientations).

Competition is not completely open but segmented. Scholars have sometimes a two-layered choice: first, with which paradigm they want to be affiliated; second, how to compete within the paradigm chosen. In some important sense, there is competition among paradigms, but each of them is organised into its own scientific and academic structure (often with dedicated journals, conferences, scientific societies). Competition within the paradigm is not open but controlled by the leaders who contributed its creation. The relationship between paradigms is a matter of academic power, or maybe of paradigmatic change in the long run. On the other hand, in SSH peer review is not universally adopted. The identity of authors is generally known by those who make editorial decisions in journals and book series. Since books are the most important source, the control of editorial decisions is more easily controlled than in journals. This makes competition among researchers even more restricted.

As it appears from the above discussion, there are clear counterparts of this situation in the field of industrial organisation in economics. The kind of competition experienced in science is similar to the situation of competitive markets, in which entry is open, incumbents never get a monopoly position, and it is not possible for an incumbent to manipulate strategic variables to its own advantage.¹ This is why, in my

¹That is why I find most arguments about the possibility of manipulation of bibliometric information, such as the Impact factor, rather pointless. Science itself is manipulable. There are many examples of fake discoveries or misbehaviour of scientists. The truly interesting question is not why these things happen, but why they happen so infrequently and how it happens that they are almost invariably discovered and punished.

opinion, scientists are not Foucauldian (see below). They find that the representation of commensuration as a form of hidden power is not appropriate for the way in which science works in their fields. It is not a matter of lack of reflexiveness, or pragmatic orientation, as opposed to the kind of critical work advocated in social sciences. Even scientists acutely aware of the social implications of their activity never subscribe to a Foucauldian argument. Simply put, competition is so harsh and the rate of knowledge production so overwhelming that no power coalition is stable.

This is not necessarily the case in SSH, where the competition is more of a monopolistic type, or even collusive oligopolistic. In other words, due to the fragmentation of disciplines and paradigmatic pluralism, the possibility of controlling a discipline for long periods is not negligible.

3 Epistemic Differences Within SSH

Yet this picture is still incomplete. On the one hand there are disciplines in SSH that have historically emulated STEM disciplines. On the other hand there are internal differences within SSH that also have implications on the orientation towards research evaluation. Thus we are faced with the challenge to examine differences *within* SSH disciplines.

In recent years the methodological foundations of social sciences (Sayer 1992; King et al. 1994; Goertz 2006; Moses and Knutsen 2007; Della Porta and Keating 2008; Brady and Collier 2010; Goertz and Mahoney 2012) and the position of social sciences with respect to general issues raised in the philosophy of science (Sayer 2000; Delanty 2005; Delanty and Strydom 2003; Benton and Craib 2011; Steele and Guala 2011) have been investigated thoroughly. A few cross studies (Steinmetz 2005; Walliser 2009a, b; Camic et al. 2011) have examined the differences across disciplines, while some other studies deal with the impact of social sciences in society (Flyvbjerg 2001; Brewer 2013; Bastow et al. 2014).

In parallel, a similar process started in Humanities, though somewhat less articulated, and partly as a response to the academic decline of these disciplines (Kernan 1997; Bate 2011; Belfiore and Upchurch 2013; Small 2013; Brooks 2014). Here a few historical comparative studies are also available (Bod 2013).

From this methodological and comparative literature, associated with related disciplinary studies, I have obtained a clear picture of the main epistemic problems addressed by various disciplines in SSH. In a recently published book (Bonaccorsi 2015) I presented a quite detailed reconstruction of the epistemic debate on four disciplines in SSH (history, political science, anthropology and English literature), combining historical material on the process of institutionalisation of the discipline in the academic system with an analysis of the main theoretical and methodological controversies. I strongly believe that this is a promising direction for research. Comparative studies that combine epistemic issues with institutional details will illuminate the way in which valid scientific knowledge is created. By taking disci-

plines as object of analysis I recognise that there are also internal distinctions within disciplines (Becher 1989; Abbott 2001) and try to take them into account.

This approach is not only useful to address the controversial issue of evaluation. It is my contention that entering into the epistemic black box of disciplines in SSH is also the only way to build up rigorous arguments to defend them vis-à-vis other disciplines, funding agencies and policy makers. There is a need to build up an argument about the *scientific* nature of SSH, based on a thorough recognition of the way in which they build up valid knowledge, though with epistemic processes that are completely different from the ones used in STEM. The damage to SSH generated by the wave of theorising that has suggested that they are just another way of producing texts instead of a truly scientific endeavour is currently underestimated. It is not enough to underline the pragmatic value of SSH in society. What is needed is a demonstration of the *intrinsic* validity of the knowledge produced by SSH scholars.

In this section I will sketch the main results of the detailed analysis carried out in Bonaccorsi (2015) and add other prominent disciplines in SSH. I review the four disciplines discussed at length in the book (history, political science, anthropology and English literature) and add other large disciplines in Humanities (philology, art history, psychology) and Social sciences (economics). The discussion below will be very concise. Interested readers are referred to references quoted in Table 1 and to the extended discussion and long reference list in the book.

I suggest that the orientation towards research evaluation is a function of four constructs, which combine historical factors with epistemic dimensions:

- (a) History of the academic institutionalisation of the discipline
- (b) Main methodological orientation
- (c) Position with respect to neo-positivism after Second World War
- (d) Position with respect to post-structuralism in the 1960s and 1970s.

Let me explain the building blocks of the model. By academic institutionalisation I mean the way in which a discipline comes to be separate from others, receive an academic label, is taught at universities in a separate way, academic positions are created and hence learned societies are formed. The institutionalisation process may be very long, taking decades (Becher 1989; Abbott 2001; Hyland 2012). A discriminant factor is whether a discipline is recognised from the beginning or is separated from previously existing disciplines. In the former case, newly created disciplines maintain the “memory” of their institutionalisation by keeping existing disciplines at a distance. They even challenge existing disciplines, either methodologically or substantively.

Thus for example English literature in the US academic system is the outcome of a separation process which took most of the nineteenth century to be completed. Literary studies were initially compressed between philology, which was the dominant discipline in US universities that followed the German educational model, and low level literature reading courses (Baldick 1983; Court 1992). The institutional separation took place in the early twentieth century and was associated with a deep

Table 1 Epistemic differences across disciplines in SSH

Discipline	Institutionalization process	Methodological orientation	Position with respect to neopositivism	Position with respect to post-structuralism	Selected references
Economics	<p>Historical origins in Moral philosophy</p> <p>First re-definition with marginalist revolution (XIX century)</p> <p>Second re-definition in XX century with the creation of the <i>Econometrica</i> journal (adoption of formal modelling in mathematical language)</p>	<p>Mathematical language</p> <p>Axiomatic foundation of behavioral assumptions</p> <p>Law-like generalizations (nomothetic)</p> <p>Non-mainstream economics committed to modelling</p>	<p>Practical acceptance</p> <p>Axiomatic foundations linked to high level debate in logic and philosophy</p> <p>Empirical issues regarding behavioral aspects addressed via the “as-if” approach (Friedman)</p> <p>Criteria for validity of causal propositions observed via econometric techniques</p>	<p>No impact</p> <p>Use of mathematical language shielded economics from the impact of post-structuralism, mostly based on the critique of texts</p>	<p>Lerner and Lasswell (1951)</p> <p>Friedman (1952)</p> <p>Bell (1982)</p> <p>Hausman (1992)</p> <p>Blaug (1992)</p> <p>Guala (2006)</p> <p>Walliser (2009a, b)</p> <p>Szenberg and Ramrattan (2014)</p>

<p>Psychology</p>	<p>Historical origins in Medicine and/or Philosophy Initial re-definition in late XIX century through Physiology and Anatomy Disciplinary autonomy gained through experimental protocols Definition of experimental protocols in the US context after Second World War</p>	<p>Experimental and quasi-experimental approach Contextual generalizations (nomothetic)</p>	<p>Practical acceptance of the regularity principle Complex psychological constructs built up by aggregation of more observable variables Laboratory setting</p>	<p>No impact Dissolution of subject (Derrida, Foucault) not accepted Strong distance from Psychoanalysis helped to maintain disciplinary integrity Acceptance only in minority positions (mainly in Psychotherapy)</p>	<p>Hilgard (1987) Danziger (1990) Ross (1991) Platt (1996) Haskell (2000)</p>
<p>History</p>	<p>Initial conceptualization of the work of historians strongly rooted in the philological analysis of written sources No need to differentiate from related disciplines</p>	<p>Broad convergence around the archive model Enrichment of the methodology via: - microhistory (Ginzburg, Levi) - collective memory (Halbwachs) - debate on revisionism (vs. Shoah) as lacking scientific foundations</p>	<p>Rejection of the neopositivism (historical truth not subject to causal confirmation) but strong emphasis on the scientific nature of historical work</p>	<p>Rejection of the post-structuralist thesis of no difference between fiction and historical texts (White) Fierce reaction based on methodological criteria of archival research</p>	<p>Carr (1961) Momigliano (1984) Novick (1988) Farge (1989) Appleby, Hunt and Jacob (1994) Eley (1996) Iggers (1997) Evans (1997, 2001) Chartier (2009) Tosh (2010) Jenkins (2003) Potin (2013)</p>

(continued)

Table 1 (continued)

Discipline	Institutionalization process	Methodological orientation	Position with respect to neopositivism	Position with respect to post-structuralism	Selected references
Anthropology	<p>Historical origins in the study of civilizations without written records</p> <p>Progressive separation from Sociology</p>	<p>In-depth case studies with strong theoretical orientation (Boas)</p> <p>Thick description (Geertz)</p> <p>Participant observation and ethnographic techniques</p> <p>Idiographic orientation</p>	<p>Initial orientation towards positivistic approach (Boas), then rejected in favour of the study of unique, non-inferential objects of analysis</p> <p>Rejection of nomothetic approach by using the case study and thick description methodology</p>	<p>Some limited impact</p> <p>Some acceptance in minority positions of hermeneutics and critical thinking</p> <p>Overall orientation aimed at preserving the material constitution of societies (against relativism)</p>	<p>Geertz (1973, 1983)</p> <p>Thoresen (1975)</p> <p>Fowler (1975)</p> <p>Clifford and Marcus (1986)</p> <p>D'Andrade (1995)</p> <p>Barnard (2000)</p> <p>Eriksen and Nielsen (2001)</p> <p>Darnell (2001, 2008)</p> <p>Barth et al. (2005)</p> <p>Clifford (2005)</p> <p>Kuklick (2008)</p>

<p>Political science</p>	<p>Historical origins in post-Second World War theory of democracies and their constituent elements (party systems, elections, government etc.)</p> <p>Need to sustain the comparison with hard sciences in US departments</p> <p>Need to differentiate from Sociology (on one side) and Economics (on the other side)</p>	<p>Three main methodological orientations</p> <ul style="list-style-type: none"> - empirical/quantitative approach - rational choice - historical/comparative 	<p>Challenge by positivism strong in the US academic environment after Second World War</p> <p>Initial acceptance of quantitative approach in US political science</p> <p>Some reaction from European continental tradition rooted in history and political philosophy</p> <p>Strong controversies around rational choice and game theory modelling</p>	<p>Minor impact</p> <p>Minority positions use hermeneutics to study political orientations of social movements</p>	<p>Lerner and Lasswell (1951)</p> <p>Bernstein (1976)</p> <p>Liteberson (1985)</p> <p>Easton, Gunnell and Graziano (1990)</p> <p>Gunnell (1990)</p> <p>Ross (1991)</p> <p>Platt (1996)</p> <p>Lindblom (1997)</p> <p>Smith (1997)</p> <p>Haskell (2000)</p> <p>Apter (2001)</p> <p>Scott and Keates (2001)</p>
--------------------------	--	--	---	--	--

(continued)