Received: 5 July 2024 | Accepted: 6 July 2024

DOI: 10.1111/spsr.12618

SPSR Swiss Political Science Review
Schweizerische Zeitschrift für Politikwissenschaft
Revue Suisse de Science Politique
Rivista Svizzera di Scienza Politica

DEBATE

The Mis-conceptualisation of Societal Impact: Why the Swiss Approach to Societal Impact is Productive and not Inexistent

Michael Ochsner

FORS & University of Lausanne

Correspondence
Michael Ochsner, Université de Lausanne, FORS, Michael Ochsner, International Surveys, Géopolis, CH-1015 Lausanne.
Email: ochsner@enressh.eu, michael.ochsner@fors.unil.ch

Funding information
European Cooperation in Science and Technology, Grant/Award Number: CA-15137 European Network for Research Evaluation in the Social Sciences and Humanities (ENRESSH)

Abstract

Societal impact as a buzz word is high on the agenda of policy-makers around the world. Often, the UK Research Excellence Framework (REF) is cited as the initiator of making the societal impact of research relevant and visible. Equally often, it is said that in Switzerland societal impact is not yet considered in higher education policy. In this paper, I show that both claims are as wrong as they are common. I argue that the UK REF's "impact agenda" is strongly linked to a specific ideology and does not represent the only approach to valorizing and fostering the research-society nexus. By pointing out some major issues in impact evaluation and presenting how the research-society nexus is discussed in Swiss science policy as a contrasting case, I sketch an approach to impact evaluation based on the role of research in society that considers different forms of knowledge generation and dissemination.

KEYWORDS

International comparison, Research quality and metrics, Responsible research assessment, Societal impact

Zusammenfassung

Gesellschaftsrelevanz oder «Societal Impact» steht ganz oben auf der Agenda der Wissenschaftspolitik. Oft wird das britische Research Excellence Framework (REF) als erster und komplettester Ansatz für die Evaluation von Auswirkungen von Forschung auf die Gesellschaft genannt. Ebenso häufig wird behauptet, dass in der Schweiz «Societal Impact» in der Wissenschaftspolitik keine Rolle spiele. In diesem Beitrag wird gezeigt, dass diese

Behauptungen ebenso falsch wie verbreitet sind. Es wird argumentiert, dass die «Impact Agenda» des REF auf einer bestimmten Ideologie basiert und weder den ersten noch den einzigen Ansatz darstellt, die (Wechsel-)Wirkungen zwischen Forschung und Gesellschaft aufzuzeigen und zu fördern. In diesem Beitrag wird ein alternativer Ansatz für die Evaluation von «Societal Impact» skizziert, der auf der Rolle der Forschung in der Gesellschaft aufbaut und verschiedene Formen der Wissensgewinnung und -dissemination berücksichtigt. Dafür werden zuerst einige Probleme der aktuellen Ansätze aufgezeigt und dann die Diskussion in der Schweizer Wissenschaftspolitik zum Gesellschaft-Forschungs-Nexus präsentiert.

Résumé

L'impact sociétal est un terme en vogue dans l'agenda de la politique de la recherche. Le Research Excellence Framework (REF) britannique est souvent cité comme étant à l'origine de la mise en évidence de l'impact sociétal. Tout aussi souvent, il est dit qu'en Suisse, l'impact sociétal n'est pas encore pris en compte dans la politique de la recherche. Dans cet article, je montre que ces deux affirmations sont aussi erronées que courantes. Je soutiens que le REF est fortement lié à une idéologie spécifique et ne représente pas la seule approche permettant de valoriser l'impact sociétal. En soulignant certains problèmes de l'évaluation d'impact et en présentant la discussion du lien entre la recherche et la société dans la politique scientifique suisse, j'esquisse une approche de l'évaluation d'impact basée sur le rôle de la recherche dans la société qui prend en compte différentes formes de génération et de diffusion des connaissances.

Riassunto

L'impatto sociale è in cima all'agenda dei responsabili delle politiche di ricerca in tutto il mondo. Spesso si cita il Research Excellence Framework (REF) britannico come l'iniziatore del processo di valutazione dell'impatto sociale della ricerca. Altrettanto spesso si dice che in Svizzera l'impatto sociale non sia ancora considerato nella politica della ricerca. In questo articolo dimostro che entrambe le affermazioni sono tanto sbagliate quanto comuni. Sostengo che l'"impact agenda" del REF britannico è fortemente legata a una specifica ideologia e non rappresenta l'unico approccio per valorizzare e promuovere il nesso ricerca-società. Evidenziando alcuni problemi principali nella valutazione d'impatto sociale e presentando come il nesso ricerca-società sia discusso nella politica scientifica svizzera, delineo un approccio alla valutazione d'impatto basato sul ruolo della ricerca nella società che considera diverse forme di generazione e diffusione della conoscenza.

INTRODUCTION

Societal impact is an important buzzword in science policy, internationally as well as nationally. Moreover, it gradually trickles down to the researchers themselves, as they increasingly need to provide indications of the societal impact of their research in institutional evaluation exercises or, even more often, estimates for the prospective impact of their research projects when they apply for funding.

But what exactly is societal impact? And how can we identify it? Surprisingly little is known about it even among experienced researchers and evaluators (see, e.g., Wróblewska, 2021) — and even less do those who ask for it have a clear idea: the definitions in evaluation procedures and funding schemes remain more than vague. Therefore, a discussion as to what societal impact means and how it can be identified is urgent, given that societal impact is already a criterion in many evaluation procedures and funding streams on the national as well as the supranational level (de Jong & Muhonen, 2020; Donovan, 2014; Oancea, 2014). The criterion "societal impact" in research evaluation has been introduced so rapidly that its theorisation lags behind (Donovan, 2019). Existing research on the conceptualisation of societal impact, especially its historical dimension (Frodeman, 2017; Gedutis et al., 2023), remains largely unknown to scholars, evaluators, and policymakers. Due to this lack of theorisation, we can only deduce the shortcomings of the practice in impact evaluation from a series of negative steering effects identified in evaluation research (see, e.g., Derrick et al., 2018), a situation that is similar to the application of simple bibliometric measures like Impact Factors or h-index, the shortcomings of which societal impact evaluation is thought to compensate (e.g. with the establishment of the Coalition for the Advancement of Research Assessment, www.coara.eu).

In this contribution, I acknowledge that the societal impact of research is an important topic, but I argue that the current discussions lack conceptualisation. While restricted in space and, thus, necessarily incomplete, the paper presents some conceptual reflections on societal impact that help to understand the research–society nexus with a special focus on improving the evaluation of societal impact.

The paper is structured as follows. Firstly, I will analyse the main assumptions of the current dominant interpretations of societal impact and the so-called "impact agenda". In the second section, I will present practical issues in the evaluation of societal impact. Thirdly, I will situate the Swiss way of impact assessment in the international discourse and show that, contrary to what is often argued, there are more approaches to fostering the relationship between research and society than the UK's Research Excellence Framework (REF) or the Netherlands' Standard Evaluation Protocol (SEP, now named the Strategic Evaluation Protocol). Finally, I will conclude with pointing out the relevance of a rigorous conceptualisation of societal impact and close with a few recommendations regarding the evaluation of societal impact.

THEORETICAL ISSUES OF THE "IMPACT AGENDA"

The idea of societal impact has entered academia not only quickly but also sustainably. It has become such an important buzzword in science policy, so that experts in research evaluation use the term "impact agenda", introduced by Times Higher Education in 2009 (Donovan, 2014). The term refers to the fact that under governmental pressure, Research Councils UK started to ask all research grants to describe ex ante the expected impact that research would have on the economy and society. This was expanded around 2011 to also include impact on culture and was subsequently introduced to the ex-post evaluation of impact on the economy, society and culture in the UK REF (see Donovan, 2014). The term "impact agenda" was later used widely in the context of the implementation of economic and societal impact as an important criterion in other national evaluation procedures and

the EU Framework Programmes. For this paper, I distinguish two main lines of argument as to why societal impact needs to be included in research evaluation. Both are interlinked by the notion of taxpayer accountability and, thus, reflect a particular mode of institutional or organisational management (Dahler-Larsen, 2012; Deem et al., 2007). The first is more theoretical and based on the idea of the emergence of a knowledge society and the duty of research to advance the economy as a link between research evaluation and societal impact. The second does not consider *why* research needs to be evaluated; rather, it just starts from the fact that it *is* evaluated and that current evaluation methods have unintended effects because they exclusively gratify academic relevance but ignore societal relevance. In this section, I will outline the theoretical issues surrounding the two lines of argument for the inclusion of societal impact in research evaluation.

"The New Production Of Knowledge" Under Conceptual Scrutiny

The impact agenda (Donovan, 2014) is strongly influenced by Gibbons et al.'s (1994) seminal work *The New Production of Knowledge*, in which they suggest that at the end of the twentieth century, a new way of conducting research emerged (p. 3). Besides the traditional way of conducting research (what they call "Mode 1"), research now also addressed current problems in society (what they call "Mode 2"). In their own words, "Mode 1 addresses problems set and solved by academics, is disciplinary, its actors are homogeneous, the organisation is hierarchical and preserves its form, and quality control is self-referential, remaining within the academic realm. […] Mode 2, in contrast, addresses problems set in a context of application, is transdisciplinary, characterised by heterogeneity, the organisation is heterarchical and transient, and is socially accountable and reflexive in its quality control" (Gibbons et al., 1994, p. 3). The two modes, of course, exist alongside each other, but the new feature was that research and academia were not self-sufficient anymore. The European Framework Programmes as well as the REF in the UK, to name only two examples, draw strongly on the ideas presented in that book, but much research on societal impact is also following this line of thinking. The terms applied, however, differ, such as "knowledge co-creation", "triple helix", "societal relevance" or "productive interactions". These terms were sometimes developed independently, sometimes as a result of criticism of the notion of Mode 2. However, despite the differences, the main arguments remain the same (for an overview see Sivertsen & Meijer, 2020, p. 68). Therefore, I will use the example of Gibbons et al. (1994) to illustrate the conceptual issues surrounding the theoretical — or rather ideological — foundations of the impact agenda.[1]

Gibbons et al.'s (1994) main claim is that there has been a historical change in knowledge production[2] since the late 1980ies (e.g., p. viii) from a) academia-centred to practice-oriented, b) disciplinary to interdisciplinary, c) full academic freedom to external influence, and d) academic merit to external accountability. They assert that this change is irreversible (p. 11). Interestingly, these clearly historical theses are presented without much historical evidence: the earliest reference dates back to World War II and most arguments are simply presented as claims remaining unsupported by academic references. Gedutis et al. (2023) demonstrate that there is ample evidence prior to the 1980s on their four theses, showing that their line of argument is invalid. Therefore, we need to reflect on what this means for the evaluative procedures of societal impact if they are influenced by this line of argument.

[1]This section is a summary of a more in-depth analysis published in Gedutis et al. (2023).

[2]They focus mainly on science, technology, engineering and medicine, but present also a chapter on the social sciences and humanities.

The first claim, *historical change towards practice orientation*, is currently a very popular argument: that the time of *research for research* is over and that today is the era in which researchers finally leave the ivory tower (see, e.g., Oliver-Lumermann & Drori, 2021). As the reader might well know, this discussion is, however, as old as European[3] science (e.g. debates between sophists and philosophers) and has returned several times since, e.g. with the importance of practice orientation during the formation of modern universities in Europe and the US (see, e.g., Bruston, 1994, pp. 9–10; Kerr, 1991, p. xi), as well as the irreversibility of the primacy of useful research over the quest of searching for truth, especially in authoritarian regimes (e.g. Krieck, 1941, pp. 1, 8). This tells us that the idea of an irreversible "change" has already been there before and it had been reversed. In fact, it rather looks as though the post-WWII period was the exception (if there ever was research for research without any expectations from politics and society).

The second claim, *the novelty of interdisciplinarity*, is based on the argument that until recently (i.e. 1980s), research has been disciplinary, but today's complex problems need to be addressed by several disciplines together (Gibbons et al., 1994, p. 3). The two underlying assumptions of this claim are easily proven wrong: the first assumption, namely that a discipline is something clearly defined, is in stark contrast to the diversity of disciplinary classifications across countries (see, e.g., Guns et al., 2018). The second assumption, namely that interdisciplinarity is something new, disregards that most disciplines today have been interdisciplinary fields in their early years, e.g. sociology or environmental sciences. This suggests that the idea that disciplinary research does not suffice to solve current problems, which is a distinct characteristic of Mode 2, is nothing new, but rather the way in which science advances.

The third claim, *the emergence of external influence*, states that until the late 1980s, scientific revolutions "never posed a threat to the existing order" (with the exception of the social sciences and humanities, though), but would start doing so (Gibbons et al., 1994, p. 100). Evidence against this claim is easy to find. In fact, the claim shows incredible disrespect towards so many scholars who suffered for defending their scientific breakthroughs. Besides the obvious examples from Socrates to Galileo, Eremeeva (1995) provides examples of external influence on the discipline of astronomy through history. Besides such prohibitive influence, Kerr (1991) shows examples of prescriptive influence in post-Civil-War US (1870–1910) (p. xi).

The fourth claim, *evolution from self-referential peer review to accountability*, states that new criteria for evaluation (i.e. accountability) are needed because traditional academic peer review is disciplinary, while Mode 2 research needs to be marketable. However, I would argue that accountability is nothing new (which follows on from the sections above), although I would agree that what is new is how "accounts" are interpreted, i.e. with marketability and return on investment. Putting aside the simple question as to whether accountability is new or not, the fact that research needs to prove its effectiveness and efficiency is not a convincing argument for a change from Mode 1 to Mode 2. It applies to Mode 1 research as well. Rather than resulting from a shift in knowledge production that would, for epistemological reasons, require more accountability, the expectation that research must show its efficiency comes from a political turn towards new public management (NPM), as Deem et al. (2007) and Bulaitis (2020) show in detail. That is to say, whereas academic quality assurance was based on professional standards and intra- and interorganisational collaboration, and professionals saw themselves as independent practitioners holding up best practices, quality assurance after NPM posits that everything is marketable, and professionals are service providers with entrepreneurialism and innovation as predominant values. The detail needed for such evaluations suggests an

[3]Going beyond Europe, while interesting, is not necessary for the argument.

objectivity of measurement and asks for indicator-based evaluation (Dahler-Larsen, 2012), i.e. bibliometrics and scientometrics. Such suggested objectivity masks the values that underlie any evaluation and, thus, the political aspects that come with decision making and control (Dahler-Larsen, 2012, pp. 234–238). Furthermore, it replaces trust in institutions with distrust. However, the people do not gain more control; instead, they simply need to trust the institutions conducting the evaluations (Deem et al., 2007, p. 24). Thus, the claim that such evaluation regimes and managerial governance lead to democratisation (see also Nowotny, 2003) is lip service, because it has little to do with democratisation, but rather with demagogy: one gives the impression that the population receive more power through service orientation and its evaluation, whereas, in fact, procedural quality controls are lifted and the decision power is shifted away from the academic community, which is very diverse in itself, to evaluative agencies and scientific policy managers, i.e. technocrats with specific ideological backgrounds. A true democratisation of institutional control would instead imply that the population have a word in what values count and what procedures seem to be adequate to assure the upholding of those values (Dahler-Larsen, 2012).

The Insufficiency Of Bibliometric Research Evaluation

The second line of argument is, by nature, less theoretical or, rather, marked by a lack of theory. It starts from the observation that current research evaluation focuses strongly on bibliometric assessments and, thus, only takes academic relevance into account, whereas the role of academia goes beyond self-sufficient knowledge production. Focusing only on academic relevance or reputation in evaluation produces incentives to concentrate uniquely on scientific communication, or even on a small part of it represented by specific indicators. This argument is present among different stakeholders, such as policy advisors and policymakers (see, e.g., KNAW, 2005; Spaapen et al., 2007), bibliometricians and evaluators (see, e.g., Leiden Manifesto, Hicks et al., 2015), scholars (see, e.g., DORA Declaration, www.sfdora.org), and university administrators (see, e.g., the CoARA principles, www.coara.eu). The argument seems to be valid. Indeed, simple bibliometric assessment using Impact Factors and/or h-index clearly leads to negative steering effects. For example, researchers evaluated by Impact Factors tend to focus on writing for journals with a high impact factor, which means reducing their work to specific topics, theories and methods published by these journals; restrict the reporting of their findings to significant results, not reporting non-significant results as they are usually not published in such journals; concentrate on English language and on communication to academics; reduce their efforts in teaching etc. Summarising the literature on effects of indicator-driven evaluations, de Rijcke et al. (2016) distinguish several types of negative effects, such as goal displacement, task reduction, bias against interdisciplinarity on the individual level – but there are also effects on the institutional level, such as head-hunting prolific "star scientists", changes in institutional collaboration etc. However, this is not only because bibliometric measurement does not account for societal impact. Rather, it is because bibliometrics is data-driven and it is not clear as to what bibliometric measures actually do measure, given the weak connection between indicators and concepts (Brooks, 2005; Donovan, 2007). This leads to the fact that only what is measured by a few specific indicators becomes visible and is valued. However, research quality is a complex, multidimensional concept, as conceptual research shows, and only about half of the relevant aspects of research quality are open to quantitative measurement (Ochsner et al., 2012).

Similarly, the approach to societal impact evaluation is data-driven, rather than being based on a convincing theoretical foundation. The statement that research ultimately needs to improve society does not represent a theory. A theory would be making a link between knowledge production and effects on society and having ideas as to how "improving society"

could be defined. Moreover, reflections on why evaluation must take place would be needed: what is the role of evaluation and what is the "role of science" as the object to be evaluated? Evaluation is not a God-given necessity but is there to fulfil a role that differs across contexts. For example, if in one country, academics should be pushed to do research as traditionally they have been teaching, an evaluation will introduce incentives to do research and to engage with other researchers. In countries in which research was focused on national contexts, internationalisation would be an important part of evaluation, while in countries that are highly internationally oriented, evaluation might bring back incentives to provide local relevance to research. Evaluations are linked to values that they incorporate, evaluations should therefore include debates about such values rather than being just compliance exercises. Dahler-Larsen (2012) and Sigurðarson (2020), for example, provide theories on what roles research plays in society and what evaluations do or can fulfil, but impact evaluation is far from being in line with their findings or from providing alternatives. Without knowing what one wants to measure, and in the absence of reflections on the quality of the data used and a clear measurement model, the validity and reliability of the measures applied remain debateable (Ochsner, 2022). This also applies to qualitative assessment by peers, as peers seem not to be clear about what they actually judge (Ma et al., 2020; Wróblewska, 2021). For example, an interviewee in Wróblewska (2021) notes that "we realised (…) they [the HEFCE, i.e. the institution that implemented the UK Research Excellence Framework (REF) introducing societal impact evaluation] didn't have any more of an idea of this [societal impact as an evaluation criterion] than we did! It was almost like a fishing expedition" (p. 3). Ma et al. (2020) report that reviewers in ex ante impact evaluation tend to opt for tangible economic impact instead of intangible general societal impact and for short-term instead of long-term impact because they are easier to evaluate, even though they were asked to evaluate prospective societal impact (p. 435). Thus, instead of evaluating whether research contributes to the role research should fulfil in society in the respective discipline, both metric and review-based evaluations of societal impact focus on easily available data or evidence, leading to a similarly reductionist approach to evaluation as the use of simple bibliometrics in academic impact evaluation: the focus lies on data availability instead of evaluating what one conceptually wants to evaluate. This opens the doors wide for negative steering effects, such as goal displacement or focus on immediately visible impact to the detriment of more significant but also more diffuse impact, achievable only through long-term interaction and collaboration (see, e.g., Watermeyer, 2016).

Finally, with regard to the SSH, the idea that research needs to be in relation with society is not new. In fact, in the social sciences and humanities, the introduction of bibliometrics has led to devaluing the societal interactions of researchers. Writing in local languages, publishing with non-scholarly publishers, and focusing on local topics were disincentivised, as they were not covered by bibliometric databases (Nederhof, 2006). It therefore might be a good idea to examine already existing practices in those disciplines that are used to interact closely with society and where the evaluation of research has already included such aspects — but in entirely different ways, abstaining from the idea that everything has to be given a manifestable "worth" (see, e.g., Bulaitis, 2020). Arguments that the SSH "have failed to characterize the ways in which these sciences generate public value" (Esko & Miettinen, 2019, p. 295) seem to be ironic but are often used so that the SSH do not fall off the bandwagon in the current research policy discourse (Benneworth, 2015). The question is rather whether the concept of "public value" needs to be rethought than the alleged need for the SSH to integrate a concept that is actually not clearly defined. Importantly, teaching is often excluded from societal impact of research as research and teaching are often seen as separate by policy-makers, not the least because in many countries the two "missions" of universities are governed by different ministries. However, teaching should not be seen as separate from research, at least in SSH disciplines, such as political science (Hug et al., 2013) as one cross-fertilises the other. Furthermore, teaching includes the education

of future civil servants, politicians, managers etc., thus bringing the concepts and findings from research into practice.

Consequently, it becomes clear that the addition of societal impact to academic relevance without any clear conceptual scrutiny does not represent a solution to the problem that evaluation does not reward good research and its diffusion into society, but rather worsens it. It adds complexity and more possibilities for negative steering effects.

CONCEPTUAL ISSUES IN PRACTICAL EVALUATION OF SOCIETAL IMPACT

The section above showed that the theoretical basis of impact evaluation is weak. Thus, it is not surprising that there are conceptual issues in identifying societal impact. If the theory is lacking, measurement or identification is like stabbing in the dark. In the following I mention a few core issues.

Focus On Tangible Products

Many societal impact evaluations overemphasise tangible impacts, which has led to much criticism from scholars and policy advisors alike (e.g. Bulaitis, 2020; Swiss Science and Technology Council, 2013). What is useful and who decides it remain in the dark, leading to systemic inequality and epistemic injustice (Fricker, 2007), i.e. that some disciplines, sub-disciplines, or even research questions are favoured over others, simply because their impacts are tangible and easy to show.[4]

Negative Impact Identified After Progress In Knowledge

A focus on proving positive impact, e.g. through impact stories or simple metrics, such as patents or licensing, might cause the perception to deviate from the potential negative impact of research. The discussion on this issue has led to the term "Grimpact" (Derrick et al., 2018; Frodeman, 2017). Using three case studies, Derrick et al. (2018) show that there are different reasons as to why research might have a negative impact. For example, it might be that a result from a single study is well publicised and has an impact but is later shown to be wrong. However, as the result has already been taken up by society, it has a lasting impact, such as vaccine rates decreasing after a study on the measles, mumps and rubella (MMR) combined vaccine showing side effects that were later shown to be unrelated.

Opportunity Cost Of Not Funding

Funding decisions impact what objects (and subjects) are studied. Of relevance is not only what *is* studied, but also what *is not* studied. It might seem to be more attractive to study how to optimally design a specific artificial intelligence service because a concrete, tangible outcome can be shown. It might be less attractive to fund research investigating how people use artificial intelligence tools, as the use of such tools might be secondary to the actual tool (or new

[4]See Ma (2022) for bibliometrics; Sigurðarson (2020) goes further and suggests that unmasking epistemic injustices in societies is one form of impact that SSH research can have.

tools). However, not funding such research leads to artificial intelligence services that (re)produce social stigma and discrimination (Leavy, 2018).[5]

External Mechanisms (Impact Is Not Inherent In Research)

Why a societal impact may be identified can lie (and most likely does lie) well outside of the reach of researchers. For example, Sivertsen and Meijer (2020) describe a case of impact that is notable. An international team including Norwegian scholars documented the archaeological site of Palmyra just before the Islamic State destroyed the site. The case was promoted as an exceptional impact by the Norwegian Research Council in an exercise of impact evaluation. Obviously, had the site not been destroyed, the story would most likely have gone unnoticed, like many other archaeological documentations. Thus, what is relevant is decided by context (rather than scientific facts). A focus on evaluating manifest societal impact thus pushes researchers to prioritise topics anticipating such impact, preferring research focusing on already known problems (instead of identifying new ones). This results in science always being a step too late, as it will be reactive (rather than proactive).

Perceptions Of What Is "Good" For Society Change Over Time

The current impact agenda assumes that what is "good" for society is clear and stable. But what is seen to be "good" can, of course, change over the course of time: nuclear weapons can be used as an extreme example, some methods in psychiatry, such as the treatment of hysteria in women, as another. Obviously, a change of ideology will change what is perceived to be "good". However, the impact agenda seems to avoid exactly such a discussion on what is "good" and why — a discourse typical to SSH and important for society (Frodeman, 2017).

Wrong Promises

Especially ex-ante impact evaluation, i.e. evaluation of the potential societal impact of research projects not yet conducted but applying for funding, comes with the risk of exaggerated promises as to what will result from the project (see, e.g., Holbrook & Frodeman, 2011). A prominent example is the Human Brain Project, which promised to simulate a human brain and, thus, to find cures for many brain issues and to further artificial intelligence. However, after funding was received, the interdisciplinary project was downsized to a technological project by its leader, excluding psychology and the neurosciences, and the leader admitted that the promises made to obtain funding were impossible to achieve. This led to a debate, not only in academia but also in public, on the sense and nonsense of huge projects (see Ochsner, Balaban et al., 2023). The way research funding is distributed, especially when focusing on societal impact, leads researchers to give risky promises for strategic reasons to gain funding, such as in the Human Brain Project (Panese, 2015). This can lead to mistrust in science as the public gets used to the fact that promises by scientists are exaggerated (Frodeman, 2017).

[5]Note that this contribution was initially written before the introduction of ChatGPT. It is interesting to see how the implementation of ChatGPT quickly raised the awareness of the problem and led to quite a lot of research on the topic. However, before it impacted society, the imbalance between research developing AI models and tools versus studying potential negative impacts was massive.

Attributability

The discourse surrounding the societal impact of research assumes that societal changes are a consequence of an identifiable research project. However, it is very difficult to prove or identify such a direct link, which is the most discussed conceptual issue surrounding impact evaluation in the literature. Four problems are often mentioned: attribution, causality, internationality and timescale (Sivertsen & Meijer, 2020). Attribution means that it is almost always impossible to separate the impact of research from other inputs and activities; causality refers to the relationship between research and impact being complex and often nonlinear; internationality reminds us that the impact of research is international, as the value chains are global and interconnected; timescale tells us that impacts of research normally extend over long time periods. I subsume all four problems under attributability, as they all thematise the difficulty of attributing an effect in society to a research project. These issues are linked to a basic misperception of what science is: science is not a single project, but rather a discourse. Each piece of a puzzle might add up to an impact that may occur after a specific project, but many projects all over the world have contributed to it. If a single project, isolated from others, presented a result that has impact, it could not be considered scientific. Let us assume that a research project finds a link between a certain diet and long COVID; if this project was presented to authorities and the respective food was prohibited (societal impact), it would not be on scientific grounds. Knowledge first needs to be discussed and confirmed by other studies using different methods and approaches to be considered scientific (see Latour, 2022). Peer review cannot replace scientific discourse — it can only verify whether the study is of sufficient quality — but different theoretical or methodological approaches, contexts, and samples can, and usually do, lead to different results. Establishing scientific knowledge thus takes time.

The Importance Of Non-Impactful Research

Evaluating the impact of research ignores the collaborative nature of science. Scientific knowledge production is not based on individual efforts, even though also in science the myth and importance of "genius" are present. Still, as the saying "standing on the shoulders of giants" suggests, scientific knowledge production is linked to previous research. Contrary to the saying above and a widely held belief in science and society, it is not only ground-breaking or relevant research that advances knowledge production. As any practising researcher knows, finding new insights includes choosing the wrong path. Conducting research along promising avenues even when they fail to produce a result is relevant as well in the knowledge production process: it tells what avenue not to take. However, it also happens that an avenue that did not lead to success for one research endeavour might prove to be relevant for another. As knowledge production is inherently collaborative (even when in competition), researchers communicate failures in presentations and discussions, which is important for the research process. If evaluation focusses on impactful research only, the communication of negative results ceases to happen, which contributes to the problem of positive reporting bias (Song et al., 2010; Sterling, 1959). Formulated positively, research needs serendipity to fully develop its potential.

Need To Let Policy-Makers Take The Credit

An issue of particular relevance to the social sciences and, specifically, political science is the fact that research aiming at policy change needs uptake by policymakers to be impactful. However, in policy-making, citation does not have the same ethical standing

as in science. Often a researcher wanting to impact policy will try to influence policy-makers through small steps and constant interaction. The most successful way in which to let research influence policy is to make policy-makers believe that it was their own idea, as they will then defend the policy intrinsically. Often the best way is to stay in the shadows and let the policy-maker take the credit. It is they who take the responsibility for the political decision after all. To my knowledge, this is a point that is surprisingly rarely discussed in impact evaluation, even though any scholar active in the political realm has most likely made this experience (see Cairney & Kwiatkowski, 2017; Dahler-Larsen, 2023). This means that real impact might evade evaluation procedures focusing on demonstrable cases of impact.

Science And Policy-Making Need To Be Separate

One of the most important points, however, and one that stands totally at odds with the concept of Gibbons et al. (1994) as well as with most impact evaluation procedures, is that impact evaluation mixes the scientific and the political (Frodeman, 2017). Impact discourse perceives the idea of scientific impact to be apolitical, just as it perceives the activity of evaluation to be apolitical. The argument goes that objective results need to lead to evidence-based policy-making. However, which evidence and policy outcomes are considered relevant does not (only) come down to objectivity, but is largely a political issue (Dahler-Larsen, 2012). The implementation of research in policy or society at large needs negotiation and political deliberation in a democracy. If research is not implemented because society is not ready (yet), it is beyond scholars' control. Science and politics need to be kept separate (see, e.g., Heinich, 2021) for two reasons: firstly, in democracies, decisions require political legitimacy; secondly, evidence needs to be independent from politics if the goal is evidence-based policy-making (Ochsner, Bulaitis et al., 2023). This is also one reason as to why scientists need to let policy-makers take the credit, as policy-makers should pick up results from scientists (more than one at best) and take them through the deliberation process. From the information from scientists to the implementation, policy-makers are responsible for the political process that science cannot bypass.

Many Pathways To Impact

There are manifold pathways to impact. A linear conceptualisation from research to impact is certainly too simplistic, and the concept of productive interactions does not cover all scenarios of how research might impact society. An empirical analysis of 60 European impact cases alone resulted in the identification of 12 pathways to impact, besides the classical linear one (Muhonen et al., 2020). For example, the “seize the day” and “anticipating anniversaries” pathways show that expected or unexpected events in society can lead to an impact of research that is conceived or recognised because of those events, or the “knowledge creeps into society” pathway shows that constant information can influence societal discourse and that this discourse can be taken up by research and, over time, impact creeps into society.

THE PRACTICE OF IMPACT EVALUATION: SWITZERLAND IN THE INTERNATIONAL LANDSCAPE OF IMPACT EVALUATION

When speaking about impact evaluation, two examples are usually used in Europe: the Research Excellence Framework in the United Kingdom (REF) and the Standard Evaluation

Protocol from the Netherlands (SEP). They are different in many aspects but similar in their ideological foundation. In both cases, societal impact is an important goal of research, and, in the name of accountability, impact is to be demonstrated or at least predicted in a plausible way. What is less known is that there are other approaches to impact evaluation in other countries.

The Two Best-Known European Examples

For reasons of space, describing the REF and the SEP in detail is not possible. Furthermore, it is not necessary because there are many scientific articles and books written on them. However, they illustrate two approaches to impact evaluation. The REF represents an implementation of the idea of accountability to society as argued in Gibbons et al. (1994) but conceptualising all research as Mode 2, while the SEP represents the approach that bibliometrics is not sufficient and needs to be complemented by societal impact.

The concept of impact was introduced to the national evaluation procedure in the UK for the REF 2014, defined as "an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment, or quality of life, beyond academia" (HEFCE, 2011, p. 26).

For the REF, case studies must be presented that show a societal impact of the research submitted to the exercise. The case studies are evaluated by review panels. Societal impact accounts for 20% in the REF 2014 and 25% in the REF 2021. The results of the REF are directly linked to the funds that universities receive according to their achievement in the exercise.

The SEP works differently. Started in 2003, it is, like its predecessors,[6] an evaluation exercise that is not directly linked to funding. Rather, "improvement and accountability are the main objectives of this system of quality assessment" (VSNU et al., 2003, p. 7). Moreover, it is noteworthy that the exercise does not only examine the performance of researchers; rather, "the system is directed toward both the research and its management" (VSNU et al., 2003, p. 7). Research is evaluated according to four criteria, among which is *relevance*, that is, scientific and socioeconomic impact. The evaluation is to be quantitative and qualitative. In the 2015–2021 SEP, the criteria were reduced to two: *research quality* and *societal relevance*, each evaluated across the following dimensions: demonstrable products, demonstrable use of products, and demonstrable marks of recognition (VSNU et al., 2016, p. 25).

What is common between the two exercises is that they focus on accountability in their argumentation and ask for demonstrable signs of a societal impact of research produced during the last cycle (around 6 years). The difference is that the REF focuses on peer review judging scientific and societal impact and, thus, suggesting that all research be Mode 2 in essence. The SEP evaluates research quality by means of scientometric indicators (and judges those indicators qualitatively) and complements this scientometric input with indicators of societal relevance, thus allowing for different missions of institutions (or institutes): some might focus more on scientific quality, while others focus on societal relevance.

Switzerland In The Landscape Of European Societal Impact Evaluation

The REF has been (aggressively) promoted in science policy and exported to Norway, Hong Kong, Sweden, Latvia, and Poland (Wróblewska, 2021). Thus, when speaking about the

[6]In the Netherlands, research evaluation has included the idea of societal impact evaluation since the early 1990s (van Drooge et al., 2013).

evaluation of societal impact, the science policy literature mainly reports REF-like procedures, sometimes with the SEP as an alternative (Sivertsen & Meijer, 2020; Wróblewska, 2021). Most analyses of national evaluation systems focus on some specific "national" evaluation procedure that represents a "coherent set" of evaluation procedures (Whitley, 2007). However, research is a complex endeavour, and researchers are not only active in their "nation" but also embedded in a regional setting (especially when thinking about societal impact), active in an international disciplinary community, evaluated by European, national and international funding agencies, evaluated in addition within the context of their institution (which usually fulfils a specific mission), and — finally — sometimes subject to a centralised national evaluation procedure (Ochsner & Peruginelli, 2022). In all countries, several evaluation procedures are in place, rarely forming a coherent set. The emphasis on impact differs across countries and evaluation procedures, as well as ranging from attributing points to each scholarly action (e.g. the Czech Republic, see Šima & Daniel, 2023) to deliberately not centralising evaluation (e.g. Ireland, see Ma, 2023). Switzerland is a particular case because it is a federal state, the evaluation of research is subject to laws on different administrative levels, and there is no centralised national research evaluation procedure. Therefore, whenever Switzerland enters the analysis, it is usually claimed that there is no evaluation procedure or, at least, no impact evaluation (see also Flinders in this Debate). However, the situation is more complex (see Ochsner, Balaban et al., 2023). It is not that societal impact does not play a role. In fact, there is a clear division of tasks between full universities and universities of applied sciences. The former focus on basic research and, therefore, societal impact plays a less important role, while the latter focus on applied research and societal impact is central. What is more, there are two different competitive funding institutions, namely the Swiss National Science Foundation (SNSF), which focuses on basic research, and InnoSuisse, which funds mainly applied research. However, and this is key, while there is a clear division of tasks, the division is not absolute. Researchers from universities of applied sciences and full universities can submit proposals to both funders; furthermore, there are funding streams in the SNSF that have an applied nature or aspect. However, the notion of "impact" is different from that of the REF, as it is not related to manifest (economic) impact, but rather refers to a topic that is currently in need of knowledge production and not just the object of pure scientific interest. Finally, the two funding organisations have common funding streams that mainly focus on the transition between knowledge production and knowledge transfer (e.g. BRIDGE, 2020).

In general, Swiss science policy does not have "no emphasis on impact" (e.g. Bandola-Gill et al., 2021, p. 227); rather, it actively decided against a simplistic concept of societal impact, as it is implemented in the REF. There are longstanding discussions on the research–society nexus, especially because the political system of a direct democracy in a federal state needs public discourses on many issues on the political agenda and, thus, needs to inform the public about the scientific state of the art. This means that the direct democratic federal political habitus is also present in science policy and, therefore, top-down approaches are often politically unsuccessful (see Ochsner, Balaban et al., 2023). Instead, bottom-up approaches are chosen, including different stakeholders in the process of policy-making, among them the scientific community. In this process, the research–society nexus is approached from different scientific perspectives in policy papers that influence the discourse (e.g. Swiss Science and Technology Council, 2013), and several conferences and workshops are organised (e.g. Nauer, 2019). At the same time, each university is obliged to evaluate its research, but there is no centralised evaluation scheme. Each university has its own evaluation procedure adapted to its mission (e.g. Loprieno et al., 2016, p. 16). Whether or not societal impact plays a role, and if so in what form, differs from institution to institution. This results in an adaptive evaluation procedure; claiming that there is "no emphasis on impact" is therefore only correct if one restricts evaluation to a centralised national evaluation scheme, "impact" to a simplistic definition of a manifest or

demonstrable short-term causal effect on society, and "university" to full universities. For the Swiss case, such a restriction would be absurd, as the whole higher education system is built on the idea of not defining those terms too narrowly.[7] Rather, the Swiss evaluation system is influenced by the federal organisation and direct democracy prevalent in Switzerland and is focused on mission orientation; in other words, each university has its particular *raison d'être*, which means different focuses, goals and roles in (local) society (see Loprieno et al., 2016). It is based on the idea that impact is an interplay of academic freedom, interaction between stakeholders, and (public) discussion — and the method is that of constant self-reflection and discourse. A recent example of how bottom-up discussions in scientific institutions are taken up in policy decisions is the adoption of recommendations of different scientific actors by the Federal Council regarding the institutionalisation of scientific consulting to political actors after the experience of the COVID-19 crisis (Bundesrat, 2022, p. 8).

NEGATIVE EFFECTS OF IMPACT EVALUATION

Why put so much emphasis on academic freedom and discourse instead of demonstrable effects? If scientists receive public money, they are surely responsible for showing that they are worth their investment. What is the difference, ultimately, from the Dutch SEP, which also focuses on productive interaction, co-production, and reflection but provides empirical evidence of claims instead of merely talking?

The differences between the Swiss (or Irish) way of (impact) evaluation and the Dutch SEP, however subtle, are key. It is the notion that societal impact is not measurable, at least not within a useful timeframe. Rewarding manifest societal impact pushes scholars to work on topics that may find impact, diffuse their research widely before the findings are confirmed, team up with questionable companies to achieve impact, etc. This is what Derrick et al. (2018) describe as "Grimpact", i.e. research having societal impact that is later understood to be negative.

However, the need to prove impact can even lead to negative impact if the research in itself and its actual impact are great. The reason for this is that a story needs to be told that convinces policymakers, the public and evaluators. I use the example of an impact movie (ESRC Social, 2017). Such movies are produced for the REF; there is even an industry for it. The movie has won several impact awards. It is about a project led by an English PI that aims at a new social work approach called *cash plus care* to reduce the risk of HIV in Sub-Saharan Africa. The movie begins by describing the issues of sugar daddies, sexual abuse, etc. It continues showing miserable young black women, then the interaction with scientists, and, finally, smiling African children running through nature while mentioning the English funder and other foundations. Whereas the research is certainly great, I would like to point to the highly problematic storytelling: African men are portrayed as sexual abusers, African women as sexual objects, Africans as being poor and miserable and not able to control their sexuality. Then, the British scientists arrive and save them from misery. This story is emphasised visually: whenever the narrator mentions a *problem*, *need* or *impact*, a black person is in focus, and whenever *research* or a *solution* is mentioned, a white person is presented. This reaches its climax at the time of 2:59 when two women are sitting at a table, and when the narrator says "the *impact* of this *research*", the focus shifts from the black woman (*impact*) to the white woman (*research*)! Note that both women are most likely from the scientific team.

[7]However, instead of "societal impact", the term "innovation" is preferred, which is not to the advantage of the SSH.

The goal is not to accuse the project — I am convinced that the research was excellent and had an important impact. What I would like to point out is the toxicity of the conception of demonstrable impact. The researchers did what they needed to do (good research, and then trying to gain recognition at their institution for it through the REF); the video directors did what they needed to do (creating an excellent video that tells a story that catches the attention and creates positive emotions); and the juries of the contests did what they needed to do (selecting a project with convincing evidence of impact). However, all of this creates a neocolonialist story that wins prizes. No word in the video mentions whether and how the British or European past and present might add to the problem, and no white person is shown when sexual abuse is mentioned. The video is exactly successful because it feeds (Western) stereotypes. A further question is how such a video was accepted by the project team, who certainly were very aware of all those issues — and they must have noticed them. I suspect, however, that the pressure to score well in the REF was too high to be critical. And this is where we are in trouble!

NOW WHAT? SUGGESTIONS FOR A SUSTAINABLE APPROACH TO THE RESEARCH–SOCIETY NEXUS

The REF approach to impact evaluation is only one of several possibilities — there are many other approaches to impact evaluation across Europe (see Ochsner & Bulaitis, 2023). As has been shown, the theoretical foundations of the approach to demonstrating the societal impact of research are weak and linked to a specific political and economic theory. In particular, the common argument of the historical novelty of the societal relevance of research is false; in fact, this argument, together with the TINA (There Is No Alternative) principle, leads to an absence of historical analysis that comes with the perception that alternatives seem not to be available, while the risks and opportunities of different approaches cannot be weighed in the decision on how to evaluate research. In fact, Gedutis et al. (2023) show that, over time, several approaches to the research–society nexus have been proposed and practised. They can be classified across the relationship between truth and power. The REF approach is characterised by "power over truth". It sets the policy-makers on top who decide what is considered to be "useful" research. This is a setting that has been used often, e.g. the pragmatist movement in philosophy, but also during the Nazi regime or the Soviet Union. However, there are alternatives that have been in use over the course of history (see Table 1 in Gedutis et al., 2023), e.g. "truth over power", which prevailed after WWII, where scientific knowledge based on academic freedom influenced policy, rather than policy influencing science (e.g. Bush, 1945). Theorising impact evaluation should start with analysing those alternative approaches.

Not only does the evaluation of demonstrable societal impact suffer from a weak theoretical foundation regarding knowledge production and the research–society nexus, it also comes with the risk of negative steering effects, some of which are already visible, as the examples of "Grimpact" (Derrick et al., 2018) and the award-winning impact video show. I think that the claim that the societal impact of research needs to be demonstrated comes with several category mistakes, leading to wrong conclusions and prescriptions as to what research should do. In the following I point out those that strike me as the most relevant.

Every project needs to have an impact. Scientific knowledge production is a collaborative action. This applies not only to the more empirical disciplines, but also to disciplines that are characterised by individual scholarship. Scientific knowledge can only be considered established knowledge when it has undergone scholarly debate (e.g. Latour, 2022). Every project is embedded in a scientific discourse; thus, it is not the project that needs to impact society, but rather the discourse, i.e. established knowledge.

All research needs to have an immediate impact. There are different types of research fulfilling different functions. Basic research is guided by scientific interest, applied research aims at solving concrete problems, regulatory research aims at addressing questions arising from policies, and practice-based research addresses practical problems and is related to professions. Given their different functions, their evaluation should differ (see also Ochsner, Bulaitis et al., 2023).

Scientists are bad at communication. Scientists are as good or bad at communication as non-scientists — there is a large diversity. But first and foremost, scientists need to communicate science to scientists. Communication to the public is a different (and highly demanding) task. Communication within science should not be confounded with the communication of established knowledge to the public (e.g. when confounding open access publications with dissemination beyond academia, as, for example, in RISIS, 2019). If scholars are good at vulgarisation, it should be rewarded. But not all scientists should be expected to be good scientific journalists.

Research needs to be freely available and formulated so that it is understood by all. Open access and open science are concepts with their own merit. But they should not be confounded with the popularisation of science. Asking that scientific papers stop using "jargon", as heard sometimes by advocates of open access, is absurd. We do not ask plumbers to stop using specialised tools, but instead use a normal hammer, merely because we want to understand what they do. The public does not need to know the internal discussions in scientific discourse (i.e. the details of every journal article), but the public needs to learn about the established knowledge, something that we had to learn the hard way during the pandemic (see Ochsner, Balaban et al., 2023).

Science needs to support evidence-based policy. Policy is only evidence-based if the evidence itself is based on independent research. However, in the context of impact evaluation, this becomes a complex story. Researchers must anticipate what policy-makers want to hear in order to provide impactful research. If research is tailored to policy-makers' expectations, we cannot talk of evidence-based policy, as the evidence is policy-based. This is a point that is particularly relevant in the context of the European Framework Programmes, in which calls are formulated that are closely aligned with desired policy outcomes.

Research needs to provide an immediate solution. Often an argument is made that research needs to provide a solution; otherwise, it would be better to spend money on practice. For example, is a research project on working conditions of nurses good enough or should the money rather be spent on nursing, i.e. actual help for persons? Of course, the issue is not limited to the single project on working conditions of nurses. The decision on how much money is spent on research or on nursing is a political one. Indeed, research might lead to changes later even if it has no immediate impact on nursing. While the effect is not immediate, it is general, and research can be considered an investment; by contrast, spending money on helping persons is money spent on single cases with no lasting effects. Both are relevant but different in nature.

CONCLUSION

Given all this criticism: can we then evaluate the societal impact of research? My suggestion is that one should think bigger than trying to distinguish between impactful and non-impactful research or researchers. Societal impact is something that a field of research, a discipline, or science as a whole achieves. The problem is not how to demonstrate or measure impact, but rather the short-sighted and toxic concept of accountability measured through key performance indicators. That is to say, the problem lies within a specific ideology that everything can and should be measured in economic terms, i.e. as value for money. It is simplistic to claim that, in the name of democracy, each researcher as a public

servant needs to prove the worth of the money spent on them (see, e.g., Gibbons et al., 1994; Nowotny, 2003). In reality, such a claim relies on the unquestioned primacy of economic thinking, confounds the global public good of knowledge with a marketable commodity (Marginson, 2013), and reduces "public good" to the policy-makers' definition of what is useful (rather than a definition that would emerge from debates in the public sphere). This is problematic if we bear in mind that public trust in politicians and businesses is much lower than trust in science or universities. In Switzerland, 74% of respondents to the high-quality register-based MOSAiCH 2020 survey trusted universities (at least a 6 on a scale of 0 to 10) in comparison to 65% for parliament, 41% regarding trust in the media, and 36% in businesses (for the data see Ernst Stähli et al., 2021). The same holds true for other European countries: in all 18 countries having taken part in the European Social Survey (ESS) Round 10 (2021) including the relevant questions, trust in scientists is higher than in politicians: the mean difference ranges between 1.5 in Switzerland to 4.5 in Portugal. 63% (Slovakia and Northern Macedonia) to 92% (Norway) of the population give a 6 or more on a scale from 0 to 10 regarding trust in scientists, while trust in politicians ranges from 7% (Bulgaria) to 56% (Norway; European Social Survey Research Infrastructure, 2023, using design and post-stratification weights). Evaluation should focus on what researchers do, rather than what society and politicians do with their research. It should serve the discourse surrounding what is relevant regarding the evaluated research and its object (see also Galleron et al., 2017). If this is reduced to a tangible or even monetary return, the research–society nexus is reduced to unidirectional impact, ignoring processes, interactions, etc. Instead of *accountability*, I therefore suggest focusing on the norm of *societal responsibility*. Empirical research on quality criteria for SSH research suggests a conceptual difference between societal impact and societal relevance, the first being demonstrable impact, in the form of change, which comes with all of the issues underlined above; the second is the idea that research is linked in some way to society (Hug et al., 2013; Ochsner, 2022). Societal responsibility would put emphasis on the second: researchers should reflect on how their research relates to society, but it also entails an ethical component that includes responsibility in the use of resources.

The evaluation of research, in my opinion, certainly should include notions of the research–society nexus. It therefore needs to take processes and activities equally into account as manifest outputs, given the many pathways to impact (Ochsner, Bulaitis et al., 2023). Most importantly, research that includes societal actors, produces outputs for the general public, results in policy advice, etc. should be seen as equally performative as research that generates many citations. However, countering the dysfunctionality of an invalid measurement of academic impact with an equally invalid measurement of societal impact does not improve an evaluation procedure. Therefore, evaluation should avoid making the same mistakes with societal impact assessment as with the focus on bibliometrics in academic impact assessment: reducing research outcomes to single outputs, focusing on what is measurable and for which data are available instead of what is relevant. Instead, evaluation should be reflexive, dynamic, and context-dependent, taking into account whatever is relevant for knowledge production and for achieving the tasks that researchers have to fulfil. Not only are evaluations an instrument of control, evaluations should also serve a learning function, by helping researchers to become better at what they do. Therefore, they need to be related to what researchers do. Research quality is a concept in the social sciences like any other concept; therefore, the same methodological requirements apply. First and foremost, for each evaluation situation, a careful definition of what quality of research and societal quality mean needs to be established and indicators must reflect the criteria derived from this definition.[8] Without a clear definition of quality, the meaning of quality remains opaque

[8]See Ochsner (2022) for methodology and examples; for a systematic review of criteria see Hug & Aeschbach (2020).

— and thus indisputable. The discussion on criteria is thus the democratic aspect of research evaluation, in which the values of research are negotiated (Dahler-Larsen, 2012) — the more inclusive the discussion on criteria is, the more we can achieve a democratic basis of research evaluation. This entails that stakeholders in evaluation, i.e., the researchers to be evaluated, the entity that evaluates (university, funder, ministry) as well as other groups involved (patient organisations, professions), need to define what is expected from the evaluated unit and clarify the criteria used (such as scientific rigour, vision of future research, applicability in professions etc.). Such criteria need to be carefully chosen and operationalised with methodologically rigorous processes, i.e. indicators must be attributed to those criteria that they can measure. Indicators that are not theoretically and empirically linked to criteria of research quality or societal impact do not validly reflect those concepts and, even though seemingly providing comparable numerical values, cannot inform us validly about research quality or societal impact.

From this perspective, the Swiss way to research quality (Loprieno et al., 2016) and the Swiss approach to societal impact, while far from being perfect, are in my opinion not an "inexistent" national evaluation system that puts "no emphasis on societal impact", but quite the opposite: a national evaluation system that puts (direct) democracy centre stage and remains adaptive and responsive to international, national, regional and disciplinary needs by focusing on a mission-oriented evaluation without top-down definitions and not too much emphasis on simplistic indicators. Through that approach, many category mistakes can be avoided or at least highlighted, and critical debates can be initiated. Such an approach gives researchers more opportunities to do what is relevant to conduct impactful research in their discipline or topic, rather than follow check-lists for successful careers that are unrelated to epistemological characteristics and impact pathways in their field. Researchers should focus on what is relevant to achieve the goals in their field rather than on feeding indicator systems or narrative logics unrelated to the practices in their field. What should be gratified in evaluations is the *diversity* of research activities and outputs that combine to (co-)produce and disseminate research and to interact with stakeholders (Ochsner, Bulaitis et al., 2023).

FUNDING INFORMATION

This chapter is partly based upon work from COST Action CA 15137 'European Network for Research Evaluation in the SSH (ENRESSH)', supported by COST (European Cooperation in Science and Technology).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in SwissUbase at https://doi.org/10.23662/FORS-DS-1232-1, reference number 1232 and in the European Social Survey Data Portal at https://doi.org/10.21338/ess10e03_2.

ORCID

Michael Ochsner https://orcid.org/0000-0002-1758-9590

REFERENCES

Bandola-Gill, J., Flinders, M., & Anderson, A. (2021). Co-option, control and criticality: the politics of relevance regimes for the future of political science. *European Political Science*, *20*(1), 218–236. https://doi.org/10.1057/s41304-021-00314-0

Benneworth, P. (2015). Tracing how arts and humanities research translates, circulates and consolidates in society. How have scholars been reacting to diverse impact and public value agendas? *Arts and Humanities in Higher Education*, *14*(1), 45–60. https://doi.org/10.1177/1474022214533888

BRIDGE. (2020). *Regulations on BRIDGE Proof of Concept grants*. BRIDGE.

Brooks, R. (2005). Measuring University Quality. *The Review of Higher Education*, *29*(1), 1–21. https://doi.org/10.1353/rhe.2005.0061

Bruston, A. (1994). Avant-propos. In F. Dubet, D. Filâtre, F.-X. Merrien, A. Sauvage & A. Vince (Eds.), *Universités et villes* (pp. 7–12). L'Harmattan.

Bulaitis, Z. (2020). *Value and the Humanities*. The Neoliberal University and Our Victorian Inheritance. Palgrave Macmillan.

Bundesrat. (2022). *Wissenschaftliches Potenzial für Krisenzeiten nutzen. Bericht des Bundesrates in Erfüllung der Postulate 20.3280 Michel vom 5. Mai 2020 und 20.3542 De Quattro vom 8. Juni 2020*. Schweizerische Eidgenossenschaft.

Bush, V. (1945). *Science, the endless frontier*. Princeton University Press.

Cairney, P., & Kwiatkowski, R. (2017). How to communicate effectively with policymakers: combine insights from psychology and policy studies. *Palgrave Communications*, *3*, Article 37. https://doi.org/10.1057/s41599-017-0046-8

Dahler-Larsen, P. (2012). *The Evaluation Society*. Stanford University Press.

Dahler-Larsen, P. (2023). Can we use deliberation to change evaluation systems? How an advisory group contributed to policy change. *Evaluation*, *29*(2), 144–160. https://doi.org/10.1177/13563890231156955

Deem, R., Hillyard, S., & Reed, M. (2007). *Knowledge, higher education, and the new managerialism*. The changing management of UK universities. Oxford University Press.

Derrick, G. E., Faria, R., Benneworth, P., Budtz Pedersen, D., & Sivertsen, G. (2018). Towards characterizing negative impact: Introducing Grimpact. *STI Conference Proceedings*, 1199–1213.

Donovan, C. (2007). The qualitative future of research evaluation. *Science and Public Policy*, *34*(8), 585–597. https://doi.org/10.3152/030234207X256538

Donovan, C. (2014). Impact Agenda. In J. B. Holbrook (Ed.), *Ethics, Science, Technology and Engineering: A Global Resource*, 2nd Edition. Palgrave Macmillan.

Donovan, C. (2019). Assessing the broader impacts of publicly funded research. In D. Simon, S. Kuhlmann, J. Stamm & W. Canzler (Eds.), *Handbook on Science and Public Policy* (pp. 488–501). Edward Elgar Publishing.

ESRC Social. (2017, June 22). *Preventing HIV risk in Southern Africa | Celebrating Impact*. [Video]. Youtube. https://www.youtube.com/watch?v=57mxuFs5bJo

Eremeeva, A. I. (1995). Political Repression and Personality: The History of Political Repression against Soviet Astronomers. *Journal for the History of Astronomy*, *26*(4), 297–324. https://doi.org/10.1177/002182869502600402

Ernst Stähli, M., Sapin, M., Pollien, A., Ochsner, M., & Nisple, K. (2021). *MOSAiCH 2020. Measurement and Observation of Social Attitudes in Switzerland. Study on Environment and related topics* [Dataset]. FORS - Swiss Centre of Expertise in the Social Sciences. https://doi.org/10.23662/FORS-DS-1232-1

Esko, T., & Miettinen, R. (2019). Scholarly understanding, mediating artefacts and the social impact of research in the educational sciences. *Research Evaluation*, *28*(4), 295–303. https://doi.org/10.1093/reseval/rvz018

European Social Survey European Research Infrastructure (ESS ERIC). (2023). *ESS10 integrated file, edition 3.2* [Data set]. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/ess10e03_2

Fricker, M. (2007). *Epistemic Injustice. Power and Ethics of Knowing*. Oxford University Press.

Frodeman, R. (2017). The impact agenda and the search for a good life. *Palgrave Communications*, *3*, palcomms20173. https://doi.org/10.1057/palcomms.2017.3

Galleron, I., Ochsner, M., Spaapen, J., & Williams, G. (2017). Valorizing SSH research: Towards a new approach to evaluate SSH research' value for society. *Fteval Journal for Research and Technology Policy Evaluation*, *44*, 35–41. https://doi.org/10.22163/fteval.2017.274

Gedutis, A., Bulaitis, Z. H., & Ochsner, M. (2023). The need for historical inquiry in societal impact evaluation: towards a genealogy of the notion of useful research. In M. Ochsner & Z. H. Bulaitis (Eds.), *Accountability in Academic Life. European Perspectives on Societal Impact Evaluation* (pp. 30–50). Edward Elgar.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzmann, S., Scott, P., & Trow, M. (1994). *The new production of knowledge. The dynamics of science and research in contemporary societies*. Sage Publications.

Guns, R., Sile, L., Eykens, J., Verleysen, F. T., & Engels, T. C. E. (2018). A comparison of cognitive and organizational classification of publications in the social sciences and humanities. *Scientometrics*, *116*(2), 1093–1111. https://doi.org/10.1007/s11192-018-2775-x

HEFCE. (2011). *REF 2014: Assessment framework and guidance on submissions (updated to include addendum published in January 2012). REF 02.2011*. HEFCE.

Heinich, N. (2021). Ce que le militantisme fait à la recherche. *Tracts* (No. *29*). Gallimard.

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Ràfols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, *520*(7548), 429–431. https://doi.org/10.1038/520429a

Holbrook, J. B., & Frodeman, R. (2011). Peer review and the ex ante assessment of societal impacts. *Research Evaluation*, *20*(3), 239–246. https://doi.org/10.3152/095820211X12941371876788

Hug, S. E., & Aeschbach, M. (2020). Criteria for assessing grant applications: a systematic review. *Palgrave Communications*, *6*, 37. https://doi.org/10.1057/s41599-020-0412-9

Hug, S. E., Ochsner, M., & Daniel, H.-D. (2013). Criteria for assessing research quality in the humanities: a Delphi study among scholars of English literature, German literature and art history. *Research Evaluation*, *22*(5), 369–383. https://doi.org/10.1093/reseval/rvt008

Jong, S. P. L. de, & Muhonen, R. (2020). Who benefits from ex ante societal impact evaluation in the European funding arena? A cross-country comparison of societal impact capacity in the social sciences and humanities. *Research Evaluation*, *29*(1), 22–33. https://doi.org/10.1093/reseval/rvy036

Kerr, C. (1991). *The great transformation in higher education. 1960–1980*. State University of New York Press.

KNAW. (2005). *Judging research on its merits. An advisory report by the Council for the Humanities and the Social Sciences Council*. Royal Netherlands Academy of Arts and Sciences.

Krieck, E. (1941). *Nationalpolitische Erziehung*. Armanen Verlag.

Latour, B. (2022). En mode Bruno Latour 2/4 : La science a-t-elle toujours raison ? In A. van Reeth (Ed.), *Les Chemins de la philosophie*. France Culture.

Leavy, S. (2018). *Gender bias in Artificial Intelligence: The need for diversity and gender theory in machine learning*. IEEE/ACM 1st International Workshop on Gender Equality in Software Engineering, Gothenburg.

Loprieno, A., Werlen, R., Hasgall, A., & Bregy, J. (2016). The "Mesurer les Performances de la Recherche" Project of the Rectors' Conference of the Swiss Universities (CRUS) and Its Further Development. In M. Ochsner, S. E. Hug & H.-D. Daniel (Eds.), *Research Assessment in the Humanities. Towards Criteria and Procedures* (pp. 13–21). Springer International Publishing.

Ma, L. (2022). Metrics and epistemic injustice. *Journal of Documentation*, *78*(7), 392–404. https://doi.org/10.1108/jd-12-2021-0240

Ma, L. (2023). Responsible societal impact (without) evaluation: reflections from Ireland. In M. Ochsner & Z. H. Bulaitis (Eds.), *Accountability in Academic Life. European Perspectives on Societal Impact Evaluation* (pp. 127–138). Edward Elgar.

Ma, L., Luo, J., Feliciani, T., & Shankar, K. (2020). How to evaluate ex ante impact of funding proposals? An analysis of reviewers' comments on impact statements. *Research Evaluation*, *29*(4), 431–440. https://doi.org/10.1093/reseval/rvaa022

Marginson, S. (2013). The impossibility of capitalist markets in higher education. *Journal of Education Policy*, *28*(3), 353–370. https://doi.org/10.1080/02680939.2012.747109

Muhonen, R., Benneworth, P., & Olmos-Peñuela, J. (2020). From productive interactions to impact pathways: Understanding the key dimensions in developing SSH research societal impact. *Research Evaluation*, *29*(1), 34–47. https://doi.org/10.1093/reseval/rvz003

Nauer, H. (2019). Was nützt Forschung? In *Meinungen und Themen aus den Geistes- und Sozialwissenschaften* [blog post]. https://wissenschaftskultur.blogspot.com/2019/06/was-nutzt-forschung.html

Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics*, *66*(1), 81–100. https://doi.org/10.1007/s11192-006-0007-2

Nowotny, H. (2003). Democratising expertise and socially robust knowledge. *Science and Public Policy*, *30*(1), 151–156. https://doi.org/10.3152/147154303781780461

Oancea, A. (2014). Research Assessment as Governance Technology in the United Kingdom: Findings from a Survey of RAE 2008 Impacts. *Zeitschrift für Erziehungswissenschaft*, *17*(6), 83–110. https://doi.org/10.1007/s11618-014-0575-5

Ochsner, M. (2022). Identifying research quality in the social sciences. In T. C. E. Engels & E. Kulczycki (Eds.), *Handbook on Research Assessment in the Social Sciences* (pp. 48–66). Edward Elgar.

Ochsner, M., Balaban, C., & Iseli, M. (2023). Societal impact, innovation, or public value? Switzerland's approach to research impact evaluation and the SSH. In M. Ochsner & Z. H. Bulaitis (Eds.), *Accountability in Academic Life. European Perspectives on Societal Impact Evaluation* (pp. 204–223). Edward Elgar.

Ochsner, M., & Bulaitis, Z. H. (Eds.). (2023). *Accountability in Academic Life. European Perspectives on Societal Impact Evaluation*. Edward Elgar.

Ochsner, M., Bulaitis, Z. H., Balaban, C., Castro-Martínez, E., Daniel, O., Gedutis, A., Giménez-Toledo, E., Iseli, M., de Jong, S., Ma, L., Mañana-Rodríguez, J., Muhonen, R., Olmos-Peñuela, J., Peruginelli, G., Sigurðarson, E. S., Šima, K., Spaapen, J., Vanholsbeeck, M. (2023). Manifesto for a Better Societal Impact Evaluation. In M. Ochsner & Z. H. Bulaitis (Eds.), *Accountability in Academic Life. European Perspectives on Societal Impact Evaluation* (pp. 10–29). Edward Elgar.

Ochsner, M., Hug, S. E., & Daniel, H.-D. (2012). Indicators for Research Quality in the Humanities: Opportunities and Limitations. *Bibliometrie - Praxis und Forschung*, *1, Article 4*. https://doi.org/10.5283/bpf.157

Ochsner, M., & Peruginelli, G. (2022). National research evaluation systems and the social sciences. In T. C. E. Engels & E. Kulczycki (Eds.), *Handbook on Research Assessment in the Social Sciences* (pp. 416–433). Edward Elgar.

Oliver-Lumermann, A., & Drori, G. S. (2021). *From Ivory Tower to Academic Commitment and Leadership*. Edward Elgar.

Panese, F. (2015). Cerveau et imaginaire sociotechnique: le Human Brain Project entre science et politique. In M. Audétat (Ed.), *Sciences et technologies émergentes. Pourquoi tant de promesses ?* (pp. 165–193). Les Éditions Hermann.

Rijcke, S. de, Wouters, P. F., Rushforth, A. D., Franssen, T. P., & Hammarfelt, B. (2016). Evaluation practices and effects of indicator use—a literature review. *Research Evaluation*, *25*(2), 161–169. https://doi.org/10.1093/reseval/rvv038

RISIS. (2019). Exploring knowledge production in Europe. The RISIS-KNOWMAK Tool. *RISIS Policy Brief Series* No. 2.

Sigurðarson, E. S. (2020). Capacities, capabilities, and the societal impact of the humanities. *Research Evaluation*, *29*(1), 71–76. https://doi.org/10.1093/reseval/rvz031

Šima, K., & Daniel, O. (2023). Societal impact evaluation in SSH in the Czech Republic: in defence against the strong STEM application discourse. In M. Ochsner & Z. H. Bulaitis (Eds.), *Accountability in Academic Life. European Perspectives on Societal Impact Evaluation* (pp. 81–95). Edward Elgar.

Sivertsen, G., & Meijer, I. (2020). Normal versus extraordinary societal impact: how to understand, evaluate, and improve research activities in their relations to society? *Research Evaluation*, *29*(1), 66–70. https://doi.org/10.1093/reseval/rvz032

Song, F., Parekh, S., Hooper, L., Loke, Y. K., Ryder, J., Sutton, A. J., et al. (2010). Dissemination and publication of research findings: an updated review of related biases. *Health Technology Assessment*, *14*(8), iii, ix-xi, 1-193. https://doi.org/10.3310/hta14080

Spaapen, J., Dijstelbloem, H., & Wamelink, F. (2007). *Evaluating research in context: A method for comprehensive assessment* (2nd ed.). Consultative Committee of Sector Councils for Research and Development.

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance – or vice versa. *Journal of the American Statistical Association*, *54*(285), 30–34. https://doi.org/10.2307/2282137

Swiss Science and Technology Council. (2013). *"Economization" of Science. Recommendations and Proceedings of the Seminar Held in Bern by the Swiss Science and Technology Council on April 23, 2013* (No. 4/2013). SSTC Report.

van Drooge, L., de Jong, S., Faber, M., & Westerheijden, D. (2013). Twenty years of research evaluation. *Facts Figures*, *8*, 1–21. The Hague.

VSNU, NWO, & KNAW. (2003). *Standard Evaluation Protocol 2003–2009 for Public Research Organisations*. VSNU, NWO, KNAW.

VSNU, NWO, & KNAW. (2016). *Standard Evaluation Protocol 2015–2021. Protocol for Research Assessments in the Netherlands*. VSNU, NWO, KNAW.

Watermeyer, R. (2016). Impact in the REF: issues and obstacles. *Studies in Higher Education*, *41*(2), 199–214. https://doi.org/10.1080/03075079.2014.915303

Whitley, R. (2007). Changing governance of the public sciences. In R. Whitley (Ed.), *The changing governance of the sciences. The advent of research evaluation systems* (pp. 3–27). Springer.

Wróblewska, M. N. (2021). Research impact evaluation and academic discourse. *Humanities and Social Sciences Communications*, *8*, Article 58. https://doi.org/10.1057/s41599-021-00727-8

AUTHOR BIOGRAPHY

Michael Ochsner works as a senior researcher at FORS, the Swiss Centre of Competence in the Social Sciences. He is president of the international association "European Network for Research Evaluation in the Social Sciences and Humanities" ENRESSH. His research spans social policy, survey methodology, research policy and research evaluation.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ochsner, M. (2024). The Mis-conceptualisation of Societal Impact: Why the Swiss Approach to Societal Impact is Productive and not Inexistent. *Swiss Political Science Review*, *30*, 223–243. https://doi.org/10.1111/spsr.12618