

On the Shoulders of Giants: Tracing Back the Intellectual Sources of the Current Debate on “GDP and Beyond” to the 19th Century

Nils aus dem Moore*

RWI (Berlin)

Christoph M. Schmidt

RWI (Essen), Ruhr-Universität Bochum, IZA (Bonn), and CEPR (London)

JEL A13; B16; E01

GDP; well-being; quality of life; sustainability; indicator systems.

Received: 21.06.2012

Revision received: 08.10.2012

Accepted: 13.10.2012

Summary

An eternal motive of human existence is the search for guidance. While values and beliefs retain their high relevance, today's enlightened societies also tend to rest their aspirations and decisions on the actual facts and on a sober assessment of possible courses of events emerging from different choices. Given the complexity of modern life, it is by now well understood that this strategy requires objective, comprehensive and accessible statistical reporting. Today, the desire to provide such a valuable basis for individual decisions and policy-making finds one of its most important expressions in the international debate on “GDP and beyond”. In contrast to similar efforts displayed in previous decades, the current projects emphasize sustainability issues and focus on the accessibility of the information, using modern tools of measurement and presentation. Yet, there is ample evidence that even by the mid-19th century economists aspired to use the objectifying power of statistical analysis as an instrument to improve policy-making and to achieve societal progress. Many of the approaches entertained today have thus to be viewed as an extension of attempts started at that time.

1 Introduction

In these early years of the 21st century, the world is searching for new and better guidance. The second global recession that resulted from a financial and banking crisis in the United States, and the growing stock of evidence that the global economy puts substantially more pressure on the ecological systems of our planet than those are able to carry have contributed to a widespread conviction that “business as usual” is no longer a sensible option. The definite direction that the course of events should take is, naturally,

* We are grateful to Andrea Cassel, Claudia Schmiedchen, and Hartmut Westram for their support, and to two anonymous referees and to the editor for their constructive comments and suggestions.

highly controversial. But one lesson from the global economic crisis and the multitude of ecological damages that are looming or already occurred seems to be widely accepted: A compass is needed that provides policy makers, economy and society with a reliable bearing in matters of economic performance, quality of life and sustainability.

Across the globe, governments, researchers and social activists are currently discussing better individual indicators for specific aspects of our present condition in material as well as in non-material dimensions. They are also working on their coherent integration into comprehensive but still accessible measurement systems that should form the basis for evidence-based public deliberation and policy-making. This movement benefits from recent advancements in statistical techniques and indicator research and is based on the widespread utilization of powerful and yet still improving information and communication technologies. The ultimate aim of these initiatives is the establishment of user-friendly statistical monitoring systems that are at the same time comprehensive and comprehensible.

But in a year in which the *Journal of Economics and Statistics* (*JNS – Jahrbücher für Nationalökonomie und Statistik*) completes its first 150 years, it seems worthwhile to look back and search for traces of this debate in earlier times. This article documents that this search turns out to be much more successful, though, than just detecting some vestiges of the current debate. Instead, many of the ideas and concepts discussed today were already present in the late 19th century. We document these intellectual roots meticulously, both with respect to economic reasoning and with respect to the potential of statistical methods to support this endeavor.

And in addition to the intriguing parallels, we also discuss the most important differences that distinguish the current and the past contributions to the issue. Most importantly, we find sustainability issues to have conquered much of the current attention, aspects which were arguably of a less urgent character in the late 19th century. Moreover, for the adept user, the potential to access a wealth of information is bigger than ever quite in contrast to the late 19th century statistician who could only have dreamed of these possibilities. This potential is creating a new challenge, though, since now the task is to utilize the technological capabilities to make the information accessible to a wider public.

The outline of the paper is as follows. The second section presents an overview of the current status and the evolution of the international debate on “GDP and Beyond”, giving an outline of the guiding principles of this work and concrete examples of measurement systems that have been recently proposed. The third section traces the intellectual sources of the current debate back to the second half of the 19th century, focusing for obvious reasons on publications in the *JNS*. Becoming more concretely, section 4 documents for the three spheres of the current debate, i.e. for (i) material well-being and economic performance, for (ii) non-material well-being and quality of life, and for (iii) various aspects of sustainability, that many of the indicators currently being discussed, and the respective reasoning behind them, can be traced back to 19th century precursors debated in the early volumes of the *JNS*. Finally, the concluding fifth section draws some conclusions from this voyage into the history of economic thought.

2 “GDP and beyond”: Evolution and current status of the debate

Despite the great variety of new indicators and dashboards that have been devised to support policy-making and public deliberation, participants in the current debate on improving the measurement of societal progress and human welfare share a common point of departure. Their overwhelming consensus is that in recent decades policy makers, business leaders and economists, as well as the media and society at large, placed too much emphasis on short-run economic performance. Concentrating on measuring the growth rate of Gross Domestic Product (GDP), they neglected to adequately address the long-term. Although economists have pointed out since its inception during WWII that GDP is a measure of economic activity and not of economic well-being, apparently they did not do this fervently enough. Consequently, in the public perception it rose to a quasi-hegemonic status and was often interpreted as a gauge of the overall welfare of economies.

There are good reasons for the assessment that GDP alone is not sufficient to convey a reliable impression of a society’s material wealth and non-material well-being. Inter alia this reservation reflects that (i) in the *economic domain* all non-market activities, such as housework and parenting or unpaid voluntary activities, and a substantial part of publicly provided services in education, health and security remain unaccounted for by GDP, (ii) in the *social domain* the distributions of income and wealth are not documented and inequalities of opportunity are not captured, (iii) in the *ecological domain* damages and losses as well as the consumption of non-renewable resources are not factored in adequately, and (iv) aspects of the *political and institutional domain* that are clearly relevant for overall welfare, like the extent of democratic participation, accountability and the rule of law are not considered whatsoever.

Notwithstanding the dominance of GDP in the public realm, the work on a broader framework for measuring human progress and well-being continued in the background. Scientific expert communities and international institutions have addressed these questions at least since the publication of “The Limits to Growth” by the Club of Rome in 1972 (Meadows et al. 1972). Starting with Nordhaus and Tobin (1972) and their calculation of a Measure of Economic Welfare, a wide array of alternative measurement frameworks and concrete indicators have been proposed, as is documented today in a vast and well-surveyed literature (Fleurbaey 2009). Perhaps the most prominent framework that has been designed to measure well-being, quality of life, human development and sustainable development in the last three decades is the Human Development Index which ranks countries by their level of “human development” through a composite indicator that takes GDP but also health and education into account (Hall et al. 2010: 7).

An important step towards a broader understanding of welfare and progress and the development of adequate indicators has been the establishment of sustainability as a guiding policy principle, at least in rhetorical terms. The famous Brundtland report (WCED 1987) defined “sustainable development” as the kind of development “that meets the needs of the present without compromising the ability of future generations to meet their own needs”. While this definition highlights the intertemporal aspect of the sustainability concept and focuses on the ethical requirement of intergenerational equity, most of the numerous implementations followed a more pragmatic interpretation, aiming at the reconciliation of environmental, social equity and economic demands which came to be known as the “three pillars” of sustainability. Recent implementations often

add governance as a fourth dimension that takes into account aspects of democratic participation, accountability and the rule of law.

By now, many advanced countries have established indicator systems along these lines. As a prominent example, the biennial Monitoring Report of Eurostat covers 111 indicators which are linked to the key challenges and objectives of the EU Sustainable development strategy (European Commission 2011). A consolidation of these numerous indicators into 11 so-called headline indicators and the visualization of their current trends allow a quick overview of the progress made towards the specific goals. Following the same approach, the German government has established a National Sustainability Strategy that is monitored biennially by a progress report released by the Federal Statistical Office. It contains 35 indicators that have been politically agreed upon to cover 21 subject areas (Statistisches Bundesamt 2012).

There is a second reason, in addition to the emergence of the sustainability paradigm, for the recent multiplication of efforts in many countries, by governments on all levels as well as in the civil society, to go “beyond GDP”. This is the growing body of evidence that the tight co-evolution of economic performance with other aspects of material and non-material welfare might have weakened considerably in recent decades, at least in some societies (GCEE/CAE 2010: 6). As long as economic performance and general welfare evidently marched in lock-step, the focus on GDP growth as a proxy for broader welfare was completely warranted. Yet today, ever mounting work pressure, reduced security of employment, the breakdown of family structures, rising poverty rates and the spread of mental-health problems and diseases of civilization have been identified as potential welfare reducing phenomena that are not captured in traditional measures, let alone by GDP (Wahl et al. 2010: 9).

The consequence is obvious: GDP alone is not sufficient to measure adequately the sustainable wealth of nations. This insight is confirmed by popular assessment. Two thirds of the respondents in a Eurobarometer poll of the European Commission said in 2008 that indicators for social, economic and ecologic dimensions should receive equal value in measuring progress. Governments and international organizations tied in with the popular dissatisfaction of traditional welfare measurement and started several initiatives towards the establishment of broader concepts. The Organisation for Economic Cooperation and Development (OECD) has without doubt given a decisive impetus to these endeavors through a series of high-level conferences in Palermo (Italy 2004), Istanbul (Turkey 2007) and Busan (South Korea 2009) and accompanying publications. Most importantly, the “Istanbul Declaration” acknowledged “an emerging consensus on the need to undertake the measurement of societal progress in every country, going beyond conventional economic measures such as GDP per capita” with the aim of producing “high-quality, facts-based information that can be used by all of society to form a shared view of societal well-being and its evolution over time” (OECD 2007).

Within Europe, the European Commission acted as an extra pacemaker. At the conference “Beyond GDP”, Commission President Barroso declared that, in measuring welfare and well-being, the EU must aim at “the sort of breakthrough that we saw in the 1930s, a breakthrough that adapts GDP, or complements it with indicators that are better suited to our needs today, and the challenges we face today” (cited in Kroll 2011: 16). The ensuing report “GDP and Beyond: Measuring Progress in a Changing World” (EU Commission 2009) contained a roadmap with the following five key steps towards a comprehensive and comprehensible reporting on progress: (i) complementing GDP with environmental and social indicators, (ii) providing near real-time information

for decision making; (iii) generating more accurate reporting on distribution and inequalities, (iv) developing a European Sustainable Development Scoreboard; and (v) extending National Accounts to environmental and social issues (Kroll 2011: 16).

The final breakthrough into the public view came with the publication of the so called Stiglitz-Sen-Fitoussi-Report (SSFC 2009). It was conducted by a commission compiled around the two Nobel prize winners Joseph Stiglitz and Amartya Sen, and the French economist Jean-Paul Fitoussi. The commission concluded that “those attempting to guide the economy and our societies are like pilots trying to steering a course without a reliable compass. [...] For many purposes, we need better metrics. Fortunately, research in recent years has enabled us to improve our metrics, and it is time to incorporate in our measurement systems some of these advances” (SSFC 2009: 9).

As an immediate follow-up to the SSFC report, the Franco-German Council of Ministers asked the German Council of Economic Experts (Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung, SVR) and its French counterpart, the Conseil d'Analyse économique (CAE) to develop a concrete proposal how the recommendations of the SSFC report could be implemented. The resulting report was published in December 2010 and proposed a comprehensive set of indicators to measure economic performance, quality of life and sustainability (GCEE/CAE 2010).

In both countries, the quest for a better measurement “beyond GDP” has continued since then with various activities: In France, the national statistical institute INSEE (Institut national de la statistique et des études économiques) has begun to enrich its surveys with questions to include new dimensions that were recommended in the SSFC report (Kroll 2011: 9). In Germany, the Bundestag established the Study Commission on “Growth, Wellbeing and Quality of Life – Paths to Sustainable Economic Activity and Social Progress in the Social Market Economy” that took up its deliberations in January 2011 (German Bundestag 2010). In its interim report the respective working group argued against the construction of a holistic composite index and proposed instead to work out a concrete proposal for an indicator suite that would be as large as necessary and as compact as possible, thereby aiming to strike the ideal balance between comprehensiveness and comprehensibility (Deutscher Bundestag 2012).

Similar activities are under way in many other countries, and a growing number of them has already completed the development phase of their indicator suites or composite indices and presented them to the public (Kroll (2011) provides a comprehensive overview). We restrict ourselves here to examples from four countries (Australia, Canada, UK, US) to document the breadth of different approaches that are currently pursued.

Australia is without doubt among the pioneers for a broader welfare measurement. As early as 1996, there was a Senate Inquiry into “National Wellbeing: A system of national citizenship indicators and benchmarks”, followed by a national conference on measuring progress in 1997 and the release of the publication “Measuring Progress: Is life getting better?” in 1998 (Eckersley 1998). Building on these initiatives, the Australian Bureau of Statistics (ABS) has published the indicator system “Measures of Australia's Progress” on a regular basis since 2002. The official brochure contains 17 headline indicators that are grouped into the three pillars Society, Economy and Environment (ABS 2010).

A second national initiative that has attracted international attention is the “Canadian Index of Well-Being” (CIW) that was started by scientists and social activists at the University of Waterloo in 2001. Ten years later, the first index brochure “How are Canadians *Really* doing?” was released (CIW 2011). The CIW is a composite indicator

calculated by two steps of aggregation: Firstly, 64 baseline indicators that depict the evolution of different aspects of quality of live between 1994 and 2008 are normalized to percentage values and summarized in indicators regarding eight distinct dimensions. Secondly, the overall CIW is calculated as the average of these dimension indicators. Although the authors are well aware of the weighting problem in the construction of their composite indicator (Michalos et al. 2011: 29-31), they nevertheless interpret the difference between the GDP growth of 31 per cent and a substantially smaller CIW rise of only 11 per cent as evidence that GDP overdraws the real gain in quality of live: "(...), our economic performance outpaces our quality of live." (CIW 2011: 12).

In the UK, more emphasis than anywhere else is placed on the integration of comprehensive measures of the country's subjective well-being into the official statistical apparatus. The heavy weight that economists around Sir Richard Layard have given to so-called happiness research at the London School of Economists and the apt popularization of its findings and potential policy implications (Layard 2005) laid the ground for this orientation. In November 2010, Prime Minister David Cameron launched a large-scale initiative for the introduction of a measure of "general well-being" (GWB) that should complement GDP, based on surveys of well-being. Starting in April 2011, 200 000 Britons have been asked the following four questions each year in the Integrated Household Survey, answering on a scale of 0 to 10: (i) "Overall, how satisfied are you with your life nowadays?", (ii) "Overall, how happy did you feel yesterday?", (iii) "Overall, how anxious did you feel yesterday?", and (iv) "Overall, to what extent do you feel that the things you do in your life are worthwhile?" (Matheson 2011).

In the United States, President Barack Obama signed the Key National Indicators Act into law in March 2010, which should lead to the creation of a Key National Indicator System (KNIS). Congress authorized 70 Million US-Dollars of public funding for the project that is carried out by the interdisciplinary National Academy of Sciences in collaboration with the newly founded non-profit institute "State of the USA". The ambition of KNIS is to provide US citizens with statistical information in their country and region on a broad variety of issues in an accessible way on a user-friendly website. The comprehensive system will eventually include around 300 individual indicators to cover the 14 topic areas, even on the disaggregated levels of states, regions and social subgroups (Kroll 2011: 10f.).

3 Intellectual sources of the current debate in the 19th century

Across the globe we currently find ambitious projects that are aiming for a comprehensive and internationally comparable account of the state of economic performance, well-being and quality of life. These initiatives have arguably led to remarkable progress in the quest of statistical analysis to generate deeper insights and to pave the ground for better policy. Yet, they are certainly not the starting point of this endeavor, but rather its most recent expression. Indeed a recurring theme in the current debate regarding the design of indicator systems and the organization of the statistical groundwork behind their publication is the insight that many of the aspects that capture our attention today have been discussed before, sometimes even for several decades. Examples are the social indicator systems devised in the 1960s and 1970s and the indicator systems for environmental sustainability that have been around for more than a decade.

But, what might perhaps come as a very surprising discovery for the protagonists of the current debate, its intellectual sources can be traced backed much further, at least

to the second half of the 19th century. To provide convincing evidence for this claim, our focus in this article will selectively lie on contributions made to the *JNS* during its first decades of existence, along the lines of the following thought experiment: If some of the early contributors to the *JNS* could have used a time machine to arrive in the year 2012, the year preceding the 150th anniversary of this journal, what would they have said to the current state of the debate? More concretely, what would be their assessment of the motives behind these initiatives, of the conceptual ambition of this endeavor, and of the statistical issues pertaining to the construction of the various systems of indicators?

3.1 The motives behind “GDP and beyond”

With respect to their reaction to the motives behind the current discussion, one might fear that we will only be able to speculate. After all, when the *JNS* was founded, German per capita GDP – even though the concept as such was yet to be developed – was arguably much lower than it is today in any OECD country. While contemporaneous observers were rightfully in awe of the technical achievements already reached in their era of “steam and lightning” (Hildebrand 1864: 136, our translation) and the fast growth in economic prosperity observed after the industrial revolution had started, they would probably not have been prepared whatsoever to the exponential growth in prosperity which we have seen since, in particular during the post-WWII era. Maddison (2001), for instance, demonstrates more than convincingly that (not only) in the developed world real per capita income today is multiple times the income experienced one and a half centuries ago.

Thus, as speculative as this has to be, it seems quite likely that our time travelers’ first reactions to the current debate on human welfare would be expressions of astonishment: How could societies which are that rich by any historical standard display such concern about their precarious condition? And, correspondingly, can it actually be true that outside of the current “GDP and beyond” debate the typical focus displayed by economic reports published by international organizations, such as the *Organisation for Economic Co-Operation and Development* (OECD) and the *International Monetary Fund* (IMF), by advisory groups such as Germany’s *Council of Economic Experts* (GCEE), and by basically all economic research institutes lies on deviations of aggregate economic activity from its trend value? That is, the major attention is not given to the level of prosperity, but its growth relative to an upward trend? How rich must human societies have become in the developed world, and how regular and stable must past growth have been?

But after this initial shock subsided, the intellectual basis of the current debate would undoubtedly receive endorsement. Hildebrand (1863a: 14, 17), for instance, acknowledges the dangers to the coherence of any society caused by fast economic development that is carried by strong market forces, rapid investments into the capital stock and turbulent scientific progress. From this vantage point, the major questions of the current debate, (i) “Is the accumulation of material welfare really generating more happiness?”, (ii) “Are there insurmountable, albeit distant limits to growth?”, and (iii) “How could well-being be ascertained without economic growth?” are merely the most recent expression of age-old questions posed by uncountable analysts of human prosperity and of the functioning of society. In addition, 19th century social scientists would certainly be well acquainted with the phenomenon that crises always tend to bring these fundamental questions on human prosperity to the surface (Hildebrand 1863a: 17).

What perhaps most distinguishes the current discussion most from earlier debates on the connex between economic prosperity and well-being is the strong emphasis given today to the theme of economic, social, and environmental sustainability. But since they were confronted with quite similar problems at their time, earlier economists would probably quite easily be convinced that the accumulated stocks of private and public debt currently provide a serious challenge to the objective of ascertaining sustainable public finances, requiring intense analysis and discussion (Müller 1912: 372, 375). Somewhat different, although probably not less supportive, might be the time travelers' reaction to the recent concerns about environmental quality and sustainability, as environmental degradation and, in particular, global challenges such as climate change, have not been an issue during their times. But even though the emphasis on these questions is predominantly a modern development, they also find their reflection in earlier writings (Hildebrand 1863a: 19).

The rest of this section will argue that we find important roots in 19th century thinking for both dimensions of the current debate as well, the rather abstract discussion about what should be the ambition of and the concepts behind such an indicator system on the one hand, and the predominantly applied discussion about the statistical principles and approaches to fulfilling this task on the other.

3.2 The conceptual ambition: “Nationalökonomie”

Much of what we can infer about the position the earlier contributors to the *JNS* would have taken, derives from the articles written by Bruno Hildebrand, the founding editor of the *JNS*. His contributions are acknowledged impressively in an obituary by Johannes Conrad (1878: IV, XI, XIII), his comparably important successor as *JNS* editor, documenting the extraordinary respect Hildebrand enjoyed among his peers. Reading Hildebrand's articles reveals that, irrespective of his personal position on these questions, he would adamantly have maintained that only careful and encompassing statistical analysis could bring about the desired enlightenment in these matters. Already the programmatic “mission statement” (our term) of the *JNS* (Hildebrand 1863d) outlines his vision of the economy as one of the fundamental elements of the life of any society, equal in importance to its language, its literature, its legal system and its art (Hildebrand 1863d: 3), making it not only worthwhile but critical to explore its functioning on the basis of solid statistical analysis.

Hildebrand (1863d: 3) reckons that the most demanding intellectual challenge for economic analysis is the fact that, although the functioning of an economy follows some important regularities, it nevertheless does not obey any fundamental laws. This, he realizes, distinguishes economics from the – by then already comparatively well-established – natural sciences. For him, the reason behind this contrast is clear: It derives from the fact that aggregate economic activity is the condensation of a multitude of individual actions and decisions. Consequently, economics as a science would need to utilize the experiences made in the past in order to identify the regularities characterizing human behavior and the causes and consequences of human actions, with the aim of supporting the principles of individual freedom and individual responsibility. Hildebrand leaves no doubt in his contributions that this objective would necessitate a plenitude of statistical work.

This summary assessment is highly reminiscent of the ambitions behind “GDP and beyond”, and, thus, deserves a closer look. In several highly programmatic articles in the *JNS*, Hildebrand (1863a, 1863c, 1866) carefully outlines the role of economics

and statistics in their historical context in a way that in many respects could serve as a guide to the current debate. As a starting point for his observations, Hildebrand characterizes the – by then – relatively new discipline of economics (“Nationalökonomie” to contemporaneous observers) as the result of the enlightenment movement of the late 18th and early 19th century (1863a: 5-6, 1866: 2). In its early expression, this new discipline had the ambition to equal the natural sciences in its quest for discovering the laws governing human behavior, a perspective that went hand in hand with the idea of an “atomistic society” in which the pursuit of individual utility was the exclusive driving force of all activities (Hildebrand 1863a: 6-7, 1873: 2-3).

Hildebrand quite openly reveals his frustration with the obstinate resistance of many contemporary economists to a more nuanced view on the importance of moral sentiments (Hildebrand 1863a: 8-9), but also offers very little sympathy for the contrasting position taken by contemporaneous socialists that the pursuit of individual utility rather be the root of all evil (Hildebrand 1863a: 9-10). For him, the evidence seems overwhelming that free choice of occupation and division of labor, the uninhibited allocation of production factors according to their marginal returns, steady investments into the stock of productive capital, and incessant scientific progress had become powerful driving forces of economic progress. These developments, which were intimately associated with the intellectual success of contemporaneous economic reasoning, were lifting human productivity to unprecedented heights – and were even viewed to enhance the ability for education and enjoyment in the population (Hildebrand 1863a: 10-14).

But despite his highly critical assessment of the socialist arguments against contemporaneous economic thinking, Hildebrand also explicitly rejects the notion emphasized by Adam Smith and other economists that pursuing an untamed *laissez faire* – which would render any statistical analysis to become a useless husk – could be a sensible foundation of a prosperous society (1863a: 10). He rather views the tendencies of any *laissez faire*-economy to an ever-increasing concentration of economic power and wealth with a serious dose of skepticism, especially since the proponents of *laissez faire* could allegedly refer to unconditioned scientific arguments (Hildebrand 1863a: 14-17). His view is echoed in Knapp (1871: 238-239). Similarly, proponents of a balanced view on economic growth today need to defend their position both against blind believers as well as against fanatic opponents of growth.

As a first conclusion of his considerations, Hildebrand challenges his peers to engage into a critical review of the contemporaneous science of economics, culminating in the clarification of the questions whether human behavior followed strict natural laws and, consequently, economics should be accepted as a natural science, after all (1863a: 19-20). If that were to be the case, neither moral sentiments nor political interventions would have to play any role. Personally, Hildebrand rejects this notion vehemently (1863a: 21-25, 137-140). Rather, in his view individual economic freedom, as worthy of protection as it is, cannot, taken by itself, ascertain societal prosperity and human progress. In addition to the individual pursuit of utility, ethical considerations and a sense of responsibility lie at the heart of any successful society (Hildebrand 1863a: 140-143). Knapp (1871: 241-243, 247) essentially takes the same position.

Today, economics is well established as a social science, not as a natural science, and economists would definitely shy away from arguing that the regularities which their work is uncovering carry the encompassing explanatory power of a natural law. Nevertheless, both in the 19th century and today, the recipients of the empirical results derived by economists and statisticians might misunderstand the limited conclusions facilitated

by the evidence. Knapp (1871: 240), for instance, argues that the idea – prominent among contemporaneous observers – that statistical regularities were proving the notion that individuals acted according to laws of nature, unable to alter their life course by their own will, decisions, or ethical considerations, reflected the incompetence of commentators to see statistical analyses for what they are: While statisticians themselves would rightfully and soberly regard statistics merely as a set of procedures to uncover regularities and causal links, its findings might lead outside observers to inflated conclusions. This assessment is highly reminiscent of the misunderstanding of the concept of GDP as a welfare measure, a misunderstanding that has not been suggested by economists or statisticians, but has nevertheless become the epitome of the motives behind the “GDP and beyond” movement. In the 19th century as well as today, analytical concepts which are taken out of context might lead to completely erroneous conclusions. In particular, many 19th century economists apparently tended to argue that, since economic freedom was widely realized, circumstances could hardly get any better, and there was not need whatsoever for economic or statistical analysis of social inequality. By contrast, Schmoller (1873: 6-7) not only argues that there was actually a deep societal divide, but also blamed the overemphasis on economic performance in the short run to lie behind it. In his view, the exclusive aspiration on fostering economic performance, without giving any consideration to the potentially negative consequences of increasing material wealth for other aspects of human existence was putting social coherence at risk (1873: 10). Similarly, Knapp (1871: 247) observes that the regularities in income inequality and poverty provide evidence enough for concern about the future prosperity of society. Schmoller (1873: 9, 11-12) advocates a strong state as a factor correcting unbearable social inequality, without interfering in the general functioning of the market economy. It is not difficult to detect the current competition between various forms of capitalism, most prominently a US-style free market economy and a continental European-style social market economy, as the present-day expression of this intellectual conflict between the concept of *laissez faire* and its critics. And the balance advocated by Hildebrand and Schmoller in their writing is clearly reflected in the setup of the modern social market economy as well.

Hildebrand’s second, closely related, conclusion regards the agenda for statistical research, and finds its direct reflection in today’s initiatives aiming at the construction of indicator systems measuring the state of human welfare and progress: From the perspective of the current discussion on measuring the state of human welfare, the highly practical consequence of such a balanced position would be that collecting evidence on the genuine state of affairs and their development over time has to be viewed as both worthwhile and necessary. Most importantly, rejecting the notion that human behavior follows natural laws and exploiting the fact that human behavior displays a high degree of regularity are no contradiction whatsoever. Rather, these systematic patterns reflect both regularities in human experiences and similarities in the environment in which individuals make their decisions, and, thus, they offer the opportunity to statistical researchers to discover patterns that can form the basis for good economic policy (Hildebrand 1863a: 142-143, 1863c: 482, 1872: 9-10).

This position on the proper role of the discipline of statistics (“Moralstatistik” to contemporaneous observers) contrasts the scattered statistical attempts that characterized earlier times (Hildebrand 1866: 2) and is emphatically shared by Knapp (1871) and von Neumann-Spallart (1885). In these contributions, statistics is ultimately an instrument that allows abstracting from anecdotes and identifying the general pattern behind

observable phenomena (Hildebrand 1866: 3; Knapp 1871: 248; von Neumann-Spallart 1885: 225). Thus, it objectifies experiences and facilitates the formulation of hypotheses and the test of their implications against the evidence, far beyond the potential offered by mere plausibility considerations (Hildebrand 1863c: 482, 1866: 3, 5). Taken together, these insights on the appropriate role of statistical analysis and the arguments regarding economics as a social science build up a fervent plea for pursuing – in today’s words – evidence-based economic research and policy advice (Schmidt 2007).

In essence, this 19th century literature already appreciates that statistics can be viewed as the reporting system of the economy as a whole, whereby it is learning about the true state of its economic circumstances, based on meticulous accounting and routinely conducted closure of accounts (Hildebrand 1863c: 482). And given that this work is continued on a regular basis, without major gaps, the resulting time series will offer even more opportunity to learn about patterns of stability and instability and perhaps even causal mechanisms underlying the results (Hildebrand 1866: 9). This understanding forms the basis for the systems of national income accounting which were subsequently developed in the 20th century and which are the staples of statistical reporting on human welfare today.

3.3 The construction of indicator systems: “Statistik”

Today it is well understood that any researcher approaching the task of constructing an encompassing system of indicators of human welfare should first consider a range of statistical issues. The most important items on the agenda concern (i) the practical setup of the analysis and the governance of the institutions involved in this work, (ii) the distinction of correlation and causality, and (iii) a clear grasp on sampling error, hypothesis testing, and measurement error. This basis for the statistical work has necessarily to be secured before the applied work is starting, or otherwise either the statistician or the recipients of the results will be headed for disappointment. Again with a focus on the contributions to the *JNS* during its early years in the late 19th century, it will be documented here that many of these issues have already been discussed in the articles published by our statistical forefathers.

Turning, first, to the practical setup of the analysis and the governance of the institutions involved in the construction of indicators and indicator systems, the simple, yet easily forgotten starting point is necessarily a careful conceptual delineation of the task. As a general rule of thumb, the more ambitious the indicator system is conceptually, the more likely it will run into serious problems of measurability and comparability. In particular, one needs to ascertain whether the desired statements to be derived merely capture the current status in a snapshot perspective or should be extrapolated to outline future conditions in terms of a projection. Classical statistical work concerns the former (Hildebrand 1864: 137), while the modern-day emphasis on sustainability issues provides additional challenges.

Most importantly, one needs to clarify at the outset, whether the desired comparability is inter-temporal or even international in nature. The fact that international comparability is not easy to ascertain was a matter of intense discussion among statisticians of the 19th century. Von Neumann-Spallart (1885: 223), for instance advocates the formation of an international statistical institute to reach more uniformity in statistical concepts and procedures. After all, he goes on to argue, statistics can only fulfill its tasks when it is understood as an international science (1885: 223), especially when the

objective is explaining the consequences of regulations or institutions which vary across, but not within nations (1885: 225). This insight had already been behind the formation of national statistical offices in the new nation states of the 19th century and of the formulation of the rules governing their operation (N.N. 1870: 111). Regarding the governance of such institutions, it seems obvious that these statistical agencies need to be independent of their respective governments (von Neumann-Spallart 1885: 229).

Another trade off concerning the organization of the task regards the recurring nature of the desired report. In particular, the current “GDP and beyond” initiatives are aiming at the construction of indicator systems which are published in regular, not all too distant intervals, which are published timely and not with all too long lags, and whose messages are robust and not altered too severely, if definitions, timing or other practical issues involved in the derivation of their indicators are varied slightly. Timeliness was a concern already 150 years ago. Hildebrand (1864: 136, 137), for instance, judged it to be unacceptable that in his fast-living times, one had to wait for several years until important statistical results were published. The pressure for short lags between the realization of the economic phenomenon and its publication has not become smaller since that time.

A further issue which most of the “GDP and beyond” literature has been silent about (an exception is Kassenböhrer/Schmidt 2011), concerns the trade off between the conceptual desirability of candidate indicators and their measurability. As a general rule of thumb, those items which, in principle, would offer important additional information on human prosperity on top of GDP, are typically difficult to measure precisely. Examples would be individual freedom or social contacts. By contrast, items which are typically measured quite precisely, such as life expectancy, are often highly correlated with GDP and, thus, unable to provide valuable additional information. In fact, the close resemblance between economic prosperity and life expectancy already occupied the literature 150 years ago (Hildebrand 1863b).

Finally, all statistical work involves time and monetary resources. This almost trivial, yet easily forgotten insight has already been a matter of concern for 19th century statisticians (Hildebrand 1864: 136; von Neumann-Spallart 1885: 225). Thus, as one guiding principle of all applied statistical work, one should make use of the generous reservoir of previous conceptual work and already established data collection wherever possible. Only if applied statisticians follow this principle, will they be able to avoid redundant efforts and to trade-off the marginal benefits and costs of collecting additional data or constructing another indicator properly.

Regarding the organization of statistical work, Hildebrand (1863c: 487, 1866: 5, 7-8) argues that the work of statistical offices will always be indispensable, most importantly, since the statistical analysis should cover the nation as a whole, not merely a specific section of it, and frequently needs to rest on official registers. But he also acknowledges the resistance which respondents might develop when being interviewed by government agencies for statistical purposes, for reasons of privacy and fear of oppression, and the limits of the willingness of civil servants in fulfilling statistical reporting duties. These ideas are certainly reflected in the modern consensus that statistical offices, research institutes and universities, and private data providers might work fruitfully together to derive an encompassing portrait of the actual state of affairs.

The second statistical issue to be discussed is the distinction between correlation and causality. While it is most obvious to experienced practitioners and academic economists and statisticians, mere correlations are all too often mistaken as causal relations in the political debate. But one has to be absolutely adamant in maintaining that the indicator

systems which are constructed in any of the various expressions of the current “GDP and beyond” movement are exclusively serving descriptive purposes. That is, their concentration is on positive questions such as “What is the true state of human welfare?” or “What are the connections between material and non-material aspects of quality of life?”. A sensible normative debate “What should be done about the state of affairs?” can only be conducted after these questions have been answered – and it cannot rest its answer on these indicator systems alone.

This was already clear to 19th century economists and statisticians (Hildebrand 1866: 9), and yet it is often forgotten in the discussions in the political and public sphere even today. Knapp (1871: 242-243) explains brilliantly how the conclusions derived from statistical results – here the detection of regular patterns of behavior – might differ substantially, depending on the direction of causality implicitly presumed by their commentators. While the contemporaneous proponents of economics as a natural science identified – in modern words – the similarities in behavior as the outcome of some external driving forces, their critics emphasized that inherent similarities between individuals and contextual effects generated homogeneous behavior instead. From the perspective of modern-day literature on social interactions (Manski 1995), Knapp’s position was highly sensible.

While the potential and the limits of observational analyses to facilitate causal statements are the matter of intense academic discussions today, with randomized controlled trials serving as a hypothetical benchmark, already Hildebrand (1866: 3) indicates that, by contrast to the natural sciences, the social sciences suffer from the impediment that they will typically not be able to resort to experimental evidence. The modern-day evaluation literature cautions that, even if causality might have been established in the analysis, one might be reluctant to extrapolate the result from the situation under scrutiny to a more general setting. This is due to a potential conflict between the internal and the external validity of evaluation studies. Typically, social experiments are characterized by a high internal validity and low external validity. Being unaware of the experimentalist movement of today, reservations about external validity are already present in the discussion of causal issues by Hildebrand (1866: 10).

As a third precondition that has to be fulfilled, both, statistical researchers and the recipients of their work need to find a good grasp on the essential statistical issues of sampling error, hypothesis testing, and measurement error. Statistical analysis is always attempting to reduce complexity by abstracting from the multitude of individual influences on the phenomenon, which are irrelevant for the question at hand. This task necessarily involves the construction of average figures, representing relative frequencies of the condition in question within a sample that represents the population. In that sense, all statistical work is abstracting from anecdotes, that is, individual cases, to derive a summary picture of the state of affairs (Hildebrand 1866: 3-4).

To receive a more detailed portrait of the state of affairs in a population, one might resort to stratification, and construct relative frequencies in the corresponding sub-samples. But, as Hildebrand (1866: 7) stated in his words already some 150 years ago, all statistical work needs to rest on some – in modern words – identification assumptions. That is, in order to characterize the contrasts between the relative frequencies across sub-samples, one has to decide, in the first place, how to distinguish these sub-samples. Statistical analysis can only detect relative frequencies when analysts are deliberately looking for them, that is, it is pursuing to discover the known unknown, not the unknown unknown.

The reported figures in any statistical analysis can never be anything else, but an approximation to the underlying population concept. While it was a comparatively recent development at the time, 19th century statisticians started to understand that probability theory is providing a sound analytical basis for addressing the question whether numerical deviations of the derived average figures have to be assessed as relevant or not (Westergaard 1885: 1). What made the idea somewhat unpalatable to many contemporaneous observers, was obviously the abstraction involved in its underlying thought experiment – which is well-established as the intellectual basis of all statistical work today – of imagining the current calculations (in the “sample”) as being embedded in a multitude of hypothetical calculations of the same kind (Westergaard 1885: 2).

Following this abstraction, it becomes possible to judge whether some movement of average figures over time or some deviation across sub-samples are indeed indicating anything substantial. It all depends on the precision of the estimates involved, and “large” or “negligible” have to be understood in relation to the unavoidable residual uncertainty, which is inherent in any such approximation (Westergaard 1885: 23). Most importantly, given the same phenomenon, large samples tend to produce more precise approximations than small samples (Westergaard 1885: 2). The concept of statistical testing is intimately related to this notion of statistical precision. Its principal idea is to devise a decision rule according to which a deviation is large enough to be assessed as “significant”. Again, *ceteris paribus* precision tends to increase with sample size (Westergaard 1885: 4). Correspondingly, any causal effect itself can only be approximated with more or less precision, not with certainty (Westergaard 1885: 22). And any statement about causality necessitates tailoring the analysis to the situation, instead of following a one-size-fits-all approach (Westergaard 1885: 12).

One of the central insights emerging from the current-day literature on the evaluation of treatment effects is that identification and sampling error are intimately related. On the one hand, it is clear that proper causal analysis is resting on the principle of comparing the comparable, and that homogeneity within a sub-sample can more easily be ascertained, if the sub-samples are defined according to more demanding stratification rules. On the other hand, given that the overall sample size is limited, an extensive stratification geared towards ascertaining within sub-sample homogeneity, corresponding to less-demanding identification assumptions, will lead to smaller and smaller sub-samples, and ultimately to highly imprecise average figures. Thus, there is no escaping from finding the right balance between identification assumptions and potential sampling error. The 19th century statistical literature was well aware of this intricate problem (Westergaard 1885: 22).

As it is a problem inherent in any statistical work, measurement error has always been regarded as a serious obstacle to empirical analysis. By contrast to the multitude of irrelevant influences on the phenomenon under scrutiny, which are fruitfully subsumed under the notion of sampling error, measurement errors tend to bias the results systematically (Westergaard 1885: 23). But as if that was not challenging enough, while condensation always requires averaging across individual observations, the current discussion regarding “GDP and beyond” also involves the idea of summary measures of welfare, as an amalgamation of the indicators representing the multitude of facets of human existence. If one was indeed pursuing this idea, this would require researchers to average even across different types of (quantitative and qualitative) indicators. Potentially, this would lead to an exponentiation of the many measurement problems which tend to plague the analysis even in the most benign circumstances.

Thus, even if statisticians are completely aware of the intricate measurement problems involved in their work, the attempt to construct a single, encompassing welfare indicator might face conceptual obstacles, if indicators are difficult to compare across persons in terms of content (“apples and oranges”). Even more dangerously, it might run into problems of manipulability, if the practical application of the concept is open to interpretation and suggestive interviewing, as might be the case, for instance, with issues of self-assessed well-being. In the face of such problems, the current consensus reached in the “GDP and Beyond” debate seems to be that an enlightened society has to be able to deal with a (limited) wealth of information, if the complex reality is to be captured adequately. Thus, instead of a single, encompassing welfare index, one should construct indicator systems instead. It very much seems that time-travelling 19th century economists and statisticians would have understood.

4 Themes and indicators today – and their precursors

One way to delineate the task of measuring human welfare comprehensively that is chosen by both SSFC (2009) and GCEE/CAE (2010) is to take standard measures of economic performance as a starting point, and to improve upon these standard measures in three directions. Regarding (i) measures of economic performance and material welfare, one should aim at advancing the traditional well-matured measures, such as GDP and the unemployment rate, while retaining their well-known strengths. To address (ii) non-material aspects of human welfare, one should enhance the indicator set regarding various facets of quality of life, respecting the conceptual limits to measurability of emotions and preferences. And to acknowledge that there are important (iii) forward-looking aspects of sustainability, one should construct projections of possible courses for the future state of affairs.

It is the latter set of sustainability indicators that distinguishes the current discussion most visibly from earlier attempts at the construction of encompassing indicator systems. Gauging sustainability typically entails new conceptual challenges as well: Quite importantly, these projections always necessitate an underlying assumption of behavioral stability. They are projections of what could happen under a specific set of circumstances, business as usual, for instance, not forecasts of what is likely to happen. By contrast, forecasts also attempt to consider how economic agents or policy-makers might react to a given tendency, thereby potentially altering the course of events altogether. In addition, and not at all less relevant, many of the issues of concern in these considerations have an important international dimension, precluding a sensible reporting that would be confined to the national level.

Figure 1 documents the dashboard that was developed on request of the Franco-German Ministerial Council by the German Council of Economic Experts and the French Conseil d'Analyse économique to monitor economic performance, quality of life and sustainability (GCEE/CAE 2010). Its three pillars “Economic Performance”, “Quality of Life” and “Sustainability” that together contain a total of 25 indicators closely follow the reasoning discussed above. The following sub-sections demonstrate that at least some of these indicators had precursors already in the statistical work of the late 19th and early 20th century. Documenting this early statistical work leads to impressive evidence that these ideas and concepts left their imprint in the volumes of the *JNS* which were published during this era.

Economic Performance (A)	Quality of Life (B)	Sustainability (C)
GDP per capita	Health: Potential years of life lost	Private sector net fixed capital formation (% of GDP)
GDP per hours worked	Education: Students (ISCED 1-6) aged between 15 and 24 years	R&D investment (% of GDP)
Employment rate (15 - 64 age group)	Personal activities: Employees working on shift work	Cyclically adjusted fiscal balance (% of GDP)
Net national income per capita	Political voice and governance: Voice and Accountability	Fiscal sustainability gap S 2
Final consumption expenditure per capita (including government consumption)	Social connections and relationships: Frequency of time spent with people at sport, culture, communal organization	Total private credit to GDP gap
Distribution measure of net income per consumption unit (income quintile share ratio (S80/S20); internationally harmonized)	Environmental conditions: Urban population exposure to air pollution by particulate matter	Real equity price gap
	Personal and economic insecurity: Not-at-risk-of-poverty rate	Real property price gap
		Level of greenhouse gas emissions
		Greenhouse gas emissions per capita
		Resource productivity (GDP relative to non-renewable Domestic Material Input, DMI)
		Resource consumption (non-renewable Domestic Material Consumption - DMC per capita)
		Biodiversity (preliminary indicator: bird index)

Source: GCEE/CAE 2010: 27.

Figure 1 Dashboard for Monitoring Economic Performance, Quality of Life, and Sustainability

4.1 Economic performance

What can today be regarded as “classic” measures of economic performance, most prominently, the growth rate of (per capita) GDP, are serving a range of practical purposes. They are, first of all, a gauge of the current state of the economy, they indicate whether monetary and fiscal policies have to be set into action, and they serve as the basis of reliable fiscal planning. Of course, as stated in section 2, measuring economic performance is at best an approximation to capturing the state of human welfare. Well-known weaknesses of these measures are twofold. They might suffer from problems of observability, as it will be the case, for instance, for activities in the shadow economy, and from problems of valuation and measurement, a problem which tends to plague measurement in the field of public services. And without doubt, distributional questions have not been addressed sufficiently in national accounting up to now.

Against this backdrop, the SSFC report recommended to (i) measure income or consumption in per capita terms, (ii) emphasize the household over the individual perspective, (iii) analyze also wealth and not only current income, (iv) analyze the distributions

of income, consumption and wealth, and (v) give more emphasis to capturing non-market activities (SSFC 2009: 12-14). After careful evaluations weighing the desirable versus the feasible, the GCEE/CAE report proposed the six indicators that are depicted in Figure 1 to monitor material well-being.

And yet, in the face of the critical undertone of the current discussion, one might easily forget the important achievements which have led to the comfortable situation modern analysts have been handed down by their forefathers. In fact, the contributors to the *JNS* during its early decades obviously would have dreamed of a measure with the qualities of GDP, surely not only to assess the economic performance but also as a tool for approximating the welfare of the population. The development of coherent national accounts and the calculation of an all-encompassing indicator for the performance of the market economy along the lines of the later concept of GDP were indeed already present as a latent ideal in their writings.

Yet for several decades, they still had to content themselves with indirect approaches to estimate the national product and to draw conclusions about the situation and evolution of material welfare from these estimates. The most important basis for such work were the tax statistics of the time as is documented by several contributions from Soetbeer (1879, 1882) in the *JNS* that are concerning the national income of Prussia. Soetbeer does not confine himself to report estimates of the overall national income in Prussia. Instead, he provides further insights to his contemporaries through international comparisons and quite detailed analyses of the income distribution. His comparisons of results for Prussia rested on methodologically comparable, i. e. income tax-based estimates of national income for the Kingdom of Saxony and the United Kingdom, respectively. These comparisons served two distinct purposes: He used, first, the tax statistics from nearby Saxony mainly with the scientific purpose to check the plausibility of his results for Prussia, since it was believed that the compilation of the Saxon income tax statistic was the most accurate and the Saxon revenue authorities were the most reliable of his time (Soetbeer 1882: 235). Second, his comparisons of results for Prussia and the United Kingdom clearly follow the ambition, to provide empirical evidence for political decisions. Therefore, he does not only compare the mean values of per capita income, but also proceeds to comparisons of the respective income distributions (Soetbeer 1882: 238-239). In doing so, he was clearly aware that his data at hand allowed only approximate estimates. But he was likewise firmly convinced that even these limited pieces of statistical evidence were highly important to inform (economic) policy.

At his time, the social question was hotly debated in Prussia. The unification of the Reich in 1871 had led to the abolishment of the remains from the medieval guild system. As Schmoller observed, the guarantees of unprecedented economic freedom led to a substantial improvement in economic performance, but it also created social problems and mounting tensions in the society. His perception, that “our society is threatened to resemble more and more a ladder which grows rapidly at the bottom and at the top, but in which the steps in the middle are more and more breaking-out, allowing a sure foothold only at the very bottom or at the very top” (Schmoller 1873: 11, our translation) was widespread and closely resembles the concerns about growing inequality in many countries that are voiced in contemporary debates about the need to go “beyond GDP”.

For Soetbeer, the concerns of growing inequality first and foremost pointed towards an empirical question that he tried to resolve for the period 1872-1879 in Prussia through analyses of changes in the income distribution (Soetbeer 1879). At the outset, he con-

cedes that based on personal experiences, non-specialist observers could get the impression of a declining income in most parts of the society but ever rising incomes in the richest classes (Soetbeer 1879: 113). He then proceeds with calculations of the income shares of six income classes for the year 1879 and tabulates the yearly evolution in the three highest income classes for the period 1872-1879 (Soetbeer 1879: 114-115). He finds that the income shares of all classes remain close to constant and that the public perception of a growing inequality does not hold up in the light of empirical scrutiny (Soetbeer 1879: 115-118).

Three years later, he confirms this result for the period 1872-1881 and, as a methodological innovation that resembles the income quintile share ratio S80/S20 included in the dashboard proposed in GCEE/CAE (2010), calculates the ratio of the income in the two highest income classes relative to the whole national income as a summary indicator to track the evolution of income inequality over time (Soetbeer 1882: 233-234). Comparing the evolution of the two classes with the highest incomes in Prussia and the United Kingdom, Soetbeer is puzzled by the fact that the introduction of a progressive income tax is hotly debated in Prussia although the data show no significant widening of the social gap, whereas in the UK with a much higher and ever growing income share of the two highest classes, this tax seems to be a non-issue (Soetbeer 1882: 239-240). Soetbeer comes to the conclusion that “there is no worse illusion than the opinion that an artificial complication of big capital accumulations could raise the general welfare and especially better the lot of the working class.” (Soetbeer 1882: 240; our translation).

With respect to one prominent recommendation of the SSFC report regarding economic performance, namely to “consider income and consumption jointly with wealth” (SSFC 2009: 13), the contrast between the desirable and the feasible has barely changed over the centuries. Already Hildebrand (1863c: 479-480) deplores that so little was known about the true value of real estate and the corresponding proceeds, but also about the debt associated with real estate. A comparable frustration is provoked in contemporary endeavors to go “beyond GDP” when politicians with a high preference for social equity have to accept the sobering fact, that a wealth inequality indicator comparable to the income quintile share ratio S80/S20 included in the GCEE/CAE dashboard is not (yet) feasible, due to the lack of reliable wealth information especially for high and highest fortunes.

4.2 Quality of life

Undoubtedly, a comprehensive perspective on quality of life comprises both material and non-material aspects. Facets of material quality of life are therefore indispensable pieces of information when aiming at an encompassing indicator, and one should not exclude material aspects altogether. Quite the contrary, we need supplementary information that augments the “classic” indicators of economic performance and material well-being. But addressing the non-material facets of human existence poses its own conceptual challenges. Most specifically, the aggregation of the various non-material aspects of quality of life confronts a twofold problem of aggregation: Not only do we have to condense the information by aggregating across people, as in any statistical approach, but we also have to find aggregate representations regarding the various facets of quality of life for any individual.

Even more importantly, one might assess the typical impact of deviations in a specific aspect of life, living with and without a physical impairment, say, more or less convincingly, depending on the information content of the data utilized and the econometric

approach chosen to extract this assessment from the data. For this purpose, one would construct contrasts between self-assessed levels of well-being, for instance, for individuals which are observed at two different points in time, some of which change their status (from “without an impairment” to “with an impairment” in the example). But as fruitful as such an approach of using the resulting difference in an outcome measure could be in understanding the typical reaction to misfortune, measuring the *level* of well-being is substantially more complicated: There is simply no normal state of affairs which could allow statisticians to calibrate different answers of self-assessed well-being.

In particular, utilizing surveys on happiness to engage into international comparisons is as tempting as it will necessarily be misleading. Too substantial are the many problems of measurement and observability, since long-term happiness and short-term affects are different, cognitive problems and strategic behavior might distort the answers to the interview questions, and there is a danger of political manipulability. Instead of such a “top-down approach” based on overall measures of self-assessed well-being, GCEE/CAE decided in favor of a “bottom-up approach” that rather collects information on seven different spheres of non-material quality of life. The seven dimensions included in the second pillar of the GCEE/CAE dashboard (Figure 1) follow the respective recommendation in the SSFC report quite closely.

The “capability approach” (Sen 1999) provided the conceptual framework for the choice of these dimensions. The core focus of the capability approach is on what individuals in a given society are able to do, i. e. which so-called “functionings” they can freely choose to promote and achieve. If our time-traveling contributor to the *JNS* from the 19th century could have a look at this second pillar of the dashboard, he would surely not be too surprised, neither by the selected dimensions nor by the chosen indicators. That dimensions like health, education, working conditions and security were already seen as playing a crucial role for the quality of life in a society is well documented for instance in the contributions from Neumann (1872), Schmoller (1873) and Sartorius von Waltershausen (1882a, 1882b).

The rapid evolution of applied statistics as a discipline during the first 50 years of the *JNS* and its expansion to ever more subject areas is by itself a proof that the statisticians of the late 19th and early 20th century already had the clear ambition to go “beyond GDP”, even though the tool of GDP was yet to be introduced. The progress made towards this aim becomes visible through a comparison of Neumann’s (1872) account of the social situation in Prussia that was still plagued with the non-availability of many fundamental statistics (e. g. on wages, p. 284) and had to make do with more or less suitable “proxy variables” (like counts of the numbers of dress- and shoemakers or of physicians in a given region as a proxy for the wealth of the population, p. 283, p. 294). A comprehensive overview of the statistical infrastructure and its latest results is provided by Kollmann (1912).

4.3 Sustainability

The standard conceptualization of sustainability encompasses the three essential dimensions of social, economic and environmental sustainability. Social sustainability addresses issues like fairness in distribution and opportunity, adequate provision of social services, gender equity, and political accountability and participation. Economic sustainability demands that the economic system is able to produce goods and services on a continuing basis under the side conditions that a manageable level of government and

external debt is maintained and extreme imbalances between the different sectors are avoided. Environmental sustainability requires (i) that a stable resource base is maintained, (ii) that the over-exploitation of renewable resource systems or environmental sink functions is avoided, (iii) that non-renewable resources are depleted only to the extent that investment is made in adequate substitutes, (iv) that biodiversity, atmospheric stability and other ecosystem functions are maintained which are ordinarily not classified as economic resources. Only if the social, economic and environmental requirements are satisfied simultaneously, is a sustainable state achieved (GCEE/CAE 2010: 102).

From a methodological perspective, sustainability is a concept regarding long-term developments. Essentially, it requires that we answer the question “whether we can hope to see current levels of well-being at least maintained for future periods or future generations” (GCEE/CAE 2010: 101). Statements regarding future developments can never be deterministic – their construction requires identification assumptions on future paths of behavior. Here we are not talking at all about forecasts, though. The desired projections are “What would happen, if ...” statements instead. These questions are especially difficult to answer in the realm of social sustainability. Consequently, the indicators in the third pillar of the dashboard concentrate on aspects of economic and environmental sustainability (Figure 1).

To answer the question whether economic growth can be assumed to continue in an uninhibited fashion, the first two measures address the maintenance and improvement of the productive capital stock. The following two indicators are both concerned with the budget restrictions of governments, but measure fiscal sustainability with different time horizons (GCEE/CAE 2010: 111-112). Another set of three indicators constitutes a small set of early-warning indicators that could alert policy makers and the general public to the build-up of excessive developments in different spheres of the financial sector. Their construction follows the principle of “cumulative gaps”, i.e. the deviations from the respective trends during a moving time window are summed up. If for any indicator the cumulative gap exceeds a pre-determined threshold, this is interpreted as a signal that a crisis in the financial sector could be looming and counteractions might be warranted (GCEE/CAE 2010: 115-125).

The remaining five indicators of the third pillar in the GCEE/CAE dashboard (Figure 1) cover the environmental domain. The inclusion of ecological indicators in a dashboard that uses nations as the unit of observation generates a serious conceptual problem: Many environmental challenges cannot be treated from a purely national perspective in any meaningful way. Keeping this important caveat and, thus, the necessity of tracking the corresponding international developments closely, GCEE/CAE decided to focus on three dimensions: greenhouse gas emissions, resource productivity and the maintenance of biodiversity. To respect the global budget of 705 gigatons of CO₂ that could presumably be emitted until 2050 without jeopardizing the aim of keeping global warming below 2 °C compared with the pre-industrial level, the CO₂ budget per capita was calculated to be 2.7 tons annually between 2010 and 2050. The fact that Germany emits currently around 10 tons of CO₂ per capita documents the size of the challenge and the urgency to resolve it (GCEE/CAE 2010: 128-132).

To monitor the use of non-renewable resources in the national economy, GCEE/CAE recommend the inclusion of indicators for resource productivity and for resource consumption. Due to a number of remaining methodological problems that limit their precision and international comparability, these two indicators can merely serve as warning signals concerning the qualitative trends of resource use (GCEE/CAE 2010:

132-139). Last but not least, the preservation of biodiversity is also addressed in the GCEE/CAE dashboard. Biodiversity, conceived as the totality of genes, species and ecosystems of a region and all their interactions, can be seen as a form of capital that is essentially needed for food and nutrition security, medical progress, the chemical industry, industrial raw materials, as well as ecosystem services like the absorption of carbon dioxide by oceans and forests (Baumgärtner 2006). However, economic research on biodiversity is still in its infancy and the few biodiversity measures that are also based on economic reasoning are still far from being operational.

The two areas of sustainability that are prominently addressed in the third pillar of the GCEE/CAE dashboard, namely economic and environmental sustainability, have been of concern already to economists and statisticians in the 19th century, as a respectable number of contributions in the early decades of the *JNS* reveal. In the economic realm, the analysis of sovereign debts and debates concerning their longer-term sustainability was already a prominent topic. In the environmental realm, questions of the sustainable consumption of renewable and non-renewable resources were also discussed.

That states face something like an intertemporal budget constraint was a well-known fact of life to the political economists of the 19th century. At their time, naturally, the fiscal sustainability was not put at risk by the implicit liabilities of a welfare state, but instead by the very explicit liabilities that were caused by the high costs of financing a standing army or even by waging war. In Prussia, public expenditures for civilian purposes exceeded the military budget for the first time in 1841 (Gerstfeldt 1883: 47). Although the use of public financial means changed considerably, and for the better, since the founding years of the *JNS*, the problem of excessive debt financing and the resulting burden of a high debt service were a topic already back then as it is now (Gerstfeldt 1883: 48-49; von Kaufmann 1887: 97-99).

The practice of comparing the structure of national budgets, so common today due to international organizations like the OECD and the European Commission, was still in its infancy time during the early decades of the *JNS*, but first steps in this direction were already made. Gerstfeldt, after he finished a meticulous compilation of summary financial statistics for Prussia whose administration did not yet publish these figures on a coherent and regular basis, compared the most relevant ratios to the respective values for France and the United Kingdom (Gerstfeldt 1883: 46). Likewise, von Kaufmann (1887) compares financial indicators for the six European Great Powers of his time (Germany, United Kingdom, France, Austria-Hungary, Italy, and Russia). The fiscal situation in the United States, which received considerable attention by 19th century economists and statisticians due to its rapid progress in many areas, were regularly a topic in the early volumes of the *JNS* (e.g. N.N. 1864b, N.N. 1879).

The most obvious difference to contemporary statistics concerning fiscal sustainability, apart from the patchy data sources in the late 19th century, is the fact that GDP was still lacking as the natural reference figure on which ratios of debt levels and debt burdens could be based. Hence, debt levels and burdens were usually calculated with reference to the total sum of revenues and made comparable across nations via per capita terms. The inadequacy of this method, given that the debt bearing capacity does not depend on the population size of a country but on its productive capacity, is only but one proof for the significant progress that the establishment of coherent national accounts and the concept of GDP provided to the disciplines of economics and economic statistics (see Bos 2011 for a comprehensive account of three centuries of macro-economic statistics).

In the ecologic realm, modern assessments of the importance of ecological sustainability find an early precursor in the statements by Hildebrand that man is merely a part of nature, ultimately acting at the mercy of nature and its unyielding laws (1863a: 19, 143). However, today's ecological concerns about the transgression of planetary boundaries and the destruction of ecosystems, obviously did not yet play a role during the early decades of the *JNS*. Nowadays, they are still on their way from the niche of ecological economics into the centre of the discipline (Leipprand/aus dem Moore 2012). At the time, environmental questions were addressed exclusively from a perspective that saw nature for and foremost as a resource that man can use.

Among the examples for this resource view on nature in the early volumes of the *JNS* are the short analysis on the range of hard coal reserves in England (N. N. 1864: 300-301), the discussion of a map that depicted production, consumption and transport of brown and hard coal in Prussia (Laspeyres 1863: 230-231) as well as treatments on the principles of forestry (Mayr 1864) and their nexus with the than emerging scientific discipline of economics (Eggert 1883). Indeed, the concept of sustainable economic activity was developed in the context of forestry and spelled out explicitly under the denomination of sustainability in the *JNS* by Eggert: "It is the concern over a sufficient use in the future that leads to a well-regulated economic activity in which a sustainable mode of production aims at using only so much in a year or period, as can be compensated for by natural growth." (Eggert 1883: 306; our translation).

5 Conclusions for the current debate

In a year in which the *JNS* completes its first 150 years, it seems worthwhile to look back and search for the intellectual roots of the debate on "GDP and beyond" in these far earlier times. This article documents meticulously that many of the ideas and concepts discussed today were indeed already present in the late 19th century, with the *JNS* being an important marketplace of ideas. This assessment regards both the tough contemporaneous debates on how to advance economic reasoning and the argumentation regarding the potential of statistical methods to support this endeavor. In fact, what most economists and statisticians consider to be the natural perspective on their fields of study today, was ground out by their predecessors many decades ago.

In addition to these remarkable parallels, in this paper we discuss the most important differences that distinguish the current and the past contributions to the issue. Most importantly, the three essential aspects of sustainability, economic, social and environmental sustainability enjoy a lot of emphasis today. As the world community seems to be seriously testing its planetary boundaries at the present time, it is no surprise that these issues were not as important in the late 19th century. In addition, where these early-day economists and statisticians were dreaming of a better statistical information base, current users of statistical reports and indicator systems rather face the danger of drowning in a sea of information. Consequently, now the task is making this information accessible to a large, potentially uneducated public.

We might be able to master this assignment for the very same reasons which have created the nearly overwhelming wealth of information, namely on the basis of advances in information technologies. We can indeed be hopeful that better statistics and better reporting techniques might combine to promote the state of information in the population, prepare better individual decisions and policy-making, and ultimately provide a foundation for democracy. These are the motives behind the many present-day endeavors

for the formulation of combined indicators and comprehensive and at the same time well-structured and compact indicator systems and their website presentations.

This spirit was already present some 150 years ago. Engel, for instance, states “Statistical analysis that should be useful for the present does not only need to present its results as quickly as possible after the facts it describes, but also needs to find the utmost dissemination, since the public is the inspiring and corrective element for statistical analysis.” (cited in Hildebrand 1864: 137, our translation). In conclusion, the recent attempts at constructing encompassing indicator systems reflecting the state of human welfare and societal progress, such as the prominent attempts by SSFC (2009) and GCEE/CAE (2010), are indeed “standing on the proverbial shoulders of giants” (GCEE/CAE 2010: 3). Since its early years, the *JNS* has been an indispensable place for the publications of these great minds of the late 19th century, and now, some 150 years later, economics and statistics owe them a tremendous amount of respect.

References

- ABS – Australian Bureau of Statistics (2010), Measures of Australia’s Progress – Is Life in Australia getting better? Internet: www.abs.gov.au/about/progress.
- Baumgärtner, S. (2006), Measuring the Diversity of What? And for What Purpose? A Conceptual Comparison of Ecological and Economic Biodiversity Indices. Internet: <http://ssrn.com/abstract=894782> (last accessed: 14.11.2012)
- Bos, F. (2011), Three centuries of macro-economic statistics. MPRA Paper No. 35391. Internet: <http://mpra.ub.uni-muenchen.de/35391/> (last accessed: 14.11.2012)
- Bötzow, C. (1883), Die in Europa und den Vereinigten Staaten von Amerika erteilten Patente 1842-1882. *Jahrbücher für Nationalökonomie und Statistik* 41: 571.
- CIW (2011), How are Canadians really doing? Highlights: Canadian Index of Wellbeing 1.0. Canadian Index of Wellbeing and University of Waterloo, Waterloo, Ontario.
- Conrad, J. (1878), Bruno Hildebrand †. *Jahrbücher für Nationalökonomie und Statistik* 30: I-XVI.
- Deutscher Bundestag (2012), Arbeitsbericht der Enquete Projektgruppe 2 “Entwicklung eines ganzheitlichen Wohlstands- bzw. Fortschrittsindicators”. Enquete-Kommission Wachstum, Wohlstand, Lebensqualität, Kommissionsdrucksache 17(26)72 neu, Berlin.
- Drobisch, M. W. (1871), Ueber die Berechnung der Veränderungen der Warenpreise und des Geldwerths. *Jahrbücher für Nationalökonomie und Statistik* 16: 143-155.
- Eckersley, R. (ed.) (1998), *Measuring Progress: Is life getting better?* CSIRO Publishing, Collingwood, Victoria, pp.3-34.
- Eggert, U. (1883), Die ersten Beziehungen zwischen Forstwirtschaft und Volkswirtschaft. *Jahrbücher für Nationalökonomie und Statistik* 40: 305-343.
- European Commission (2009), GDP and beyond – Measuring progress in a changing world. Communication from the Commission to the Council and the European Parliament, COM 2009-433, Brussels.
- European Commission (2011), Sustainable development in the European Union – 2011 monitoring report of the EU sustainable development strategy. Luxembourg.
- GCEE/CAE – German Council of Economic Experts and Conseil d’Analyse économique (2010), Monitoring Economic Performance, Quality of Life and Sustainability: Joint Report as Requested by the Franco-German Ministerial Council. Wiesbaden: Statistisches Bundesamt.
- German Bundestag (2010), Setting up of a Study Commission on “Growth, Wellbeing and Quality of Life – Paths to Sustainable Economic Activity and Social Progress in the Social Market Economy”. Printed paper 17/3853, Berlin.
- Gerstfeldt, P. (1883), Beiträge zur Statistik der Finanzen in Preußen. *Jahrbücher für Nationalökonomie und Statistik* 41: 1-61.
- Hall, J., E. Giovannini, A. Morrone, G. Ranuzzi (2010), A Framework to Measure the Progress of Societies. OECD Statistics Working Paper 2010-05.

- Hildebrand, B. (1863a), Die gegenwärtige Aufgabe der Wissenschaft der Nationalökonomie. *Jahrbücher für Nationalökonomie und Statistik* 1: 5-25, 137-146.
- Hildebrand, B. (1863b), Die neuesten Untersuchungen über die mittlere Lebensdauer. *Jahrbücher für Nationalökonomie und Statistik* 1: 605-631.
- Hildebrand, B. (1863c), Die statistische Aufgabe der landwirtschaftlichen Vereine. *Jahrbücher für Nationalökonomie und Statistik* 1: 478-488.
- Hildebrand, B. (1863d), Vorwort. *Jahrbücher für Nationalökonomie und Statistik* 1: 1-4.
- Hildebrand, B. (1864), Die amtliche Statistik Preussens und ihre neuesten Publicationen. *Jahrbücher für Nationalökonomie und Statistik* 2: 136-154.
- Hildebrand, B. (1866), Die wissenschaftliche Aufgabe der Statistik. *Jahrbücher für Nationalökonomie und Statistik* 6: 1-11.
- Hildebrand, B. (1872), Die Verdienste der Universität Jena um die Fortbildung und das Studium der Staatswissenschaften. *Jahrbücher für Nationalökonomie und Statistik* 18: 1-11.
- Hildebrand, B. (1873), Die Eisenacher Versammlung zur Besprechung der socialen Frage und Schmoller's Eröffnungsrede. *Jahrbücher für Nationalökonomie und Statistik* 20: 1-6.
- Kassenböhmmer, S. C., C. M. Schmidt (2011), Beyond GDP and Back: What is the Value-added by Additional Components of Welfare Measurement? *Ruhr Economic Papers* No. 239.
- Knapp, G. F. (1871), Die neuern Ansichten über Moralstatistik. *Jahrbücher für Nationalökonomie und Statistik* 16: 237-249.
- Kollmann, P. (1912), Die Statistik in Deutschland nach ihrem heutigen Stand. *Jahrbücher für Nationalökonomie und Statistik* 99: 414-423.
- Kroll, C. (2011), Measuring Progress and Well-Being – Achievements and Challenges of a New Global Movement. Friedrich-Ebert-Stiftung, Berlin.
- Laspeyres, E. (1863), Karte über die Production, Consumption und Circulation der mineralischen Brennstoffe in Preussen während des Jahres 1862. *Jahrbücher für Nationalökonomie und Statistik* 1: 230-232.
- Layard, R. (2005), *Happiness – Lessons from a New Science*. New York: Penguin.
- Leipprand, A., N. aus dem Moore (2012), Die natürlichen Grenzen der Ökonomie: Plädoyer für eine ökologische Schuldenbremse. *Leviathan – Berliner Zeitschrift für Sozialwissenschaft* 40(2): 173-201.
- Maddison, A. (2001), *The World Economy: A Millennial Perspective*. Paris: OECD.
- Manski, C. F. (1995), *Identification Problems in the Social Sciences*. Harvard University Press.
- Matheson, J. (2011), National Statistician's Reflections on the National Debate on Measuring National Well-being. London: Office for National Statistics.
- Mayr, G. (1864), Ueber Forstverwaltungsgrundsätze. *Jahrbücher für Nationalökonomie und Statistik* 2: 433-454.
- Meadows, D. H., D. L. Meadows, J. Randers, W. W. Behrens III (1972), *The Limits to Growth*. New York: Universe Books.
- Michalos, A. C., B. Smale, R. Labonté, N. Muharjarine, K. Scott, K. Moore, L. Swystun, B. Holden, H. Bernardin, B. Dunning, P. Graham, M. Guhn, A. M. Gadermann, B. D. Zumbo, A. Morgan, A.-S. Brooker, I. Hyman (2011), *The Canadian Index of Wellbeing*. Technical Report 1.0, Canadian Index of Wellbeing and University of Waterloo, Waterloo, Ontario.
- Müller, J. (1912), Versuch einer Statistik des Vermögens deutscher Städte. *Jahrbücher für Nationalökonomie und Statistik* 98: 344-375.
- N. N. (1864a), Nachhaltigkeit der englischen Steinkohlen-Ausbeute. *Jahrbücher für Nationalökonomie und Statistik* 2: 300-301.
- N. N. (1864b), Die Staatsschulden der Vereinigten Staaten v. Nord-Amerika. *Jahrbücher für Nationalökonomie und Statistik* 2: 301-302.
- N. N. (1870), Erlass, betreffend die Zusammensetzung, Stellung und Geschäftsführung der statistischen Centralcommission im Königreich Preussen. *Jahrbücher für Nationalökonomie und Statistik* 15: 111-113.
- N. N. (1879), Brutto- und Netto-Schuld der Ver. Staaten, ferner Gesamt-Bevölkerung, Schulden und Zinsen pro Kopf. *Jahrbücher für Nationalökonomie und Statistik* 33: 340-341.
- Neumann, Fr. J. (1872), Unsere Kenntnis von den socialen Zuständen um uns. *Jahrbücher für Nationalökonomie und Statistik* 18: 278-341.

- OECD (2007), Istanbul Declaration,
Internet: <http://www.oecd.org/newsroom/38883774.pdf> (last accessed: 14.11.2012)
- Sartorius von Waltershausen, A. (1882a), Arbeitszeit und Normalarbeitstag in den Vereinigten Staaten von Amerika, I. Arbeitsbedingungen in der Vergangenheit und der Gegenwart, im Osten und Westen. *Jahrbücher für Nationalökonomie und Statistik* 38: 461-473.
- Sartorius von Waltershausen, A. (1882b), Arbeitszeit und Normalarbeitstag in den Vereinigten Staaten von Amerika, II. Agitation für die Abkürzung der Arbeitszeit und für den Normalarbeitstag. *Jahrbücher für Nationalökonomie und Statistik* 39: 107-146.
- Schmidt, C.M. (2007), Policy Evaluation and Economic Policy Advice. *AStA – Advances in Statistical Analysis* 91: 379-389.
- Schmoller, G. (1873), Die Eisenacher Versammlung zur Besprechung der socialen Frage und Schmoller's Eröffnungsrede. *Jahrbücher für Nationalökonomie und Statistik* 20: 6-12.
- Sen, A. (1999), *Development as Freedom*. University Press, Oxford.
- Siebert, H. (1975), *Ökonomie der Umwelt: Ein Überblick*. *Jahrbücher für Nationalökonomie und Statistik* 188: 119-151.
- Soetbeer, A. (1879), Umfang und Vertheilung des preussischen Volkseinkommens im Jahre 1879. *Jahrbücher für Nationalökonomie und Statistik* 34: 112-119.
- Soetbeer, A. (1882), Preussisches Volkseinkommen i. J. 1881. *Jahrbücher für Nationalökonomie und Statistik* 39: 230-240.
- SSFC – Stiglitz, J.E., A. Sen, J.-P. Fitoussi (2009), *Report by the Commission on the Measurement of Economic Performance and Social Progress*. Paris.
- Statistisches Bundesamt (2012), *Nachhaltige Entwicklung in Deutschland – Indikatorenbericht 2012*. Wiesbaden.
- von Kaufmann, R. (1887), Finanzstatistisches zu den Schulden der europäischen Großmächte. *Jahrbücher für Nationalökonomie und Statistik* 49: 97-150.
- von Neumann-Spallart, F.X. (1885), *Ein internationales statistisches Institut*. *Jahrbücher für Nationalökonomie und Statistik* 45: 220-237.
- Wahl, S., M. Schulte, E. Butzmann (2010), *Das Wohlstandsquartett – Zur Messung des Wohlstands in Deutschland und anderen früh industrialisierten Ländern*. Denkwerk Zukunft – Stiftung kulturelle Erneuerung, Bonn.
- WCED – United Nations World Commission on Environment and Development (1987). *Our Common Future*, Oxford.
- Westergaard, H. (1885), *Zur Theorie der Statistik*. *Jahrbücher für Nationalökonomie und Statistik* 44: 1-23.
- Dipl.-Vw. Nils aus dem Moore, Rheinisch-Westfälisches Institut für Wirtschaftsforschung, Büro Berlin, Hessische Straße 10, 10115 Berlin, Germany.
nils.ausdemmoore@rwi-essen.de
- Prof. Dr. Christoph M. Schmidt, Rheinisch-Westfälisches Institut für Wirtschaftsforschung, Hohenzollernstraße 1-3, 45128 Essen, Germany, and Ruhr-Universität Bochum, IZA (Bonn), and CEPR (London).
praesident@rwi-essen.de