Introduction to the Skeptical Turn in Evaluation:¹

An Overview

Peter Dahler-Larsen

¹ I have retained part of the title of my inaugural lecture at University of Copenhagen. My work since then has largely been a “Copenhagen product.” I thank my colleagues at the Department of Political Science, its leadership past and present, and its administrative staff for the stimulating and productive atmosphere I have found here. I also wish to thank Thomas Schwandt and Ole Martin Høystad for their intellectual inspiration and friendship over many years, as well as for constructive remarks on earlier versions of this text.
PREFACE

Applying for the conferral of the *dr. scient. pol.* (“upper doctoral”) degree, I hereby submit this overview along with a package of five texts:


I will refer to these texts by title and otherwise follow conventions regarding references.

Each of the above texts speaks for itself and reflects the context in which it was written. Rather than summarizing the works I have submitted, the purpose of this jacket is to convey a sense of continuity and coherence across these contributions together with the progression which allowed me to arrive at an overarching pattern of thinking, which is best termed the *skeptical turn in evaluation*. The purpose of this overview is to account for the network of ideas uniting these publications into a single project. This project, however, must be a dynamic one, since evaluation itself is socio-historically moving target.

PART ONE: FINDING ONE’S FEET IN THE EVALUATION SOCIETY

1.1. An evaluation situation.

The socio-political history of evaluation can be told in many ways. Consider the following observations.

In 1985, which was more or less the same year I first became interested in evaluation, Rossi and Freeman (1985: 19) offered the following definition:

*Evaluation research is the systematic application of social research procedures in assessing the conceptualization and design, implementation, and utility of social intervention programs ...* [It]
involves the use of social research methodologies to judge and to improve the planning, monitoring, effectiveness, and efficiency of health, education, welfare, and other human service programs.

In 2004, an international OECD panel visited Denmark. After a few days in the country, the panel produced a report stating that the single factor most needed in the Danish school system was an evaluation culture. This led to the installation of a number of evaluation mechanisms, including quality reports, individual action plans—and alas, national testing. Within a few years, my society has become an evaluation society.

Now, we are in in medias res in Denmark 2019. Some children in the second grade are reportedly crying before taking national tests in schools. There is much debate about whether the culture of the “performance society” has taken over, and national tests have triggered numerous controversies over the years. Decision has recently been made to evaluate the national testing system; in other words, evaluation has apparently gone so far that only another evaluation is capable of stopping it.

An advisory board has been established to advise the Minister of Education on how to frame evaluation questions and interpret evaluation results. This board consists of teachers, school principals, civil servants, test developers, parent and pupil representatives, and researchers. I find myself being the chairman of the advisory board.

Many viewpoints are represented in the advisory board. Some focus on the technical quality of tests (which is lacking). Some emphasize the (lack of) usefulness of test results for teachers or even the unfortunate consequences of testing for pedagogical processes and for the very meaning of education. Others argue that despite technical deficiencies, test results are still useful for school managers. Some argue that teachers should be better educated to understand the statistical properties of the test results. And others yet claim that teachers already have enough to do. There is no consensus about the use of test results at the individual, class, school, municipal, and national levels. The advisory board produces a lengthy list of potential evaluation questions.

How has evaluation come so far? What thinking about evaluation is responsible and useful?

1.2. The need for a skeptical turn in evaluation

I would not have referred to Rossi and Freeman’s definition, were it not for the fact that it has been the cornerstone of textbooks and courses which were used to educate generations of evaluators, including myself. Evaluators learned that their business is to improve programs. The definition is also
significant because it is symptomatic of a self-understanding of evaluation that could conveniently justify the subsequent widespread diffusion and institutionalization of evaluation. Who can be against improvement?

In the situation described above, what should I tell the many stakeholders in the evaluation of national tests? Should I tell them to put aside their worries, because evaluation is in principle *improvement*, regardless of whether children allegedly cry or not? And that there is no alternative to this view, because it is part of the official definition? Could I do that, honestly and with integrity, already being aware of how much fuss evaluation has created? What should I do with the great discrepancy between evaluation ideology and evaluation practice that I have observed with my own eyes since 1985?

The early definition neither problematizes the term “improve” nor does it problematize what else might happen when social research procedures are applied in society. It does not inspire evaluators to be curious about the social ramifications of the institutionalization of evaluation beyond what is contained in judgment and in improvement. It is blind regarding the socio-political ramifications of its own operations.

Much controversy already takes place within the advisory board described above—even before the evaluation itself has begun. Improvement still remains to be seen.

The Rossi and Freeman definition is now more than 40 years old. Since 1985, the socio-political status of evaluation has developed and continues to do so. And clearly, many other definitions have emerged. I merely used a particularly succinct one for the sake of the argument.

Evaluation is becoming more widespread in time, space, and across policy areas. It is varied in form and shape, many of which are becoming more systematic, mandatory, and institutionalized. It is a policy input and a policy instrument unto itself in an era where the borders between politics, administration, and society are becoming increasingly unclear (Rosanvallon 2009: 50). It has significant political cache, not only in policy-making but also in quality assurance, risk management, performance management, resource allocation, and even promotion and branding. The language of evaluation (e.g. evidence, concepts, indicators, effectiveness, performance, achievement, benchmarking, metrics, etc.) shapes the construction of social understandings of what is of value. Evaluation is thus much more than a form of applied research. Evaluation is a factor of considerable force in society. It is a principle of legitimacy, which demands many instruments and practices (reports, rankings, tests, etc.) regardless of whether the term “evaluation” is used in relation to each
of these instruments. It is in this broader context we can understand why the members of the advisory board have so much to discuss.

The present status of evaluation in society is historically unprecedented. Evaluation and its consequences are moving much faster than our ability to capture and reflect hereon. Just because evaluation promises social betterment, improvement, and learning, and just because social scientists invented it and continue to be involved in it, it should not be protected from the kind of critical reflection that otherwise characterizes how social science examines important contemporary phenomena.

It is therefore time for a skeptical turn. While this particular term might be my own, it is comforting to see how numerous researchers and leading thinkers regarding evaluation are moving in similar directions, including Power, Schwandt, Leeuw, Mathison, Segerholm, Kauko, and Wouters.

To me, a skeptical turn entails a thorough, paradigmatic revision of the status of the key building blocks of evaluation, such as the purposes of evaluation, quality, causality, methodological choices, and consequences of evaluation, but it is more than a conceptual exercise; it is also a socio-existential one. It means reconsidering the mixed experiences with evaluation we already have. Here, the term “turn” is also useful. As evaluation wandered, in the case above, through the OECD, into evaluation culture, and into the classroom of the second-graders, it made some turns. We are now able to see it more clearly from other perspectives, making a new kind of reflection possible. It is the socio-historical and practical turn of evaluation itself, which makes a skeptical turn in the thinking of evaluation possible and timely.

1.3. Evaluation as a political phenomenon
The political is not only found in what is officially defined as “political systems.” The study of the political requires attention to the broad social processes of both politization and depolitization (Warren 1999). Evaluation obviously finds itself amidst these two and plays a political role as it directly and indirectly helps steer and sometimes restructure a number of key domains in modern societies and welfare states; not only education, as seen above, but also social work, health, research, etc. At the same time, evaluation often operates under a depoliticized veil of indicators and metrics, serving apparently unquestionable social goals such as “improvement,” “development,” “learning,” “enlightenment,” “quality assurance,” and “informing the public.”
Of particular interest is the role of ambiguous concepts in relation to these processes (Freeden 2005). The observation of a concept (e.g. quality) that helps transfer an issue related to *res publica* into something naturalized or reified is itself a crucial element in an analysis of the political (Hay 2013). Such analysis should pay attention to categorization and classification processes (Hay 2013: 110) as well as valorization (Julnes 2012; Stark 2009). These processes are central to all evaluative work. But given the complexity of evaluation systems and given the distance in time and place between, say, international evaluation agencies and local professionals (for not to mention citizens), to what extent is it meaningful to say that evaluation helps society function as a democracy? The question is pressing, because the very phenomenon of evaluation was originally conceived of as a way in which social research methodology could be put to the service of a society making rational decisions about policies. Evaluation was to understand itself in the context of collective reasoning characteristic of *homo politicus* (Brown 2015: 221). But does evaluation no longer need democracy? Even further, does it in fact make its own particular contribution to undermining a sense of *demos*, a general process described by Brown (2015)?

Or is it possible that an official evaluation with a broad advisory board is a genuinely democratic attempt, albeit perhaps a desperate one, to catch up with an “evaluation culture” that is otherwise running away?

1.4. Calibrating the theoretical level of ambition

Mjøset (2009) provides a fine account of the multiplicity of research agendas in the social sciences. He describes numerous research frontiers at different levels. In local research frontiers, research contributes to the accumulation of knowledge in relation to pressing social problems and/or middle-range theory, but research findings in these territories do not converge into grand theory. Although I have borrowed inspiration from a number of theoretical sources, constructivism, social philosophy, institutional theory, practice theory, hermeneutic philosophy, and a bit of Actor–Network-Theory (ANT), I do not contribute to these theoretical constructions as such, nor did I plan to do so.

Instead, I am engaged in one of Mjøset’s local research frontiers. The common denominator for the different lines of inquiry I pursue in my publications is how I can best contribute to an understanding of *evaluation* and the role it plays in contemporary society. I do not use evaluation as a case to
illustrate a particular theoretical point. I use various theoretical approaches to understand evaluation itself as socio-political phenomenon.

More specifically, I ask how evaluation is constructed and how it helps to construct *in medias res*. Thus, it is no surprise that my approach is a constructivist one.

**PART TWO: A CONSTRUCTIVIST APPROACH AND THE RESPONSIBILITY OF THE EVALUATION RESEARCHER**

2.1. A constructivist approach

My engagement in the research agenda described above leads me to select, in a fairly eclectic manner, those theoretical ideas that I find most useful for an analysis of evaluation. I have no particular taste for highly abstract theory which has already turned the world into one system or configuration of systems (in the styles of Marx or Luhmann). Rather, I am more attracted to theories that portray human action as situated in particular practical and socio-historical circumstances. I have sympathy for approaches that refrain from excessive claims about the consequences of a social phenomenon until one sees what is going on in practice and in socio-historical time and place (an inclination I find among otherwise diverse intellectuals such as Becker, Schutz, Weick, Latour, Schwandt, and Vattimo).

My approach is constructivist in a fairly broad sense of the word. Let it be said immediately that, as I see it, constructivism is in no way a theory about the fictitious nature of reality (Dahler-Larsen 2015b). The point is not that a choice is necessary between constructivism or reality (Latour 2003). As a social scientist, I leave meta-statements about the ontological status of reality to philosophers. I am concerned with the social construction of what is experienced and relevant as reality in society. This reality is as real as it can get.

The best defining characteristic of constructivism in practical social analysis is simply that there cannot be an appeal to unexplained extra-social factors. There is no “human nature,” no “function,” no “purpose,” and no “laws” with an existence independent of the reality studied and which determines the social order under study (Castoriadis 1997). At this point, I find constructivism to be rigorous: “Underlying explanatory principles,” “driving forces,” “motivations,” “purposes,” and “independent variables” are all constructions and should be studied as such.
Again, this is not an ontological claim; it is merely a fruitful analytical principle. It debunks claims that there is one good reason why the social order is as it is. The constructivist response is: It can always be otherwise. Reasons and justifications are social products as much as that which they are meant to justify. As a corollary, in evaluation, there is no “hook,” “anchor,” or “foundation,” such as methodology or values or purposes outside the realm of social definitions and constructions. The justifications for evaluation and the tools, models, and concepts used by evaluators, as well as the consequences of evaluation, should be studied as meticulously as evaluation studies evaluands, if not more.

This view influences how I define the problem with evaluation as I see it. I wish to understand evaluation as something constructed for particular reasons and particular purposes in particular contexts and which also helps construct particular forms of social reality. In a constructivist perspective, evaluation at any given time and place involves a brew of assumptions, methods, typifications, practicalities, values, and myths. While many of these are taken for granted, there is no overarching guaranteeing principle, such as improvement, development, learning, or truth, which imposes its own reality upon evaluative processes without being results of social construction themselves.

Consistent with a classical motif in social constructivism, I also wish to show that although people’s official motivations and understandings do play a role, they take things for granted which are, analytically speaking, merely the available social constructions. New realities are created through imagination, language, and social interaction that people do not recognize as their own product. In a constructivist perspective, these processes are usually not under the control of any one rational mind that steers an evaluation from beginning to end.

This analytical grip is promising for two reasons. It deprives evaluation of any privileged access to truth or methodological rules or purposes of social betterment beyond time and place. Instead, evaluation always owes its existence to a construction process based on the raw materials available in the situation at hand (ideas, imaginations, data, models). Secondly, a constructivist perspective opens a perspective on all kinds of social consequences of evaluation beyond what is officially promised by evaluators. Evaluation tends to legitimize itself with reference to intended improvement (i.e. before the improvement itself has occurred). Instead, a constructivist perspective unpacks this assumption by asking: What happens? How does evaluation establish a position for itself so that it
may be taken seriously? Who says that what it produces is improvement? How do people find out what constitutes improvement? Could evaluation actually stifle improvement? What else happens?

A social constructivist approach is not without pitfalls. Social constructivists themselves do not always recognize their own constructions. As Hacking (1999) warns, social constructivism can be taken to very different logical conclusions. Just because something is analyzed as a social construction and it could have been otherwise, it does not mean that the analyst is “wiser” than the actors under study who do not “see” the constructions in the same way the analyst does; it does not mean that the construction is “bad” and that there is no guarantee that any “alternative” would be “better.”

Furthermore, as Longino (2002) shows, knowledge that is socially produced is not necessarily “corrupt,” “biased,” or “not rational.” Producers of knowledge are sometimes doing their utmost, and they are subject to a high number of (socially constructed) control and validation mechanisms. Too often, social constructivism has been used to attack the normativity embedded in existing social constructions, while in effect it is merely smuggling another normativity into the picture.

If socially constructed knowledge is all we have, it is not automatically bad in itself. Instead, the construction of constructions should be qualified by considering how we find out what constitutes better constructions (Latour 2003; 2004).

Perhaps the contemporary criticism of knowledge (often invigorated by the sociology of knowledge and STS studies) has “sent us down the wrong path, encouraging us to fight the wrong enemies, and, worst of all, to be considered as friends by the wrong sort of allies because of a little mistake in the definition of its main target” (Latour 2004: 231). There is no particular reason to attack evaluators who wish to confront totalitarian or inefficient governments or otherwise speak truth to power just because evaluators produce “socially constructed knowledge.” We live in times in which we cannot imagine a society without evaluation, and we do not wish to live in such a society (Dahler-Larsen 2015c; The Evaluation Society, 2012; The Skeptical Turn in Evaluation, 2018). Our sense of modernity and modern reflexivity is out of the bottle and cannot be pushed back (Vattimo 2004).

The question becomes what to evaluate, how, how much, and at what price. If evaluation is both constructed and helps to construct something, how can these constructive processes be set in motion in democratic ways given that principles such as truth, progress, development, or learning are constructs, not metaphysical guarantees in themselves?
Reflexivity is called for, but it is particularly demanding given that an “evaluation researcher” is caught in a double engagement in both theoretical and practical matters.

2.2. Theoretical and practical relevance structures: A double engagement

Having a constructivist view means to see all knowledge construction as something that is socially situated. The same view evidently also applies to the knowledge produced by the evaluation researcher (deliberately maintaining the double meaning of this term: a researcher doing evaluation and/or doing research on evaluation).

To unpack this double engagement, it is useful to refer to the concept of relevance structure introduced by Berger and Kellner (1981). Our knowledge and concepts are selected or shaped in particular ways, depending on how we are engaged in the world. With a particular form of engagement (e.g. a particular social role) comes a particular relevance structure that filters the kind of knowledge and thinking relevant to playing that role. Berger and Kellner (1981) follow Weber and make a sharp ideal-typical distinction between the practical and theoretical relevance structures.

The practical relevance structures are linked to tasks, projects, and interests in everyday life. They are embedded in practical contexts, which do not easily lend themselves to systematic scientific study. They take place on rough ground (Schwandt 2003).

In contradistinction, Berger and Kellner explain, a scientific relevance structure “brackets” ideologies and values. Only then can good scientific work be produced. Theory-oriented questions can be pursued and answered only if one serves no other master than pure interest in knowledge within a discipline. The resulting knowledge is abstract, depoliticized, and cumulative. The control of its development is in the hands of academics whose autonomy is guaranteed by free institutions, such as universities.

In that light, one might expect a theoretical account to be logically ordered, impartial, and free of the messiness of the experiences characterizing practical life. Although more complex things can be said about the relation between practical and theoretical engagements (Pielke 2007; Sismondo 2008; Stokes 1997), I shall mention only a few complications which pertain directly to evaluation.

The very field “evaluation” would not exist were it not for the practical engagement of social science research in real-life problems. And the very meaning of evaluation is that it should be used as directly as possible to make social improvements. Social science has itself crept out of the cage of the
“theoretical relevance structure” which Berger and Luckmann (1966) defined for it. When the literature on Mode-II knowledge developed to describe research-based knowledge produced in the social contexts of its application (Gibbons et al. 1994), evaluation was already there. As Stame (2013) correctly observes, the history of evaluation should not only comprise the “great men” (sic!) who invent Evaluation Theory, but also and even more all the “theory weavers” who help make evaluation a part of ongoing social practices.

Against a view stating that, first, robust evaluative knowledge is produced and decision can then be made to use it in practice, Paul Wouters (2018) makes a compelling observation: Evaluative decisions are already everywhere. The evaluative use of data has penetrated modern life to such a degree that it becomes an abstraction to assume that the production of knowledge and the use of knowledge should succumb to two entirely different sets of rules. Evaluation culture, evaluation systems, and the knowledge they both produce are already part and parcel of institutional operations and daily social practices. What actually happens in this domain should be seen as an important source of knowledge about evaluation. This view is reminiscent of the good old phenomenological point that everyday experiences should be taken seriously as the basis for theorizing, which, to be fair, Berger and Kellner (1981) also note.

From what independent, objective, and orderly perspective should the experiences gained from practical engagement with evaluation then be studied theoretically? It is problematic to assume without further notice that evaluation researchers can make independent and impartial studies of evaluation. Evaluation is already defined in terms of assumptions about its social purpose and potential consequences, including accountability, learning, improvement, enlightenment, and providing information to the public (Chelimsky 2006). Is it not futile to expect that evaluation researchers can neutrally position themselves in such a landscape of promises? If you promise on behalf of social science to deliver improvement, for example, is it not a painful, existential challenge to give up on that promise? A study of evaluation on the basis of the self-understanding of evaluation inherited from modernist social science is neither detached in the sense of “having nothing to do with practical life” nor in the sense of “having no interest in the subject.” It would be social science having no interest in social science. It would be parents uninterested in their children. Nobody would believe that set-up. Therefore, something more than merely an empirical study of evaluation carried out with conventional social science paraphernalia is needed.
The problem here is not that I am not interested in an “objective” study of evaluation. The problem is that if a starting point is taken in the self-understanding of evaluation (as applied social science leading to improvement), the starting point is anything but objective. Once social science commits itself to social improvement, it must face the loss of a foundationalist and authoritative voice if it wants to be reflexive. It can no longer claim conceptual control over the description of its own social implications.

Furthermore, to make the tragic paradox come full circle, social science is today itself under such enormous pressure from evaluation systems (bibliometrics, citation analysis etc), that it is becoming increasing mythical to view its position as “autonomous” (Stengers 2018).

A more credible path is the one suggested by social studies of science (Hackett et al. 2008). The increasing self-reflexivity of social science has developed as a logical consequence of studies of science (Hackett et al. 2008; Latour 2005; Longino 2002; Sismondo 2008) and a result of the “sociologization” of philosophical thinking (Vattimo 2004).

2.3. The responsible evaluation researcher

Double engagement in theorizing about evaluation and participating in the practical social life of evaluation is full of friction due to differences in problem-definitions, concepts, languages, motivations, interests, and timeframes.

**Even if the two types of engagement reflect two different kinds of relevance structures on the ideational level, however, the present societal and institutional conditions under which evaluation unfolds as a social construction simply do not allow for a corresponding separation of these two kinds of engagement as if they took place in compartmentalized parts of society.**

Consider my own thrownness into evaluation. In practice, engagements such as research, teaching, consultancy, and public speaking have more overlaps than one would expect if a clear separation of theoretical and practical relevance structures ruled the world. This is particularly pertinent in teaching. I cannot ask my students to be critical in particular circumstances and not in others. It is my duty to let them meet conflicting relevance structures in the same course, in the same lesson, in the same evaluation case. I cannot ask them to be evaluators from nine to five and social researchers in their

---

2 This perspective was deliberately kept out of an early landmark *oeuvre* in a constructivist study of the sociology of knowledge (Berger and Luckmann 1966), but could not, of course, be kept out for long.
free time. I cannot ask my students not to be citizens. Teaching exemplifies an intense coming-together of these various relevance structures.

Consequently, I cannot apologize for the lack of detachment from practical life in my engagement with evaluation. And although I recognize the dangers of practical and political pressures, I cannot merely lament the loss of “independence” and declare all knowledge and insight gained in the practical social arena as “biased.” I argue instead that I take the double engagement in theory and practical life seriously to the best of my ability. In a way I follow Latour, who suggests that when we are critical of a social construction, we should not move away from it, but instead closer to it (Latour 2004: 231).

Reflexive work is necessary. The lived experience wherein optimistic promises of evaluation are confronted with the messiness of practical experiences with evaluation is sometimes a productive predicament that forces reflexivity to occur.

However, a key problem for the academic achievements of the practically engaged evaluation researcher is that the experiences flowing from various practical engagements with evaluation do not present themselves in the form of neat packages fitting into a linear and already-defined academic research agenda on evaluation.

Let me suggest three principles that I see as relevant regarding how an evaluation researcher can handle a double engagement in these two dimensions in a responsible way.

According to the principle of situated responsiveness, the researcher is a member of society and responds to a number of problems which are socio-historically specific and do not present themselves in any particular order. One is simply asked to “do an evaluation,” “advise on an evaluation problem,” or “develop a model” in the practical world, always under specific socio-political circumstances and always under time pressure. Theoretical problems may appear to be a little more neatly organized, but they are also social constructions in the sense that they are shaped very much by shifting intellectual agendas and different institutional challenges, which help channel research money and intellectual curiosities in some directions rather than others. Situated responsiveness means that the evaluation researchers are invited or thrown into dialogical situations, the initial conditions of which are not under the control of the researcher. These situations are uneven in character as evaluation assumes many forms and shapes in many different institutional and political situations. In a small country, there are not enough evaluation researchers to allow for strict specialization, so one must be
prepared to respond in situated ways to a broad array of very different kinds of requests. In such a
country, an engaged evaluation researcher can therefore not afford the kind of thematic or
paradigmatic specialization that is so characteristic of renowned evaluation researchers in the US.

The second principle is *creative translation*. It is up to the evaluation researcher to transform the
uneven experiences gained from societal engagement into academic concepts, theories, and models.
In addition to a change in the level of abstraction, this translation also means a change of potential
audience and a change of language from Danish to English. Evidently, the transformation of
experience into academic publication happens in dialogue with existing theories, with current and
shifting debates, and it is subject to peer review. There is no algorithm explaining how a particular
experience is transformed into an academic publication, except that it involves what is known as
interpretation. Whether it has to do with articulating an analytical viewpoint, developing a conceptual
model, or discussing general trends in the field, an element of creativity is needed in such an
interpretive endeavor.

Third and finally, the principle of *undivided responsibility* maintains that although situation-
specificity and creativity are important, double-engagement does not mean that contradictory
statements can be made in different settings. Instead, a double-engaged researcher should be able to
explain how a specific translation from the practical to the academic (as well as the other way around)
is consistent with what was claimed to be the case in other contexts in which the same researcher was
involved. Admittedly, being situated in practical situations means being faced with different
audiences. It is necessary to take some social constructions for granted if you want to be seen as
credible by your audience when criticizing other social constructions. Here, communication requires
strategic and rhetorical choices. Obviously, the double-engaged researcher learns on a regular
occasion that an earlier standpoint taken in an earlier situation was not well communicated or even
not sufficiently thought-through, but in that case the experience and insight gained should be taken
up in the next moments of situated responsiveness and creative translation. Sources of contradiction
should not be glossed over but used reflexively.

Without that kind of integrity, it will be difficult for others to trust the double-engaged evaluation
researcher, and it would be difficult to trust himself/herself.

The responsible social scientist can commit to the promotion of evaluation (as e.g. understood as a
profession-to-be or a set of business interests) only if accompanied by a sufficient degree of reflection
about the role of evaluation in society (Schwandt 2002; 2017) and a sufficient dose of skepticism
about the many promises, which have been made by evaluation so far but not kept (*The Skeptical Turn in Evaluation*, Dahler-Larsen 2018). This dose would be no smaller than the dose of critical thinking that teachers in the social sciences would normally prescribe in their common practice. The best contribution of the social scientist lies not merely in methods which evaluators think to be the most important thing they can take from social science (Coryn et al. 2016); instead, the responsible social scientist would ask questions about evaluation in society building on whatever is fruitful in the full range of concepts, ideas, and forms of reflexivity circulating in social science today.

Evidently, it should also be recognized that a resulting skeptical view of evaluation is also situated and constructed. It is not a view from nowhere. The skeptical view I articulate would not have been possible without certain literatures I have read, certain teachers I had, and if I had not grown up in a country characterized by a high degree of trust in the public sector, which helps create a climate skeptical of excessive evaluation (Dahler-Larsen and Boodhoo 2019). In that sense, contributions to evaluation research can be “rooted” as social constructions and reflexive if they live up to the principles of situated responsiveness, creative translation, and undivided responsibility.

**PART THREE: THE CONTRIBUTIONS**

**3.1. Putting a constructivist perspective to work**

As a consequence of the double engagement described above, it has been impossible for me to avoid overlaps of some ideas and observations in various forms of publications. My hope is that every idea which re-occurs has been elaborated and twisted, somewhat like a spiral turning back to where it was, but on another level with a new, reflexive twist.

If I have dwelled on the issue of engagements as described above, it is because there is no happy alignment between my trajectory through these various engagements on the one hand and my publications on the other. I see each publication as an experiment. It has been a rough and messy journey, not a linear one. I have accounted for it in a way I think is consistent with situated responsiveness, creative translation, and undivided responsibility; and with the idea that evaluation is both constructed and constructing.

More specifically, a constructivist inquiry into evaluation should include:

a. Inquiries into the construction of evaluation, meaning how evaluation is formed and shaped in particular organizational and socio-historical contexts. These inquiries should include how the
“purposes” of evaluation are constructed, how evaluation enrolls concepts such as quality in its operations, and how key ideas about methodological choices result from what is socially available at the time, but also of the choices made by evaluators. It should be shown how there is no foundationalist principle that determines the form and shape of evaluation which is not also at the same time formed, shaped, made relevant, and constructed as part of the evaluation process.

b. Inquiries into how evaluation helps construct and reconfigure social realities. These inquiries should include a perspective on the consequences of evaluation which extend far beyond the official purposes, such as “improvement.”

c. Integrating thoughts about how these two aspects—the social construction of evaluation and social construction by evaluation—are related (see “the contestability differential” below).

If these inquiries are successful, they will have consequences for the self-understanding of evaluation by shedding new light on key terms in evaluation, including the purposes and use of evaluation, quality, value, causality, and methods. Whereas each of these key terms conventionally function as a guarantee for certain foundations under evaluation, my strategy is to “steal” such guarantees back and show their socially constructed (and thereby questionable) character.

Finally, the agenda described in points a)–c) being thematically open to all aspects of evaluation would not permit all kinds of conclusions. In this sense, constructivism is self-disciplining. For example, there can be no commitment to a “good use” of evaluation which is not a construction, and it cannot be backed up by a foundationalist principle which merely “falls out of the sky.” If we deal with social constructions analytically, we must do so from beginning to end.

In the work submitted here, my ambition has been to carve out a paradigm that is able to do that, whereas there is less emphasis on showing empirical variations of evaluative phenomena (although many empirical examples can be found in the submitted texts). Each of the following seven contributions (made in five publications) highlights an aspect of evaluation as social construction, thereby fleshing out the paradigm.

3.2. Is my work critical?
A “critical” view is ambivalent as our contemporary social systems incorporate various forms of critique in their functioning (Boltanski and Chiapello 2007). Evaluation is itself a form of controlled institutionalized critique. Evaluation has often responded to both external and internal critique by
developing new models, new approaches, and new forms of institutionalization. This year’s critique is often reborn as next year’s evaluation model.

In a constructivist perspective, criticality is relative to existing social constructions. To be critical is to challenge some social constructions while appealing to others. Without any appeal to some constructions that people understand, accept, or take for granted, no critical argument will be meaningful or relevant. For a situated evaluation researcher in the midst of social constructions, exactly which constructions to challenge and in which way is a strategic choice, and sometimes a difficult one. For example, one can challenge how the notion of evidence has been captured by particular methodological views in evaluation or the very notion of evidence itself. The same with the notion of causality. One can expand the ways in which causality can be grasped theoretically and methodologically, or one can challenge the very notion of causality itself. I have made different choices in each of my publications. For this reason, one of my generous reviewers of earlier versions of this text was correct in observing that my different contributions do not have the same degree of radicality in their critical edge.

It would be premature and unnuanced to categorize my work as either “critical” of evaluation or not. Another reason to refrain from using the term is how “critical” is often bluntly interpreted in the public eye as “negative.”

I prefer the term “skeptical.” Here, skepticism signifies not negativity, but critical thinking and a cultivated absence of misplaced enthusiasm. I see *The Skeptical Turn in Evaluation* as a form of culmination (thus far) to the spiraling movements described above. But it took six steps to get to the seventh step where the turn was made.

### 3.3. The contributions

*Contribution 1: The problematization of drivers behind the evaluation wave*

In the official evaluation discourse, evaluation serves alleged purposes such as learning, accountability, enlightenment, and informing the public. They are also alleged purposes of evaluation. They are promises made before evaluation takes place. They are not documented effects. Nor are they part of a theory explaining why many organizations, at a particular moment in time, find that evaluation is the way to realize these great promises.
I read such organizational promises in the light of neo-institutional theory, which explains isomorphism among modern organizations (Meyer 1994; Powell and DiMaggio 1991). In this reading, evaluation reflects rationalized myths in the environments of modern organizations. Although it is a fairly abstract theory, it produces an idea that can be tested. If the social imaginaries characterizing the dominant rationalized myths in modern environments change, then it should be possible to show corresponding changes in forms of evaluation and the mentalities undergirding them.

In “The Evaluation Society” I try to qualify this idea by showing which kinds of ideas in modernity have fostered changing imaginaries in evaluation. I place special emphasis on the transition from “reflexive modernity” to “the audit society” (Power 1996; 1997). I trace this reconfiguration of evaluation and relate it not only to external social forces, but also to ongoing internal discussions in the field of evaluation. The dynamic at play is more than merely a reflection of isomorphism with changing social expectations. Interestingly enough, evaluators have their own “good” arguments for moving in toward something similar to the audit society described by Michael Power.

If the official purposes of evaluation are no longer taken at face value, it is like stealing the idea that holds everything together. In The Evaluation Society, I make the neo-institutional reading of evaluation productive by showing that if evaluation is not driven and directed by its official “purposes” but rather by institutional logics, imaginaries, and myths, then the understanding of consequences of evaluation should change as a corollary.

Contribution 2: The conceptualization of constitutive effects

In understanding the use of evaluation, the advocates of evaluation usually take a starting point in a list of official purposes. Some “unintended” forms of use are then added; and sometimes “misuse.” I have become increasingly uncomfortable with these distinctions for obvious, normative reasons (who knows the goodness of consequences?) as well as for analytical reasons. One root of the analytical problem lies in the distinction between intended and unintended forms of the use of evaluation (Dahler-Larsen 2013). This distinction continues to refer to the official intentions behind evaluation as a starting point. If my Contribution 1 has any merit, however, the official purposes of evaluation cannot be trusted and are therefore neither good predictors of the actual consequences of evaluation nor good normative benchmarks for the same.

Instead, I have carved out the notion of “constitutive effects,” which describes how evaluation helps define the reality it claims to measure. I have suggested mechanisms which help facilitate constitutive
effects (e.g. limited attention, self-fulfilling prophecies, and institutional lock-in, as well as domains in which constitutive effects can be explored in empirical studies) (Dahler-Larsen 2013). In my most recent book, *Quality: From Plato to Performance* (Dahler-Larsen, 2019), I have further described how these consequences are sometimes already built into a network or configuration (which includes infrastructures, computers, and evaluation machineries). Constitutive effects as a concept help move the “constructed” and the “constructing” nature of evaluation closer to each other. The term constitutive effects helps us pay more attention to otherwise unrecognized ways in which evaluation restructures society. It is important to do so as the mechanisms and infrastructures which facilitate constitutive effects often escape democratic deliberation.

I have developed my thinking about constitutive effects over the years and in several publications; in the submitted work, the notion is well represented in *The Evaluation Society* and in *Quality: From Plato to Performance*.³

**Contribution 3.** Pointing out the interdependence between evaluation and an attractive concept: Quality.

Many contemporary debates take place around grand concepts such as sustainability, equality, innovation, and, perhaps above all, quality. I have analyzed quality in depth because of its prominence in evaluation and in wider discourses about public services in contemporary society. Quality is a particularly interesting concept because it lends itself to a form of rhetorical use that weaves normative and descriptive statements together. Whenever quality is normatively loaded, however, it is always to the positive side. Appeals to quality make opposition illogical.

In *Quality: From Plato to Performance*, I demonstrate that there is no eternal essence in quality, which lends itself to measurement. Instead, I show how evaluation machineries and measurement help define quality in different ways. Society’s obsession with a vague concept with ever-positive meanings and the function of evaluative inscriptions are deeply related. Contrary to what is conventionally assumed, the vagueness of the concept of quality is not a problem for the measurement of quality. Instead, because of the vagueness of the concept, quality inscriptions take on a defining

---

³ Further research: A comprehensive research agenda involves empirically mapping the constitutive effects of particular evaluative instruments under different circumstances (my present project on workplace assessments would make a small contribution here). Furthermore, given what recent studies show about multiplicity and ambiguity under multiple orders of worth, it would be interesting to theorize about whether constitutive effects of evaluation are sometimes enhanced not only under conditions of uncertainty and ambiguity, but perhaps even by uncertainty and ambiguity (I am going to cultivate that idea in a book chapter soon).
role, and the constitutive effects of measurement may even be enhanced. Inspired by practice studies and a bit by ANT, I show that conceptual ambiguities and frictions do not disappear but are displaced through infrastructures and networks. They reemerge in dilemmas in practices. The quality concept helps facilitate particular kinds of constructions.

Quality inscriptions “stabilize” particular forms of administration so that constant political renegotiations of values are avoided. Those who control the definition of quality (with built-in value-judgements) also exercise meta-control upon the formation and delivery of public services. Whether they do so “inside” administrative machineries or at great distance in international agencies, many of these constructions escape the kinds of deliberations we normally associate with ideal democracy. This point resonates with Brown (2015), who analyzes a broader wave of neoliberal discourses and practices (not evaluation in particular).

At worst, installing problem-definitions and solutions around the notion of quality works as a functional alternative to having deliberation about the meaning of the values implicated in the issue at hand. Installing quality by means of evaluation machineries can be a way of *practically circumventing* a thorny issue, which is how to determine which values should guide a particular evaluative practice.

Attention to the definition and use of the term quality is therefore an important point on the agenda for research on evaluation and its role in society. More broadly, this contribution shows that if we want to understand the influence of evaluation, it is not enough to look at methods, data, findings, etc. We must also pay attention to how evaluation makes use of popular concepts in its operations.

*Contribution 4. Illustrating the constructed nature of causal models.*

Causality is an immensely important concept since much of evaluation sets out to inquire into causal links between a social policy, program, project, or intervention and its outcomes. Many view “causality” as the most fundamental, ontologically given factor in all evaluation; or perhaps on a good day, an epistemological issue (e.g. when describing how there are approaches to causality such as “variance approaches” and “process approaches”).

Like quality, however, causality is an ambiguous concept, the meaning(s) of which are derived from practical operations. It becomes socially constructed in particular ways in particular forms of evaluation. I am interested in how causality manifests itself concretely as a part of evaluative practice.
To understand how evaluation constructs what becomes visible as causality, I turn to causal models. Causal models are context-dependent (Dahler-Larsen 2001). I have found a good case in one form of evaluation, “theory-based evaluation” or TBE (Coryn et al. 2011), where the construction of causal models is a particularly central activity. These causal models are called program theories. TBE usually views the program theory as a “common representation of the program that all parties can reference” (Donaldson and Lipsey 2006: 65).

With TBE as a framework, I trace the construction of causal models all the way down to the micropolitics of model construction. I show the difference it makes whether different causal theories held by stakeholders are included in the same model or not. I show that with a slightly innovative departure from conventional rules in TBE, evaluators can portray conflicting program theories directly in a program theoretical model, which I argue is an exercise with promising political and democratic implications. See Theory-Based Evaluation Meets Ambiguity (Dahler-Larsen 2017b).

Although this contribution is articulated in TBE vernacular, it shows how evaluators can think of their work in a broader constructed/constructing perspective, using the evaluator’s construction of causal models (program theories) as an example. This point is entirely consistent with the social constructivist view of evaluation as a “brew” which is being prepared in a contingent manner out of the available ingredients. Some new variations in the co-construction of causal models (a departure from foundationalist causal thinking) are also promising in a democratic light.

Along with the analysis of “drivers” of evaluation and of evaluation “use,” the analysis of how “causality” is contingently produced shows how one key component in evaluation after another can be conceived as a result of socially constructive processes.

Contribution 5: Illustrating methodology as choice and deliberately escaping new “foundationalist” principles.

---

4 After writing Theory-Based Evaluation Meets Ambiguity, I realized that further links could be made between the micropolitics of causal modelling in evaluation and institutionalized, socio-historically specific theories, and policy ideas shifting over time. A theory-based evaluator will therefore find that some program theories are easily socially accepted, whereas others are not in fashion. In that sense, a theory-based evaluator reproduces broader social ideas. An article which described this phenomenon could do the same with causal modelling as The Evaluation Society does with the norms, values, and purposes of evaluation. Such an article would also tie Theory-Based Evaluation Meets Ambiguity (even) closer to a constructivist research agenda on evaluation.
When evaluators turn to social science for inspiration, they seek above all *methods*. Methods are understood as tools that help to overcome evaluative problems. The focus on methodology often resonates with a modern rationalist and instrumentalist mentality. With very few expectations (Schwandt 2002; Julnes 2012), there is much stronger emphasis among evaluators (and for that matter in the evaluation field as such) on describing the foundation of evaluations in methodology rather than its foundation in the philosophy of values, although both are logically elements in proper evaluation theory (Shadish, Cook, and Leviton 1992).

I was happy to accept an invitation to contribute a chapter on qualitative methods in evaluation to the new edition of Denzin and Lincoln’s *Handbook on Qualitative Methods*, which has come to be regarded as a very influential book.

In my chapter, I wanted to show the multiplicity of ways in which qualitative methods could be adopted in evaluative situations, depending on what were interpreted to be the most pressing needs of the situation. I wanted to help evaluators realize that, at least to a point, they have a co-responsibility for the construction of the evaluative situation that paves the way for particular evaluative grips on reality.

The notion of *context* is too often interpreted as if it were an almost physical phenomenon, whereas the same context can often demonstrably be read in different ways (Dahler-Larsen and Schwandt 2012), thus opening a door for multiples ways of “constructing” an evaluation. The corollary critical distance to “foundationalist” and “positivist” understandings of methods was not difficult to sell in the context of the handbook. At the same time, however, I found that the earlier versions of the handbook had too often been captured by an ideology that identified qualitative methods with a particular social justice agenda; or more specifically, identity politics defined in very particular ways in a specific contemporary debate in the US. I could not support the liberation of qualitative methods from foundationalism just to tie them in the next moment to a narrow agenda of identity politics.

I sailed between Scylla and Charybdis by maintaining and re-emphasizing a classical link between evaluation and democracy, which must be a self-reflexive one. I thereby maintained a priori principles undergirding evaluation as far from “metaphysics” as possible. I find this strategy beautifully consistent with a theoretical view of evaluation as social construction, but also, hopefully, a constructive and helpful approach when inviting others into the field and the training of evaluators. I consider it responsible not to preach to evaluators-to-be about any single overarching metaphysical principle that all evaluation should comply with, but instead let them know that they have an
important democratic role to play as situated in a midst of complicated and unfinished social constructions of reality; or that which Thomas Schwandt (2002) calls “the rough ground” (see also “the contestability differential”). More specifically, students of evaluation should not read the “critical philosophical stuff” about evaluation, then put that aside, and then do “real evaluation.” They should see their evaluation practice as part of society-making.

In that sense, I see my contributions regarding “causal models” and regarding “qualitative methods” as very much in alignment within the same overall project, although most people would see them as thematically different. Both of these contributions illustrate that once key concepts such as “causality” and “methodology” are deprived of foundationalist and metaphysical overtones, then these key elements in evaluative practice can also be understood as both constructed and constructing.

**Contribution 6. The contestability differential**

“Contestability differential” is a term I have developed in order to theorize about the relation between the constructed and constructing aspects of evaluation. The two must be related in a particular way if evaluation is to function. The contestability differential is the concept I offer as a way of generally grasping that evaluation seeks to change something at the same time as it must take something for granted—without requiring my commitment to any particular principle as the normative and ontological best foundation for evaluation.

My notion of the contestability differential springs out of reading the self-institution of society as a fundamentally open and autonomous activity (Castoriadis 1997), and one that should be seen as no longer guided by metaphysical principles (Vattimo 2004). If tradition had determined everything, evaluation would not have been conceivable. Evaluation thrives on contingency. It contributes to the self-appropriation of a fragile knowledge society (Stehr 1994). Evaluation is accompanied by a belief that some things can be changed and organized differently. Evaluation is contestation in the sense that it brings stuff into the realm of contingency and creates the possibility of subjecting it to human purpose and intention (Hay 2013: 109). Evaluation questions a part of reality.

More specifically, in the moment that something is constituted as an evaluand, its properties and impacts are measured against goals, standards, and the like; and most likely, something is found which can be changed. *Evaluation induces contingency by placing evaluands in a position of potential contestation.*
In order to be effective, however, the evaluation itself must be trustworthy, solid, legitimate, competently made, or anchored in other principles which some find acceptable. We know from institutional theory that institutions define reality through myths, rules, scripts, norms, sanctions, etc. Institutions function most smoothly and effectively when reality is merely held in place as “taken for granted.” Evaluation cannot work if it questions everything, including itself, at the same time. Evaluation must keep some social constructions out of question in order to question other social constructions. A metaphor explains the contestability differential: If you want to turn a screw with a screwdriver, you must be firmly grounded and have a good grip on the tool. Without something firm, say, if you are weightless in space, you will turn around in midair when you turn the screwdriver, while the screw itself will remain fixed in its original position.

If the evaluation is more criticized or more socially fragile than the evaluand, the evaluation will be ineffective. In other words, in order to help reconstruct evaluands, evaluations must be anchored in more solid constructions than their evaluands. This is what I refer to as the contestability differential (Dahler-Larsen 2015c; The Skeptical Turn in Evaluation, Dahler-Larsen 2018). Evaluation thrives on the difference between social constructions that are more or contestable and those that are less contestable. Or it creates this difference in order to make space for itself. If all social constructions were equally solid, evaluation would not exist; or if it did, it would not work.

The contestability differential offers an analytical entry point to empirical analyses: How is a contestability differential established in a particular case? What are the institutional myths, back-ups, self-defenses, and rhetorical moves that are applied in order to defend a particular evaluative system, approach, and set of decisions?5

5 The different ways in which a contestability differential can be built is illustrated in a study I did of how secretaries of Nobel committees for medicine and peace, respectively, regard the “solidity” of their decisions over the years. I found a clear difference between the two. The secretary for the peace committee freely admitted that the annual identification of a peace prize was a political decision, and a very difficult one, and that mistakes are inevitable and will also be made in the future. He therefore did not have to insist that his institution was in any way perfect. On the other hand, he clearly saw the Nobel Peace Prize as an active, political intervention in the world which can make a positive difference. In contrast, the secretary of the medical committee did not see any similar future-oriented use of the prize. Instead, looking backwards, he did not leave any cracks in the description of his organization. He kept justifying all of the laureates chosen over the years, including the controversial prize awarded in 1949 to Moniz for his introduction of the lobotomy in the treatment of psychiatric patients. I invite you to read the study published in the American Journal of Evaluation under the title “Medicine, Peace, Dynamite: The Nobel Art of Evaluation” (Dahler-Larsen 2014). The two very different contestability differentials in that article are easily found, as you compare the history of the evaluative practices behind the two Nobel Prizes.
After I began conceptualizing the contestability differential, I found it theoretically interesting to list factors that I had encountered in my work on either side of the contestability differential, with an open mind to how something can be more or less contested (from the “regulatory pillars of institutions” to the socially “taken for granted”). I assumed that if I found a limited number of phenomena on either side it ought to be possible to synthesize a theory about them.

On the contested side of evaluands, I have found:

- increasing contingency and reflexivity as general sociological trends, including weakening of authority (The Evaluation Society)

- political critiques of public institutions and professionals (Dahler-Larsen 2004a)

- abstraction, quantification, and comparison, which “decomposes” objects into calculable components (Quality: From Plato to Performance)

As regards the strategies that evaluations use to make themselves less contested, I found:

- broad sociohistorical myths about learning, development, accountability, and, later: the prevention of risk (The Evaluation Society)

- aligning evaluation with an ever-positive concept such as quality (Quality: From Plato to Performance)

- technologies such as computer systems which provide evaluative infrastructure (Dahler-Larsen 2004b)

- manpower and expertise (Dahler-Larsen 2004a)

- editing recommendations into something compatible with Zeitgeist so that they will have a greater chance of being implemented (Dahler-Larsen 2004a)

- institutionalization of evaluation (The Evaluation Society)

- moving evaluation closer to management circles (Dahler-Larsen 2015a)

- integrating evaluation into complex infrastructures, networks, and configurations across time and space (Quality: From Plato to Performance).

- mechanisms of competition and financial incentives (Quality: From Plato to Performance; Dahler-Larsen and Kjær Foged 2018).
As I reviewed the phenomena on these lists, I realized that my first hunch about a “limited number of factors” was evidently naïve. In return, however, it is an analytical point in itself to recognize the number and breadth of phenomena in contemporary society which evaluation brings onboard in order to enhance its own operations. The list testifies to the infiltration of evaluation in the sinews of society; a line of thinking suggested in The Evaluation Society and further continued in Quality: From Plato to Performance.

The contestability differential might also be a productive analytical construct in situations where evaluations or evaluative frameworks are not successful. My hypothesis is that this occurs as a result of an insufficient contestability differential under the circumstances (e.g. lack of manpower, lack of resources, lack of credible measurement, lack of evaluation machinery, or a too-strong evaluand). Imagine, for example, what it would require to establish the contestability differential necessary to establish an authoritative center committed to the evaluation of royal families.

The analytical power of the contestability differential can also be gauged by its power to make predictions. I hypothesize, for example, that an evaluation institution which performs self-evaluation will run into difficulty because its evaluative work will appear simultaneously in two ways in two contestability differentials (more contested/less contested). In a case study, I found that an evaluation institution declared that it subjected itself to the same evaluative model that it usually applies to others, which in my terminology is tantamount to having the same model appear on either side of the same contestability differential. I reveal how they solved this difficult logical problem using the simplest of moves: They lied. They did not subject themselves to a model that was nearly as strong as the one they used on others. Consequently, their self-evaluation had little evaluative bite. Their main conclusion was that they needed more money (Dahler-Larsen 2011). While I had not conceptualized the contestability differential at the time, in retrospect it helps explain what was going on in the case.

In terms of publications, The Evaluation Society paved the way for this kind of thinking and informed my contribution to the Denzin and Lincoln handbook. The contestability differential is articulated clearly in Dahler-Larsen (2015c) and summed up in The Skeptical Turn.

The contestability differential recognizes the importance of values, beliefs, and convictions in evaluation but helps maintain critical distance exactly because it does not insist on any particular normativity (see Rosanvallon 2009: 44). Without this spirit, I believe I could not have written, say, my chapter in Denzin and Lincoln’s Handbook of Qualitative Research (Dahler-Larsen 2017a) with
what I regard as a fairly open mind, at least escaping the trap of conflating qualitative methods with a particular, normative agenda. The notion of the contestability differential helped me maintain a cool head.

I consider the conceptualization of the contestability differential normatively “emancipating” in the sense that evaluation can clearly be carried out without such claims as “it would be better if more stakeholders were involved;” or “social justice should be guiding all evaluations,” for not to mention “only evaluation reports printed on recyclable paper should be considered legitimate” as if these were universal principles that would guarantee better and more trustworthy evaluations across time and place.

The concept of the contestability differential has the same status in both the normative and descriptive domains. It does not portray any notion of good use as an overarching purpose which cannot be studied or questioned. It merely considers a functional relation between the factors involved on either side of the contestability differential. It is up to practical and democratic judgment to find out which kind of contestability differential should be considered legitimate enough for a particular item of evaluative work.

I would immediately emphasize “for a particular piece of work.” A democratic society has the right to dismantle a given contestability differential when that differential no longer works well enough, when it costs too many resources, or when it steps on precious plants in the common garden known as res publica.

Research on evaluation has an important mission in pointing out all of the situations in which key parts of a contestability differential struggle to live up to democratic scrutiny. Does evaluation promise too much? Have the interests of the evaluation industry become dominant? Are the constitutive effects uncontrollable and undesirable? Have we constructed a monster?

Such skeptical questions can be asked without any metaphysical backup in terms of “knowing” what the best form of evaluation would be. In sum, research in evaluation can only suggest that contestability differentials be established as soft and preliminary democratic constructions. This proposal resonates with that which Vattimo describes as “weak thinking.” In practice, it is my hope that the contestability differential concept will help prevent runaway effects in evaluation machineries.
In sum, the great advantage of this concept is that it leads to a different set of questions compared to those that often dominate the evaluation discourse. We should no longer be asking questions like, “What is the optimal way of making sure that the institutionalization of evaluation is enhanced?” Instead, we should ask whether, in a particular instance, a particular form of investment in evaluation (beliefs, imaginaries, money, effort) is well-placed in relation to specific issues of common concern. Thinking about such a concern includes respect for the values at stake and attention to the many forms of consequences that an evaluation is likely to have in the situation at hand.

In other words, if the contestability differential is taken seriously, then universal beliefs in evaluation methods and evaluation institutions would be replaced by an approach which returns our concerns to those of subject matters, substances, values, and practical–political choices in the socio-historical situations into which we are thrown.

An interesting question is what would be required of institutional arrangements to embody the principles of the contestability differential in practice.

Contribution 7. The skeptical turn in evaluation

By a skeptical turn I am not merely referring to a change of direction in the literature (similar to “the interpretive turn,” “the practical turn,” etc.); it is also a metaphor for a “turn” of an object that reveals new insights. My argument is that in its widespread and institutionalized form, evaluation has revealed new sides of itself, which in turn requires new reflective work. I describe this in *The Skeptical Turn in Evaluation* (Dahler-Larsen 2018).

This chapter comes out in a book which is part of a series written by members and hang-abouts in the INTEVAL network under the leadership of Ray Rist, a former World Bank official. INTEVAL consists primarily of experienced evaluators from international organizations, who have shaped the evaluation agenda around the world, together with a handful of academics.

Although a long-time member, I have voluntarily moved towards the edge of this network in recent years, where my skeptical views on evaluation are more comfortably situated. Interestingly enough, some of the most experienced INTEVAL members have also become increasingly skeptical over the years. The fact that they invited me to write this chapter for their book is itself a beautiful paradox and an indication that “the skeptical turn in evaluation” may be a term whose time has come.
If democracy is never the implementation of universal principle but instead a way in which society works on itself given a specific socio-historic situation (Rosanvallon 2009), maybe society in the present situation needs to “learn” how to live not only with evaluation but with the specific experiences with evaluation that have been gained over the decades, including the experiences with the institutionalization of evaluation.

_The skeptical turn in evaluation_ is a concept which allows a moderation and re-interpretation regarding the promise of evaluation and its contribution to democracy. Under this headline, it is possible to rethink how evaluation produces the concepts, results, data, and insights that society uses in its own self-description and self-appropriation.

### 3.4. Postscript

The work in the advisory board continues.

### References


