Valuation Studies Vol. 2(1) 2014

Contents:

LiU Electronic Press
http://valuationstudies.liu.se
Valuation Studies logo by Rebecca Elfast

Valuation Studies is a peer reviewed open access journal connecting several vibrant research fields working on the study of valuation as a social practice.

Editors: *Claes-Fredrik Helgesson*, Technology and Social Change, Linköping University, Sweden
Fabian Muniesa, Centre de Sociologie de l'Innovation, Mines ParisTech, France

Advisory board:
Patrik Aspers, Uppsala University, SE
Susi Geiger, University College Dublin, IE
Peter Karnøe, Aalborg University, DK
Hans Kjellberg, Stockholm School of Economics (SSE), SE
Alexandre Mallard, Mines ParisTech, FR
Bill Maurer, UC Irvine, US
Yuval Millo, University of Leicester, UK
Jan Mouritsen, Copenhagen Business School (CBS), DK
Andrea Mennicken, LSE, UK
Ebba Sjögren, SSE, SE
Jesper Strandgaard Pedersen, CBS, DK
Kjell Tryggestad, CBS, DK
Steve Woolgar, University of Oxford, UK

Editorial board:
Kristin Asdal, University of Oslo, NO
Diane-Laure Arjaliès, HEC Paris, FR
Frank Azimont, EMLYON Business School, FR
Alexandra Bidet, Centre National de la Recherche Scientifique, FR
Michel Callon, Mines ParisTech, FR
Alberto Corsin, Spanish National Research Council, ES
Barbara Czarniawska, University of Gothenburg, SE

Editorial board (continued):
Emmanuel Didier, Centre National de la Recherche Scientifique, FR
Liliana Doganova, Mines ParisTech, FR
Marion Fourcade, UC Berkeley, US
Isabelle Huault, Université de Paris Dauphine, FR
Klaus Høyer, University of Copenhagen, DK
Paul Kockelman, Columbia University, US
Michèle Lamont, Harvard University, US
Javier Lezaun, University of Oxford, UK
Celia Lury, University of Warwick, UK
Donald MacKenzie, University of Edinburgh, UK
Liz McFall, The Open University, UK
Peter Miller, LSE, UK
Liz Moor, Goldsmiths, UK
David Stark, Columbia University, US
François Vatin, Université Paris Ouest, FR

Editorial office:
Claes-Fredrik Helgesson, co-editor
Lotta Björklund Larsen, head of office
Karin Thoresson, production & layout editor
Maria Eidenskog, editorial assistant
Josefin Frilund, aide
All at Technology and Social Change, Linköping University, SE

The journal IT-infrastructure is provided and maintained by the *Linköping University Electronic Press* (LiU e-press). The operations of the journal is supported by a grant from the *Swedish Research Council* (VR).

Web: http://valuationstudies.liu.se
E-mail, editors: editors@valuationstudies.liu.se
E-mail, office: office@valuationstudies.liu.se
Twitter: @Val_Studies
Facebook: http://www.facebook.com/ValuationStudies

Valuation Studies 2(1) 2014: 1–4

Editorial note

Valuation Is Work

Claes-Fredrik Helgesson and Fabian Muniesa

The sheer effort of valuation! Valuation is not only a proliferating and multifaceted social practice with important consequences (Helgesson and Muniesa 2013) or indeed sometimes a spectacle put on display and consumed (Muniesa and Helgesson 2013). Valuation may also be work, hard work. We want, in this editorial introduction, to take the opportunity to reflect on the aspects of work and effort in the performance of valuations.

In place of solid empirical study on the efforts of devising and performing valuations, we can make use of our own and others' insight into a few of the many valuation practices that are close to us as academics. The work involved in the peer review process obviously and immediately springs to mind for two newly minted journal editors. Looking around us, we have also other such valuation practices in what is done within and around academic appointment committees, in the assessment of grant applications, in the grading of exams in higher education, and so on. When thinking about it, academia appears to be a line of professional work almost obsessed with putting time and effort into different valuation practices. As we all know from experience, these practices not only have great consequences, and sometimes becomes a spectacle, they regularly entail a labour process.

Let us here focus on the peer review process for academic journals. Sometimes this is discussed as a gate where the percentage passed through is taken as a sign of the quality of the journal (lower percentage is better, we have learned). The editorial process with peer

Claes-Fredrik Helgesson, Department of Thematic Studies—Technology and Social Change, Linköping University, claes-fredrik.helgesson@liu.se

Fabian Muniesa, Centre de Sociologie de l'Innovation, Mines ParisTech, fabian.muniesa@mines-paristech.fr

© 2014 Claes-Fredrik Helgesson and Fabian Muniesa
LiU Electronic Press, DOI 10.3384/vs.2001-5992.14211
http://valuationstudies.liu.se

review has also been discussed in terms of "voting" (Hirschauer 2010) and as a "sieve" (Maurer 2013). Another one would be to think of it as distributed work (Lamont 2009).

In a nice recent article, Neha Vora and Tom Boellstorff (2012) unveils the process by which the manuscript by Vora went through the editorial process (involving Boellstorff as editor) with peer review in *American Anthropologist* and was transformed into the finished article. With quotes from different drafts they highlight the assessments as to what was good and what needed revising, as well as the efforts made to perform these assessments and revisions. In their own words:

> The key point is that scholarly articles are not polished objects that emerge full-formed from the minds of geniuses laboring in isolation. Rather, articles are paths; they are the result of a process of not just research but also the craft of writing. They are the product of a conversation between authors, reviewers, and editors, a collaboration whose details are usually hidden from others and referenced only obliquely in an author's acknowledgments. (Vora and Boellstorff 2012, 583)

The notion of conversation used here is attractive and ties nicely to the notion of published academic work as participating in conversations. Yet, it hides the distributed and on-going efforts depicted in the article. The "revise and resubmit" verdict in editorial work appear here not so much as a deferral of a decision to a later point. Instead it appears as a commitment to keep something in labour. It means keeping it in a process of valuation, and simultaneously in a process of transformation. It is this keeping of the manuscript in a zone of effort and care that is important since it is in this zone of effort where value is accrued.

The peer review process of academic journals as depicted by Vora and Boellstorff actually challenges efforts to clearly distinguish between the processes of assessment (judgements of value) and the processes of production (accruing value) as is suggested by François Vatin in the first issue of *Valuation Studies* (Vatin 2013). In the editorial review process, the efforts that *valorize* and *evaluate* (to use his terms) are highly intermingled or even inseparable. And, indeed, it is the "maybe" verdict of "revise and resubmit" that keeps them intermingled, and, more importantly, keeps the involved parties making an effort. The efforts of judging and accruing value are indeed separated in the rejection decision (even an "accept" means continued effort to polish the piece and to evaluate where that polish is needed). Hence, the production of scholarly output as depicted by Vora and Boellstorff suggests that it thrives on the transformative capacity of the peer review process that goes on under the label of "revise and resubmit."

Moving on to other forms of valuation highly proliferated within the academic world, we could set our eyes on the grading of exams. Anyone who has ever done it immediately recognises that it is a valuation that takes a lot of effort and stamina. Yet, it is, as a rule, not as conversational as is the review process. Or, at least, the one whose exam is under scrutiny is not made part of the conversation going on in the effort of assessing it. Similar forms of separation may be at play in the assessment of grant applications or the assessment of candidates by appointment committees.

Here perhaps it might be more tenable to talk about the separation between the process of assessment and the process of production in line with Vatin's suggestion. Yet, thinking with the notion of valuation as work would clearly be to highlight that assessment amounts to some kind of productive job anyway. Perhaps it is here where an interesting rendezvous can be had between, on the one hand, an interest in valuation as a process involving people reaching compromises on what they like and want, how much and for how much (in a place often referred to as "the market"), and, on the other hand, an interest in valuation as a process involving people doing things for a living inside bureaucratic organisations (that is, in the "the workplace"). An academic journal, however modest, is definitely a nice site to penetrate "the secret laboratory of production" (as Marx would have put it) from this twofold angle. But definitely not the only one and we suggest adding this attentiveness to the labour of valuation to the evolving agenda of valuation studies.

Acknowledgements. We are indebted to Bill Maurer for drawing our attention to the article by Vora and Boellstorff.

References

Helgesson, Claes-Fredrik, and Fabian Muniesa. 2013. For What It's Worth: An Introduction to Valuation Studies. *Valuation Studies* 1 (1): 1–10.

Hirschauer, Stefan. 2010. "Editorial Judgments: A Praxeology of 'Voting' in Peer Review." *Social Studies of Science* 40 (1): 71–103.

Lamont, Michèle. 2009. *How professors think: Inside the curious world of academic judgment*. Cambridge: Harvard University Press.

Maurer, Bill. 2013. "Transacting Ontologies: Kockelman's Sieves and a Bayesian Anthropology." *HAU: Journal of Ethnographic Theory* 3 (3): 63–75.

Muniesa, Fabian, and Claes-Fredrik Helgesson. 2013. "Valuation Studies and the Spectacle of Valuation." *Valuation Studies* 1 (2): 119–123.

Vatin, François. 2013. “Valuation as Evaluating and Valorizing.” V*aluation Studies* 1 (1): 31–50.

Vora, Neha, and Tom Boellstorff. 2012. “Anatomy of an Article: The Peer-Review Process as Method.” *American Anthropologist* 114 (4): 578–583.

Claes-Fredrik Helgesson, professor in Technology and Social Change at Linköping University, and **Fabian Muniesa**, researcher at the Center for the Sociology of Innovation at Mines ParisTech, are the editors of *Valuation Studies*.

Valuation Studies 2(1) 2014: 5–41

A "Democratization" of Markets? Online Consumer Reviews in the Restaurant Industry

Kevin Mellet*, Thomas Beauvisage, Jean-Samuel Beuscart, and Marie Trespeuch

Abstract

This article examines the promise of market democratization conveyed by consumer rating and review websites in the restaurant industry. Based on interviews with website administrators and data from the main French platforms, we show that review websites contribute to the democratization of restaurant criticism, which first started in the 1970s, both by including a greater variety of restaurants in the reviews, and by broadening participation, opening restaurant reviewing to all. However, this twofold democratic ambition conflicts with the need to produce fair and helpful recommendations, leading review websites to seek compromises between these two dimensions.

Key words: market; valuation devices; reviews; online consumer reviews; democratization; restaurant industry

In recent years, the Internet has given rise to many forms of user participation. These user contributions are usually voluntary, and result in the production of freely available public information goods (encyclopedias, video or photo databases, news, expertise, etc.). Consumer rating and review systems such as Amazon's and TripAdvisor's are examples of participatory devices that have met increasing success over the past decade and are now extremely widespread. This way of assessing the quality of products takes various forms, depending on the design of the websites, but it has established a

Kevin Mellet, Orange Labs—Sense / Centre de Sociologie de l'Innovation, Mines ParisTech, France, kevin.mellet@orange.com

© 2014 The authors
LiU Electronic Press, DOI 10.3384/vs.2001-5992.14215
http://valuationstudies.liu.se

central and relatively standardized device: the so-called "online consumer review" (OCR), which comprises a rating and an associated written review. OCR systems allow any Internet users to offer their opinions on a broad range of products and services based on their personal experiences, and to thereby guide consumers' choices. Nowadays, these systems cover a wide variety of goods and services, including cosmetics, hotels and restaurants, appliances, and cameras as well as mechanics, funeral services, banking services, etc. (Beauvisage et al. 2013).

The emergence of a new consumer voice is part of a broader movement towards "empowerment" and "democratization" associated with the Internet. Such, at least, is the explicit discourse of the founders and managers of OCR websites. For instance, the head of Yelp, a consumer reviews website dedicated to local businesses, asserts that "consumers are *empowered* by Yelp and tools like it: before, when they had a bad experience, they didn't have much recourse. They could fume, but often nothing else other than tell their friends. Now the consumer has a lot more power."[1] As for the founder of TripAdvisor—a website that collects ratings and reviews of hotels, restaurants, and tourist sites—it says that "online travel reviews have hugely changed the way the travelers can plan their holidays—they add an independent view of where to go and stay giving another level of assurance that their hard-earned travel Euro is spent wisely. . . . That's the positive power of Internet *democracy* in action."[2]

What do website managers mean when they say that they are contributing to the democratization of the market? What framings do they operate to give materiality to the ideology and rhetoric of democratization? Are they able to give substance to these democratic and consumerist claims in practice? In other words, how much credit should be given to these claims?

The purpose of this article is to describe the production of evaluation by OCR websites and to gauge their claims about being part of a democratization movement. This study focuses on the restaurant industry, which offers a particularly interesting subject with which to complete this agenda. Indeed, it is in this sector, along with the hotel industry, where the effects of consumer reviews, and controversies they generate, are the most important (see Jeaclee and Carter 2011; Luca 2011; Anderson and Magruder 2012; Scott and Orlikowski 2012; Cardon 2014). On websites that specialize in the evaluation of local businesses, such as Dismoiou, Qype, or Yelp, restaurants are the most extensively reviewed category (Nomao 2012). Symmetrically, websites dedicated to reviewing restaurants, such as LaFourchette in France, get high audience ratings. Finally, the

---

[1] Jeremy Stoppelman, CEO of Yelp, in *The Guardian*, 05/06/2013. Emphasis added.

[2] Stephen Kaufer, CEO of TripAdvisor, in *Actualizado*, 20/10/2011. Emphasis added.

extension of consumer reviews has caused a stir among professionals: for example, the introduction of consumer reviews on the French version of the Michelin website in 2012 sparked an outcry from some of the greatest French chefs, who deplored an impoverishment of culinary expertise.

The above-mentioned website managers explicitly refer to a democratization movement associated with the Internet, which they claim to be part of. In his seminal work on the democratizing effect of the Internet, Benkler identifies and differentiates two types of effects: "The first is the abundance and diversity of human expression available to anyone, anywhere, in a way that was not feasible in the mass-mediated environment. The second, and more fundamental, is that anyone can be a publisher, including individuals" (Benkler 2006, 214). The first dimension of Internet democratization stresses access to information and the multiple opportunities that arise as a result, while the second one focuses on the ability of individuals to be active participants in the public sphere. These two dimensions are also emerging from research interested in the integration of democratic ideology into the economy and the market—including the restaurant industry (see Ferguson 1998; Johnston and Baumann 2007). Zukin (2004) describes the history of American consumption as a process of democratization. Her analysis of the formation of shopping as an institutional field shows that technical innovations and lower prices have systematically contributed to expanding access to consumer goods. The work of Gould (1989, 2004) follows another direction, this one grounded in political philosophy. She argues that democratic decision-making should apply not only to politics but also to economic and social life. This approach "bases the requirement for democracy on the equal rights of individuals to participate in decisions concerning frameworks of common activity defined by shared goals" (Gould 2004, 163). To sum up, we can identify two different conceptions of democratization associated with markets and with the Internet. The first one is expressed in terms of access to the market and refers to the "economic" definition of democratization: online review sites allow a much greater number of consumers to benefit from reviews on a wider range of restaurants. We can speak here of democratization-as-inclusion. A second form of democratization, highlighted in the words of TripAdvisor's and Yelp's managers, involves the opening of the public sphere to the expression of new players, namely "ordinary" consumers. Let us call this second form democratization-as-participation.

By confronting these regimes of democratization with respect to the production of lay reviews—that is, examining not only the rhetoric of OCR website managers, but also their concrete achievements—we can estimate the strength of the democratic claims made by OCR websites. In contrast to the economic fiction according to which price summarizes all available information about a product or service, economic sociology has sought to describe the concrete devices that build product quality and guide consumer choice. In the restaurant industry, consumers are likely to rely on gourmet guides (Karpik 2000; Bonnet 2004) and newspaper and magazine reviews (Naulin 2012), but they may also be influenced by word of mouth or by various forms of marketing and advertising. In this context, online review sites constitute *prima facie* an innovative calculative agency (Callon and Muniesa 2005). They especially seem to escape existing descriptions of valuation devices. Thus, compared to the typology proposed by Karpik in *Valuing the Unique* (2010), these sites combine aspects of personal judgment devices (i.e., organizing people's raw expressions) and features from impersonal devices (i.e., building a score and a unique ranking of restaurants). Similarly, in his sociology of reviews,[3] Blank (2007) distinguishes between "connoisseur" reviews—where consumers trust the expertise of a qualified individual, the critic—and "procedural" reviews—where trust is based on impersonal techniques of objectification of qualities, such as the types of technical tests for hi-fi equipment or appliances that are at the centre of consumerist expertise (Mallard 2000). Review sites today do not build up recommendations based on the expertise of a few individuals, nor on rigorous objectification procedures.

As noted by Blank, reviews "are produced by institutions with institutional memory and standard procedures" (2007, 7). These institutions, which he also called "systems," "can be thought of as the short answer to the question 'why is this assessment credible?'" (Blank 2007, 28). This definition invites us to observe in detail how assessments are built, their format, and the type of ranking they produce. It is the construction of review sites as (young) credible institutions that we focus on here. As suggested by Orlikowski and Scott (forthcoming), these valuation devices can be understood as "material-discursive practices," and the outcome of the valuation process depends strongly on the material specifications of the systems. By positioning the OCR websites as extensions of other institutions that equip the market (gourmet guides and food critics in particular), by observing how they are structured and the procedures and formats they put in place, and by noting also the tensions at work between different ways of materially constructing the reviews, we will be able to understand the nature of the agency these sites build, and to define how they produce value. Our perspective complements the work of Scott and Orlikowski (2012), who are interested in understanding how accountability is exercised online on an OCR website (TripAdvisor)

[3] Blank defines reviews as "public summaries and evaluations that assist readers to be more knowledgeable in their choice, understanding, or appreciation of products or performance" (Blank 2007, 7). Blank focuses primarily on reviews produced by professional experts.

and overflows offline in organizational practices. We evaluate the extent to which, and the means by which, the specific accountability exercised by these websites can be qualified as "democratic" or as "more democratic" than other forms of ranking and other accountability practices.

In this article, we build upon three types of empirical materials. First, we investigate the ways in which specialized market intermediaries have built devices dedicated to the aggregation and display of restaurant reviews by ordinary consumers. We list the most important websites involved in customer reviews in the French market: Cityvox, L'Internaute, Dismoiou, Nomao, Yelp, TripAdvisor, and Michelin. Assuming that the construction of OCR devices affects the type of democratization promoted by OCR websites, we first undertook an overview of their features. We reported, for example, if the website suggested particular items to be assessed, such as "food quality" or "value"; if there was a place for free comments; how the average score was calculated; if the website spotlighted intensive reviewers, etc.

Second, we used the results of this systematic exploration to create interview guidelines, which we then used to conduct in-depth interviews with the website managers. These were conducted between July and September 2012 with French managers of five websites: Cityvox, Dismoiou, LaFourchette, Nomao, and Yelp. All the interviews were recorded and transcribed. We asked questions about the history of the websites, how the algorithms were constructed, whether the websites had incentive policies to encourage reviews, how the valuation items were chosen, and whether the restaurants were selected by the website, among other things. We also attended professional meetings and roundtables dedicated to social recommendations on the Internet and analysed most of the professional press and market studies produced by the industry during a two-year period (from mid-2011 through mid-2013). The analysis of these qualitative data was a good start in describing the history, the constraints, and the strategies of the OCR websites dedicated to restaurants. It also allowed us to put the industry's democratic claims into perspective.

Third, we collected extensive empirical material on the evaluation practices in the restaurant industry: on how traditional gastronomic guides and web users assess restaurants, how many and which restaurants are rated, and what scores they receive. We developed dedicated web crawlers to gather this information from online resources: major French OCR websites—general purpose (Cityvox, Qype, TripAdvisor) and restaurant-focused (LaFourchette, L'Internaute)—as well as online versions of traditional gastronomic guides (Michelin, Bottin Gourmand, Gault & Millau). The online version of Michelin's guide publishes both its own (professional) reviews and (ordinary) Internet users' ratings and reviews. Online data were gathered between April 2012 and March 2013; the resulting data set consists of the list of restaurants, their average rating and number of ratings, and the details of the evaluation: score, written review, and contributor identification (Table 1). Based on detailed ratings, we also reconstructed basic information on contributors: number of reviews, ratings, and reviewing history.

Table 1. Basic OCR website data.

Website		Retrieval date	Number of restaurants	Number of ratings
Consumer reviews	Cityvox	04/2012	41,152	208,222
	LaFourchette	07/2012	6,799	642,549
	L'Internaute	07/2012	68,752	480,495
	Qype (restaurants)	01/2012	70,304	88,881
	TripAdvisor (restaurants)	07/2012	32,213	338,722
	Michelin—Internet users	03/2013	18,454	67,679
Professional reviews	Michelin—official reviews	03/2013	4,180	NA
	Bottin Gourmand	03/2013	5,254	NA
	Gault & Millau	03/2013	3,318	NA

The article is structured as follows. The first section depicts the history of gastronomic evaluation before the Internet and identifies two distinct movements of democratization within culinary criticism: inclusion and participation. The second section shows how the development of OCR websites extends these two movements, by achieving inclusion and by systematizing participation. The third section examines in greater detail the functioning of this assessment system. We highlight two features common to all websites: the unequal distribution of contributions, and the homogeneity and high level of average scores. These rating characteristics are constraints that platforms have to deal with, because they are in conflict with the aim of building fair and effective recommendations. The fourth section describes the different types of compromise set up by websites in order to articulate the participation of all users and the effectiveness of the recommendations. This requires either weakening the goal of democratization-as-inclusion or emphasizing a model of participation that favours intensive contributors to the detriment of the principle of equality (of contributors).

Before the Internet: An Effective but Moderate Movement of Democratization

A democratization movement has been affecting the gastronomic segment of the restaurant industry since the 1970s. This movement takes two different, although concomitant, forms, which correspond to the two conceptions of democracy defined above: democratization as market inclusion and democratization as participation. The first dimension entails the expansion of the market by including a growing number of consumers and producers. The second one involves the participation of consumers in evaluating the quality of restaurants and producing hierarchies within that sector. These two movements directly affect the devices, criteria, and procedures used to evaluate the quality of goods and services. Our purpose in this first part is not to recount the history of the restaurant industry as a whole (which is beyond the scope of the article), but rather to examine the transformation of the valuation devices that organize the market. By focusing on their material organization, we try to assess the extent to which their socio-technical organization enacts the two dimensions of the democratization process. Before examining them individually, it is worth recalling briefly the central place occupied historically by food critics and guides in the restaurant industry.

The Pivotal Role of the *Michelin Guide*

Many sociological studies provide evidence of the pivotal role played by food critics, and especially by the *Michelin Guide*, in the emergence, organization, and maintenance of the value of so-called gastronomic restaurants in France (Karpik 2000), the United Kingdom, and Germany (Lane 2013), and to a lesser extent in the United States (Johnston and Baumann 2007; Ferguson 2008).

Gastronomic guides rely on experts, professional food critics, who are responsible for producing judgments of taste and thereby assessing the quality of a restaurant following specific criteria and evaluation standards (Bonnet 2004). Born in the early twentieth century, the *Michelin Guide* (also known as the "Guide Rouge") has specialized in restaurants since 1993, and has introduced two quantitative assessments: a rating from one to three stars, representing the quality of the food, and a five-point ("fork and spoon") rating representing the level of comfort and reception (Karpik 2000).[4] These scores are accompanied by a written comment. The assessment takes a relatively stable form—which is also used by the OCR devices—articulating and combining two opposed operations: commensuration (transforming

[4] The other restaurant guides follow, with, among others, the *Gault & Millau*, rating the restaurants on a scale of 20—5 chef's hats since 2010—and the American equivalent of the *Michelin Guide*, the *Mobil Travel Guide*, founded in 1958 and later known as the *Forbes Travel Guide*, which adopts a rating of 5 (stars).

qualities into quantities) and "singularization" (highlighting the unique and thus immeasurable dimension of the product).

The expert evaluation of Michelin relies on a single scale of quality —directed towards fine dining—with two effects. First, the guide determines and publishes a ranking of gourmet restaurants, and thus helps to guide production by explaining the dominant quality conventions within the market. The *Michelin Guide* has accompanied the transformation of the market, dominated successively by "classical cuisine" (and the figure of the restaurateur), which emerged in the early nineteenth century and experienced its golden age between 1930 and 1960, and by "nouvelle cuisine" (and the figure of the chef) that emerged after May 1968 (Ferguson 1998, Rao et al. 2003). Second, the guide is very selective and thereby contributes to the dualistic structuring of the restaurant industry: on the one hand, the segment of gastronomy, consisting of hundreds to thousands of restaurants; on the other hand, a wide group involving tens of thousands of restaurants, independent or belonging to chains, which are not deemed worthy of evaluation.

To sum up, the *Michelin Guide* marks the incorporation of gastronomy into a fundamentally elitist model, where access to the market is restricted to a limited pool of consumers. The latter have enough economic capital to afford the starred restaurants' bills, as well as the necessary cultural capital to allow the expression of judgments of taste and the purely aesthetic appreciation of the pleasures of the table (Bourdieu 1984). It was in opposition to this elitist model, or at least a detachment from it, that a process of democratization of gastronomy started in the 1970s. The dynamics of this democratization are diverse, but they will always be based on offering printed guides that come as alternatives to Michelin and its emulators.

Democratization as Inclusion: Restaurants and Guides for All

Democratizing the market means, first, expanding it to include more consumers and restaurateurs. This movement is promoted by guides and critics that introduce themselves as alternatives to Michelin. It consists on the one hand of allowing new generations of consumers with limited purchasing power access to restaurants and, on the other, of pluralizing the evaluation of gourmet food to include restaurants that do not refer to the canon of traditional French cuisine.

In the first category, we can find the editorial production of culinary or travel guides, aimed at a penniless audience, which differs from the Michelin readership. The *Guide du Routard*, whose first edition was published in 1972, embodies this alternative. The guide is selective and also focuses on the quality of the food and hospitality, but it introduces budget constraints. In doing so, as explained by its founder Philippe Gloaguen, Routard immediately addressed a generation of broke young travellers. "The idea behind the Routard was to travel cheaply, by meeting people: we were searching for attractive and friendly places rather than the old buildings and museums. We have always kept this in mind. . . . The Routard came at the right time. Charter flights had started a few years earlier (1967). People with little money were travelling with cheap tickets and were looking for affordable hotels and restaurants."[5] Nearly three decades later (1997), the *Michelin Guide* itself would eventually fit into this logic by introducing the "Bib Gourmand" selection, which rewards moderately priced restaurants (Barrère et al. 2010).

Democratization as inclusion is also associated with the "pluralization" of expert judgments. This movement accompanies the evolution of elite cultural practices into a more eclectic and omnivorous consumption. It has the effect of opening market access on the supply side this time. In the field of gourmet food, Johnston and Baumann (2007) draw attention to this movement by observing the emergence of new valuation frames in gourmet food journalism in the United States in the early 2000s. Some magazines, such as *Bon Appétit* and *Food and Wine*, started to criticize the unique value scale that puts French haute cuisine at the top of the culinary hierarchy. They introduced new frames, authenticity and exoticism, pointing at them as legitimate cultural options. Johnston and Baumann interpret this shift as a way of managing the contemporary ideological tension between democracy (inclusive logic) and distinction (exclusive logic). In the context of the United States, where democratic ideology is associated with market culture and consumerism, "democratic ideology fuels the omnivorous notion that arbitrary standards of distinction based on a single, elite French notion of culture are unacceptable, and that multiple immigrant ethnicities and class cuisines possess their own intrinsic value" (Johnston and Baumann 2007, 173). In France, *Le Fooding* guide, launched in 2000, fits these dynamics of pluralizing expert judgments perfectly. It includes in its selection exotic food or even "trendy" tables. Its founder, Alexandre Cammas, was led by "the desire to cross swords with a certain idea of 'unique good taste,' and to open up a more libertarian voice in the world of French gastronomy."[6]

Whether it makes cultural consumption accessible to the masses or renews the repertoires of critics whose reviews are intended for the socio-economic elites, the dynamics of democratization as inclusion maintain the centrality granted to selective guides, and especially to the figure of the expert.

---

[5] "Guides de voyage: 40 après, le routard est une institution," *AFP*, 22/02/2013. http://www.afp.com/fr/professionnels/services/news/838624/sitemap/. Our translation.

[6] "Le Fooding célèbre ses 10 ans," *L'Hôtellerie Restauration*, 16/11/2010. http://www.lhotellerie-restauration.fr/journal/restauration/2010-11/Le-Fooding-celebre-ses-10-ans.htm. Our translation.

Democratization as Participation: Zagat

Things go differently with another dynamic of market democratization, one that requires the active participation of consumers in the valuation of goods. Ordinary consumers are invited to contribute to the making of a collective opinion on the goods and services they have experienced. This lay judgment results from the aggregation of individual judgments of multiple consumers, as opposed to expert judgments, which are assumed to be incomplete, opaque, biased, or even corrupt.[7]

Historically, the participatory logic was not completely absent from guides based on the work of experts. It was mainly based on customer feedback, in the form of letters and e-mails. Such contributions are often encouraged by guides. The Routard rewards feedback from its readers—and they are numerous; according to Lane (2013), Michelin receives 45,000 letters and e-mails in Europe each year. However, even if they are sometimes thanked at the end of the book, the consumers are not presented as contributors or reviewers. Consumer judgments are recognized only to the extent that they rectify omissions or report on developments that have occurred since the visit of the official critic.[8]

Democratization as participation requires, more radically, the disappearance of the expert.[9] The aim is to depart from "expert" judgments by appealing to the judgment of ordinary consumers. This relies on the implementation of specific procedures: voting on the one hand, and the multiplication and aggregation of experience accounts on the other. These actions are at the heart of the editorial project of the Zagat Survey, established in 1979 by Tim and Nina Zagat in New York: "The publishers kept saying that people don't want to hear from people like them, they want to hear from experts. It's sort of amazing when you look back on it now."[10] The Zagat Survey gives voice to ordinary consumers: restaurant rating is entrusted to a group of amateur critics that has continued to expand from 200 in 1979 to more than 30,000 in the early 2000s. The rating of each restaurant is the average of individual ratings given by reviewers. Publishers

[7] As stated by Raymond Postgate, founder of the *Good Food Guide* in the UK in 1951, "you can corrupt one man. You can't bribe an army." http://www.thegoodfoodguide.co.uk/news/the-good-food-guide-is-60.

[8] This is, for example, what Lane observes in her analysis of the *Michelin Guide*: "The Michelin Guide actively encourages responses from diners but still uses inspectors' judgment as the main criterion for the classing of restaurants. The considerable customer feedback . . . is said to provide useful indications but is never considered a substitute for the work by inspectors in the field" (2013, 351).

[9] Even though, according to Bonnet (2004), gourmet critics must be able to put themselves in the *average* consumer's shoes.

[10] Nina Zagat, *New York Times*, 14/11/2010.

compile consumers' written comments and summarize the opinions expressed in the form of quotes.

While some authors emphasize the important innovation that Zagat's format represents (Blank 2007), others have stressed the artificiality of its procedures, which would basically suggest a sham democratization (Shaw 2000; Davis 2009). Shaw denounces multiple breaches of democratic principles and procedures set out by managers. Year after year, the base of evaluators grows, but it is the editors—Tim and Nina Zagat—who produce: i) an average score (out of 30), while the rating of an individual is either 0, 1, 2 or 3; and ii) a survey of verbatim accounts they have received, allowing the editors to prompt consumers to say what they, the editors, want them to. In addition, the questionnaire is sent once a year and respondents do not have to document their entitlement to vote (i.e., by providing receipts). In other words, the democratic procedures fail.

Other guides before Zagat had taken parties to dine to establish the quality of a restaurant on the basis of the collective assessments of consumers rather than just the enlightened judgment of one or two experts. This was the case, for example, in the *Good Food Guide*, first published in 1951. This guide is affiliated with the British Consumers' Association and it produces a ranking of restaurants by combining ordinary consumer reviews (25,000 in 2010) and those of professional experts working anonymously. But, as pointed out by C. Lane, "it is not clear . . . how much weight is accorded to each source of judgment, nor how many inspectors are employed" (Lane 2013, 356).

The participatory process driven by consumerist associations or entrepreneurs seeks to allow lay judgments to emerge. However, printed guides do not quite make the role of the expert—or at least the active mediation of the publisher—disappear. Indeed, by multiplying judgments on the same restaurant, the question of their synthesis appears: with Zagat, it remains the output of an editorial work; OCR websites will make a difference by giving this synthesis digital and algorithmic foundations.

Online Consumer Reviews: The Second Phase of Democratization

The History and Positioning of Web-Based Platforms

We focus now on the emergence and dissemination of websites that collect and publish consumer reviews in the restaurant industry. This movement began in the late 1990s, continued during the next decade (Table 2), and led to the establishment and spread of the standardized form of "rating + written review" that is now found in many retail sectors (Beauvisage et al. 2013).

Although Amazon was a pioneer in the field of online customer feedback—the online bookstore began collecting and publishing reviews in 1996, two years after its launch—the online assessment of restaurants is also relatively old. It was initially driven by "first generation" city guides, such as CitySearch in the United States (started in 1995, but ratings and reviews start from 2000), its French equivalent Cityvox (1999), as well as the entertainment and events section of web portal L'Internaute (2000). At the same time, in France, several specialized guides in the form of web pages published by enthusiastic amateurs appeared. These initially relied on a limited pool of testers and critics and then gradually expanded to "all users," in the image of Restoaparis.fr (1999) and Resto.fr (2000).

Table 2. OCR in the restaurant industry—historical marks.

	USA	**France**
City guides	CitySearch (1995, *rating* + *reviews* from 2000)	Cityvox (1999) L'Internaute (2000)
Specialized websites	Zagat online (1999; acquired by Google in 2011, content becomes freely available)	Restoaparis.fr (1999) Resto.fr (2000) iTaste (2008) Michelin (2012)
Booking services	Opentable (1998)	LaFourchette (2007)
Online urban guides (2nd generation)	TripAdvisor (2000) Yelp (2004)	TripAdvisor (2005) Dismoiou (2007–08) iTaste (2007) Qype (2008), bought by Yelp in 2012 Yelp (2010)
Search engines, aggregators, directories	Google Maps and Google Places then Google+ Local (2005, 2006)—includes Zagat reviews (2011) YellowPages—rating and reviews from CitySearch since 2010.	Google Maps and Google Places then Google+ Local (2005, 2006) Nomao (2007–08) PagesJaunes (rating and reviews from 2010)

With the wave of Web 2.0 innovations, a new generation of websites dedicated to local searches has emerged. Yelp was launched in the San Francisco Bay Area in 2004; it took the place left vacant on the web by

Zagat[11] and differs from the first generation of city guides in that it provides more space for contributors. The site is a local directory and each "place" can receive more or less detailed information, a rating out of 5, and reviews. The site focuses on the community of contributors by developing social features, including member pages that give members visibility and allow them to connect to one another, reward systems (badges) for the most prolific contributors, and dedicated communication tools for members. The site expanded internationally after 2009, arriving in France in 2010. The European equivalent of Yelp, Qype (contraction of "Quality" and "Hype") was launched in 2006 in Germany and arrived in France in 2008.[12] Dismoiou (2007), iTaste (2007), TripAdvisor (2005 for the French version), and the booking site LaFourchette (2007) all appeared at around the same time and offer similar design and features.

Some "historical" players, finally, converted later to online consumer reviews. This is the case of the directory Pagesjaunes.fr, which, since 2010, has allowed users to submit ratings and reviews of listed professionals, with the exception of regulated trades such as doctors and lawyers. In the field of gastronomic guides, the *Michelin Guide*, which is facing a decline in sales of its printed version,[13] launched a free online version in 2012. The latter integrates two major innovations: the inclusion of restaurants not listed in the paper guide, and the possibility for Internet users to rate and review restaurants.

In sum, many websites invite ordinary consumers to evaluate restaurants. The sites differ from each other, of course, because of their particular histories, design, and positioning. However, we observe that, in general, the Internet continues and deepens both democratization movements identified previously: economic democratization and an inclusive logic on the one hand, participatory logic and the empowerment of ordinary consumers on the other. These two dimensions are now examined in turn.

---

[11] The guide, published annually in print, went online in 1999, thanks to the acquisition of the family-owned company by the investor General Atlantic. Access to content is limited, for fear of cannibalizing print sales. As a logical consequence, the website is poorly referenced and gets low audience scores. This model is an economic failure—although the print version remains very profitable—and leaves room for other free sites, starting with Yelp. In September 2011, Zagat was acquired by Google—who had offered $500M in 2009 to buy Yelp, without success. From that moment, the content became free and was made available on Google Maps and Google+ Local services.

[12] Qype is close in design to its American equivalent, Yelp. The latter acquired Qype in October 2012 for an estimated $50M.

[13] The drop in sales is massive: in French bookstores, sales of the *Michelin Guide* have fallen from 500,000 units in 1996 to 107,000 in 2010 ("Tout juste lancé, le site Michelin devra convaincre internautes et restaurateurs," *Le Monde* 06/03/2012).

Restaurant Reviews: From Selectivity to Inclusion

Online consumer reviews represent an important extension of the field of taste-making. While printed guides are based on the critical work of experts and operate on a highly selective logic, websites open up reviewing to a very wide range of restaurants. Printed guides such as *Le Routard* and the *Michelin Guide* produced a dual world, clearly separating a small collection of restaurants that meet their specific quality criteria from the vast array of other restaurants, which were excluded and ignored by evaluation. In contrast, OCR websites open up a world of lay reviews, where each platform integrates a large number of restaurants that fall under varied quality scales: on TripAdvisor, for example, three-star (Michelin) restaurants are reviewed side by side with kebab houses.

It is possible to quantify the magnitude of this movement of inclusion. Considering the number of listed restaurants, we observe that the official guides are typically ten times smaller than OCR websites (Table 3).

Table 3. Number of listed and rated restaurants on a selection of websites—online version of printed guides and OCR websites.

Category	Guide / Website	Listed items	Rated items
Printed guides (online version)	Michelin—experts	4,180	4,180 – 100%
	Bottin Gourmand	5,254	3,860 – 73%
	Gault & Millau	3,318	3,318 – 100%
OCR Website (lay reviews)	Cityvox	41,152	25,223 – 61%
	LaFourchette	6,799	4,200 – 62%
	L'Internaute	68,752	61,782 – 90%
	Michelin—consumers	18,454	14,260 – 77%
	Qype	70,304	29,672 – 42%
	TripAdvisor	32,213	31,999 – 99%

The 2012 Michelin Guide includes expert evaluations for 4,180 restaurants, both in its printed and online versions. On the website, however, more than 18,000 restaurants were reviewed by Internet users.[14] Meanwhile, TripAdvisor listed more than 32,000 French restaurants, most of which had received at least one user review. On the online portal L'Internaute, it is possible to scroll through more than

[14] This measurement was taken in March, 2013, exactly one year after the launch of the website.

68,000 restaurant pages, 9 out of 10 of which are subject to at least one rating. LaFourchette is the only exception to this inclusive logic: this selective website only lists restaurants that use its booking service and get an average rating higher than 6.5 out of 10. "Only" 6,800 restaurants were listed, of which 4,200 were rated, which is similar in number to the printed guides.

One can interpret the extension of the field of evaluation as an achievement of the democratization-as-inclusion movement born in the mid-1970s: all restaurants, even the smallest, are subject to review. Moreover, even the most poorly rated restaurants are listed online, whereas printed guides listed only the best. Thus, online guides not only point at the best restaurants, but also at the worst. This becomes apparent when examining the price ranges of the restaurants listed by each guide (Table 4).

Table 4. Price range and distribution of restaurants.*

	Guide	– €15	€15–€30	€30–€60	€60–€90	€90+	*Total*
Paper	Michelin—experts	0%	10%	71%	14%	5%	*100%*
	Gault & Millau	1%	14%	59%	19%	7%	*100%*
	Bottin Gourmand	3%	32%	52%	10%	4%	*100%*
OCR	Michelin—consumers	4%	44%	45%	5%	2%	*100%*
	TripAdvisor	14%	51%	23%	10%	2%	*100%*
	L'Internaute	16%	76%	7%	1%	0%	*100%*

* Price ranges are based on a predetermined classification for L'Internaute, and with an average of the price brackets for other guides and websites, when such information is available. The filling rate of this information is 21% for TripAdvisor, 80% for the Bottin Gourmand, 85% for L'Internaute, 89% for Gault & Millau, and 95% for Michelin.

Traditional expert guides studied here are focused on the high-end range of restaurants. In terms of price, they are focused on restaurants in the €30–€60 range (between 50% and 70% of listed establishments), and to a lesser extent on the €60–€90 range (between 10% and 20%). Inexpensive restaurants (less than €15) are virtually absent from traditional guides. In contrast, almost 15% of establishments listed on L'Internaute and TripAdvisor belong to this category. The majority of listed restaurants belong to the €15–€30 range. As a symbol of this economic democratization movement, at Michelin there is a division of labour between the professional inspectors, who focus on more than 4,000 restaurants in middle and high price ranges, and Internet users, who review more than 18,000 restaurants in the lower-priced segment.

Power to the Internet Users

Along with this movement of economic democratization, websites continue and deepen the participatory process that was initiated by printed guides such as the *Good Food Guide* and *Zagat*. The movement is progressive, going through a complete outsourcing of editorial work and the stabilization of a judgment device supposed to reflect the collective opinion of consumers about restaurants. The review sites have not so much invented the participation of Internet users in making judgments about goods and services (such opinions were already being expressed in online forums, for example), but an algorithmic form to synthesize them. This synthesis relies on a digital objectification of quality: the score given by an individual is not only the synthesis of his or her opinion, but also the implementation tool of democratization-as-participation, through the algorithm and calculation.

The recent history of websites dedicated to consumer restaurant reviews is marked by a progressive disengagement of the editorial and content production roles. This is well documented in the historical analysis of Cityvox by Weygand (2009). Launched in France in 1999, the site relied immediately on the contributions of users to enrich its content. This did not prevent the young start-up from recruiting, in June 2000, 75 employees assigned to the editorial team to provide content for the various local versions of the site. The company had around 120 employees at the end of 2001, before downsizing its staff to about 15 employees later in the decade. In addition, control over consumer reviews has remained strong, as each submitted review is always moderated by an employee before being posted, and contributors see their reviews published only after they have written three. This dissuasive logic reflects moderate distrust in the voice of Internet users, who must remain strictly controlled by the site's editorial team.

With the emergence of the second generation of city guides, there has been a disengagement of all editorial roles. Websites such as Dismoiou, Qype, and Yelp buy local business directories, which they publish online. Information about the listed businesses, which may initially be scant and sometimes even incorrect, can be corrected and enriched by the traders themselves; the websites encourage them to do so by arguing that richer information provides better rankings, and thus greater visibility, in search engine results. It is also possible to add places that had not yet been listed. Users can then, once registered, leave ratings and written reviews about the businesses, which can be published immediately. Moderation work is delegated to the algorithms responsible for identifying fraudulent behaviour—such as many reviews published at the same time from a single IP address or for a single business, for example—and also to users who may flag suspicious published reviews. Internal teams, meanwhile, focus their activity on the technical management of the platform, the running of the "community" of contributors, and the sale of advertising space.

Disengagement of the editorial work involves the enrolment of users. Fulfilling the movement of democratization-as-participation also requires the establishment of democratic assessment procedures. Websites have relied on existing "rating + review" systems to build collective opinion, by recording, aggregating, and publishing ratings and reviews produced by online contributors. In the restaurant industry, the device imposed itself on all players, even lately to market intermediaries that are not directly involved with lay assessment (such as the PagesJaunes website and especially the Michelin website). The combination of a rating and a written review follows the format established by the *Michelin Guide* in 1933. The first difference with the printed version is that the rating of the restaurant is made from the average of all ratings given by individual contributors. The analogy with the vote is enhanced by a systematic display, next to the rating, of the number of individual scores assigned to the restaurant: "voter turnout" reflects the quality of democracy-in-action.[15] The second difference is the way the written review contributes to the making of an objective judgment on the quality of the restaurant. In the case of the *Michelin Guide*, the objectivity of the review is guaranteed by its anonymity and its (supposed) conformity to strict and codified assessment procedures (Blank, 2007; Ferguson, 2008). In contrast, lay consumers can unleash their subjectivity in their written reviews. Indeed, it is the accumulation and proliferation of subjective narratives, which the user is free to browse, that guarantees the formation of an objective judgment.

This setting is common to all OCR platforms. However, it may be amended at the margin, in various ways, by site managers. Thus, on Cityvox, ratings published more than two years before are not included in the average rating—it is one year on LaFourchette. On Dismoiou, an elaborate algorithm seeks to moderate the effect of extreme ratings:

---

[15] The unequal distribution of ratings distorts the evaluation of different restaurants. Indirectly, it provides a second assessment metric, based on popularity. Website managers believe it is used as an indicator of second ranking, which provides information on the reliability of the rating. Similarly, Luca (2011) shows in an econometric study that the positive impact of a rating on the income of a restaurant is even stronger when this rating is made of a large number of individual ratings. The platforms are not immune to the "Zagat effect" criticized by Shaw (2000), according to whom the supposedly democratic rating is primarily a proof of popularity.

> We consider that negative and isolated ratings are not necessarily . . . well, when there are few ratings, there is an algorithm that brings up the average rating a little bit, we hope in the smartest way possible, but the goal is not explicitly to raise the rating, the goal is to try to mitigate the risk of competitors or people giving a very poor appraisal when it is an isolated case. When assessments are numerous, the average is quickly built, but when there are few assessments, we have something in the recommendation algorithm that will minimize extremes. But it is also true for very good ratings; if there is only one, it will not appear. . . . Basically, the algorithm flattens the extremes when there are few ratings, be they good or bad. (Dismoiou)

The publication of written reviews may also vary from one site to another. By convention, the most recent reviews are always displayed first; but more or less sophisticated filters allow users to select reviews or browse them according to different criteria: language, contributor's status, associated rating, consumption context, etc. The main issue is to ensure that reviews result from actual consumption experiences and are therefore genuine—and this is the issue that was the focus of discussions in the AFNOR working group, aimed at producing a certification of online consumer reviews in 2012 and 2013. Despite websites' efforts to ensure the authenticity of ratings and reviews, deceptive reviews akin to misleading advertising or denigration slip through the net—in proportions that are impossible to assess. For example, on Yelp, 16% of reviews are identified by the filtering algorithm as fraudulent; the proportion of fraudulent reviews actually published, however, is unknown (Luca and Zervas 2013).

Taste-Making Through Public Opinion

By broadening the array of reviewed restaurants, OCR websites have achieved the democratization as inclusion in the restaurant industry; in addition, they have developed a model of consumer contribution to this evaluation. Now we examine in greater detail the operation of this model, and characterize the nature of the rankings produced by the device. If OCR devices are indeed produced by "anonymous and distributed consumers using informal, variable and individual criteria grounded in personal opinions and experiences" (Orlikowski and Scott, forthcoming, 1), they do not produce very heterogeneous and unstable valuations.

In this section, when data processing involves contributors (not restaurants), we had to partially restrict the scope of analysis because it was impossible in some cases to connect a review to a single contributor.[16] Thus, analyses of contributors (in particular, assessment

---

[16] Indeed, among the six platforms studied, some do not require their contributors to create a profile page, or they otherwise allow them to remain anonymous. To exclude these cases, we isolated identifiable generic evaluators, for example, those without real or login names on L'Internaute ("Philippe", "Patrick C."), the "Anonymous User" on LaFourchette, and "A reviewer from Facebook" on TripAdvisor.

of the intensity of their activity) include 95% of retrieved reviews from LaFourchette, 90% from TripAdvisor, and 54% from L'Internaute.

A Heterogeneous Distribution of Participation

At first sight, synthetic participation rates on OCR websites appear to be moderate (Table 5): between 1.5 and 3 reviews per contributor on average, and an average number of reviews per restaurant between 3 (Qype) and 10.6 (TripAdvisor)—with the notable exception of LaFourchette, with more than 150 reviews per restaurant on average. There are two explanations for this. First, as a restaurant booking service, LaFourchette sends diners an e-mail asking for a review after each and every reservation. Second, as a selective guide, LaFourchette works with a smaller but active established group, in contrast to the extensive logic favoured by its competitors.

Table 5. Average rating activity.

	Number of rated restaurants	Number of ratings	Average number of ratings per restaurant	Average number of ratings per contributor
Cityvox	25,223	208,222	8.3	2.8
LaFourchette	4,200	642,549	153.0 *	2.2 **
L'Internaute	61,782	482,407	7.8	1.5
Michelin—consumers	14,260	67,679	4.7	1.8
Qype	29,672	88,881	3.0	2.9
TripAdvisor	31,999	338,722	10.6	1.9

* 40% of these ratings contain a written reviews, 60% are made of a simple score.

** On LaFourchette, the average number of reviews per contributor is probably underestimated because the base of restaurants is constantly evolving. We consider only the reviews of restaurants listed in our extraction and skip the reviews of restaurants that were previously on the site.

However, as is the case in almost all participatory platforms on the web, the observed behaviour does not follow normal distributions but power laws, and the averages reflect only imperfectly behaviour. Thus, many restaurants have few reviews, and few restaurants have many (Table 6): between 20% (L'Internaute) and 44% (Qype) of rated restaurants have received a single review, a third have two or three, and a tiny fraction have received more than 50 reviews. LaFourchette is an exception: its rating system does not list reviews for establishments that have fewer than two ratings, and the site also has a strong ability to generate feedback following reservations, so that 40% of the restaurants on this site have more than 50 reviews.

Table 6. Distribution of rating activity over restaurants.

	Number of ratings received by restaurants					
	1	2–3	4–9	10–49	50+	*Total*
Cityvox	27%	25%	25%	21%	2%	*100%*
LaFourchette	0%	3%	18%	36%	43%	*100%*
L'Internaute	20%	24%	31%	24%	1%	*100%*
Michelin—consumers	36%	30%	21%	12%	0%	*100%*
Qype	44%	32%	20%	4%	0%	*100%*
TripAdvisor	20%	23%	28%	26%	3%	*100%*

As far as contributors are concerned, the general shape is similar. On the basis of usable contributor data, between 50% (Cityvox) and 81% (L'Internaute) of contributors have left only one rating, and only between 1% and 5% have written more than 10 reviews (Table 7).

Table 7. Distribution of the number of ratings per contributor.

	Number of ratings per unique contributor					
	1	2–3	4–9	10–49	50+	*Total*
Cityvox	50.3%	32.4%	13.6%	3.4%	0.2%	*100%*
LaFourchette	66.4%	21.1%	9.7%	2.7%	0.1%	*100%*
L'Internaute	80.9%	14.0%	4.1%	0.9%	0.0%	*100%*
Michelin—consumers	71.9%	19.6%	6.9%	1.5%	0.0%	*100%*
Qype	65.8%	18.4%	10.7%	4.8%	0.4%	*100%*
TripAdvisor	67.6%	22.5%	8.3%	1.6%	0.0%	*100%*

Therefore, the active minority weighs heavily in the total production of reviews: overall, 20% of the most active contributors generate half of the reviews, and the 1% who are most active (between 300 and 2800 Internet users according to the website) have written between 10% and 20% of the reviews (Table 8).

Table 8. Proportion of ratings made by the x% of most active contributors.

	The x% most active...			
	20%	10%	5%	1%
Cityvox	56.2%	42.4%	31.6%	14.8%
LaFourchette	56.6%	42.1%	30.1%	12.0%
L'Internaute	47.0%	34.4%	25.4%	11.4%
Michelin—consumers	51.1%	37.8%	27.1%	11.2%
Qype	66.2%	52.9%	40.3%	19.2%
TripAdvisor	51.1%	37.1%	25.8%	10.2%

An active minority thus produces the majority of evaluations. Should we conclude that all assessments are biased? In the context of participatory platforms, it is necessary to examine whether each contributor "weighs" the equivalent of its content creation activity (here, ratings). Yet, individual ratings are not displayed as such; they are aggregated into average scores, which are used to determine the rankings. As a consequence, individual ratings can be more or less diluted in the average score depending on whether the restaurant has received many or few ratings. Thus, a contributor who has given only one rating would see it buried in the average if it concerned a very popular restaurant, while that contributor would instead be the sole judge if the restaurant had received only his or her rating. In other words, it is important to examine whether intensive contributors weigh as much in the mean scores of restaurants as the less active ones. To check this out, we split the contributors' database into five groups according to their reviewing activity (from 1 rating to 50 and more). We compared the groups in terms of their share in the total number of ratings, and their weight in the restaurants' average evaluation (each rating of a restaurant is divided by the number of evaluations the restaurant has received) (Table 9). For instance, in the case of Cityvox, the contributors that had posted only one rating produced 18% of the overall ratings and accounted for 21% in the final average evaluation of the restaurants.

Table 9. Contributors' weight in restaurant evaluation regarding their activity.

Contributors' activity	% of ratings	Weight (%)
1 rating	18%	21%
2–3 ratings	29%	25%
4–9 ratings	26%	23%
10–49 ratings	21%	21%
50 or more	6%	10%

Cityvox

Contributors' activity	% of ratings	Weight (%)
1 rating	30%	46%
2–3 ratings	22%	21%
4–9 ratings	24%	18%
10–49 ratings	20%	12%
50 or more	5%	3%

LaFourchette

Contributors' activity	% of ratings	Weight (%)
1 rating	42%	41%
2–3 ratings	20%	21%
4–9 ratings	16%	16%
10–49 ratings	11%	12%
50 or more	10%	10%

L'Internaute

Contributors' activity	% of ratings	Weight (%)
1 rating	40%	44%
2–3 ratings	25%	24%
4–9 ratings	21%	19%
10–49 ratings	14%	12%
50 or more	1%	1%

Michelin—consumers

Contributors' activity	% of ratings	Weight (%)
1 rating	40%	44%
2–3 ratings	25%	24%
4–9 ratings	21%	19%
10–49 ratings	14%	12%
50 or more	1%	1%

Qype

Contributors' activity	% of ratings	Weight (%)
1 rating	32%	30%
2–3 ratings	25%	24%
4–9 ratings	21%	24%
10–49 ratings	12%	15%
50 or more	10%	8%

TripAdvisor

Results are similar for the six OCR websites: whatever the level of activity, the weight in the production of reviews is quite similar to the weight in the establishment of average scores (+/- 3 points). In other words, the assessment of restaurants clearly reflects the activity of all members of each platform. The democratization of gourmet judgment brought about by online ratings and reviews is a double trigger: any Internet user can participate in the evaluation, and its weight in the ratings will reflect the user's level of contribution.

3.2. A Lenient and Very Homogeneous Rating

Contrasting strongly with the heterogeneous distribution in the number of reviews given by Internet users and received by restaurants, average scores fall within a narrow and moderate range. Moreover, they are generous. This characteristic of online consumer rating is noticeable on all of the platforms: a majority of restaurants get an average rating of 4/5 (Table 10).

Table 10. Average and median rating of restaurants.

	Average	Median	Mode	Standard deviation
Cityvox	3.75	3.83	4.00	0.78
LaFourchette (/10)	7.94	8.00	7.90	0.61
L'Internaute	3.63	4.00	4.00	0.84
Michelin—consumers	3.98	4.00	4.00	0.72
Qype	3.70	4.00	4.00	1.03
TripAdvisor	3.81	4.00	4.00	0.84

This property is very consistent from one platform to another. The average of (average) scores of restaurants is between 3.63/5 and 3.98/5 (for LaFourchette, which gives scores out of 10, it is 7.94). On all websites, the mode is 4. It is therefore a property independent of the characteristics and features of different websites: in general, Internet users' ratings produce a mild score centred on 4 out of 5. As a consequence, the distribution of scores is Gaussian (Figure 1): most restaurants have a rating closer to the average, which is also the median (4/5). In contrast, few restaurants have a very good or a very bad score.

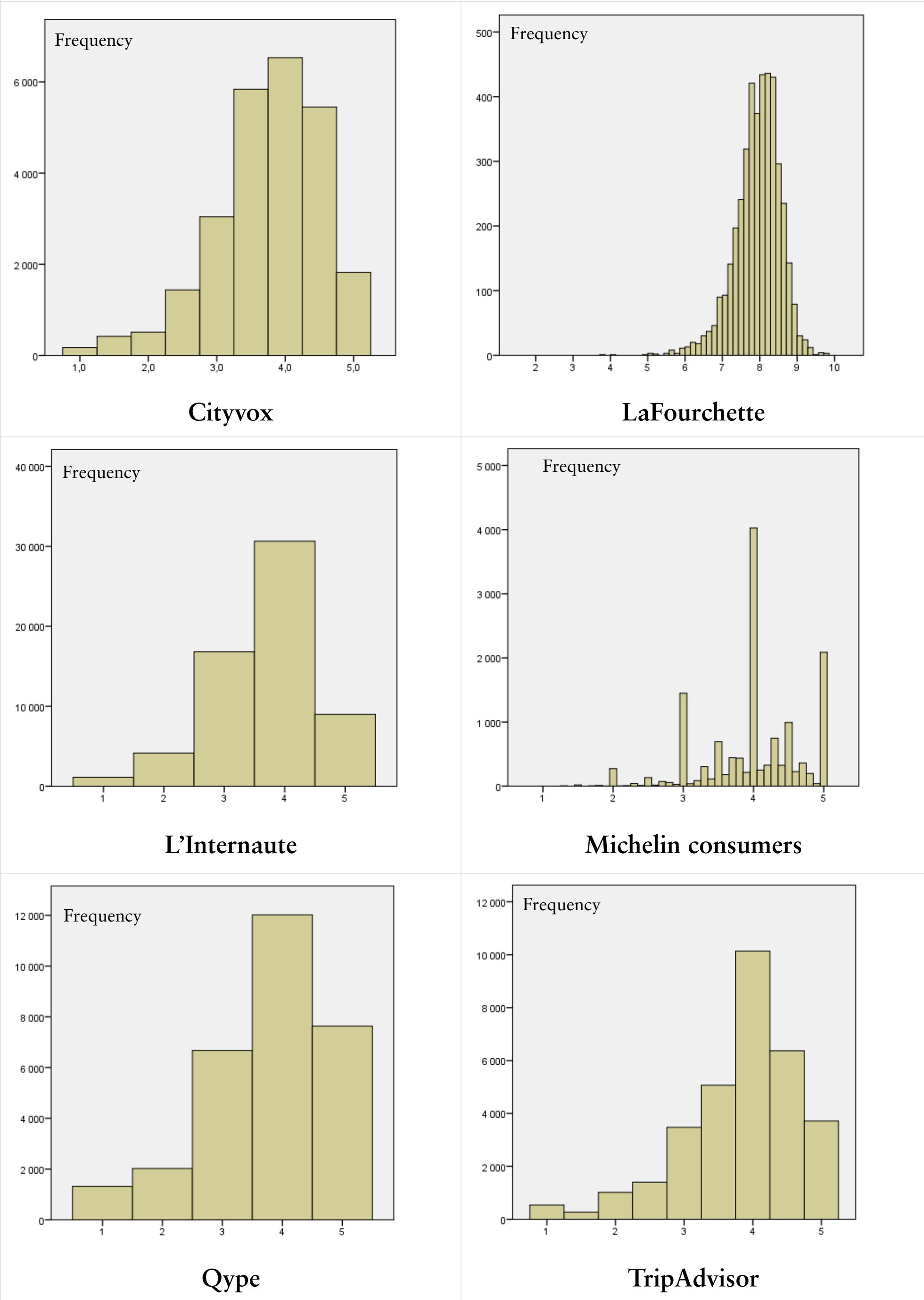


Figure 1. Distribution of average scores of restaurants.

This grading of the quality of restaurants is in stark contrast with the distribution of the ratings given by conventional gourmet guides (Figure 2). *Bottin Gourmand*, *Gault & Millau*, and the *Michelin Guide* are all extremely parsimonious about bestowing good grades, awarding the highest scores only to a handful of institutions. The result of this selective and elitist approach is a pyramidal distribution of restaurants, each upper step being more difficult to achieve than the previous one.

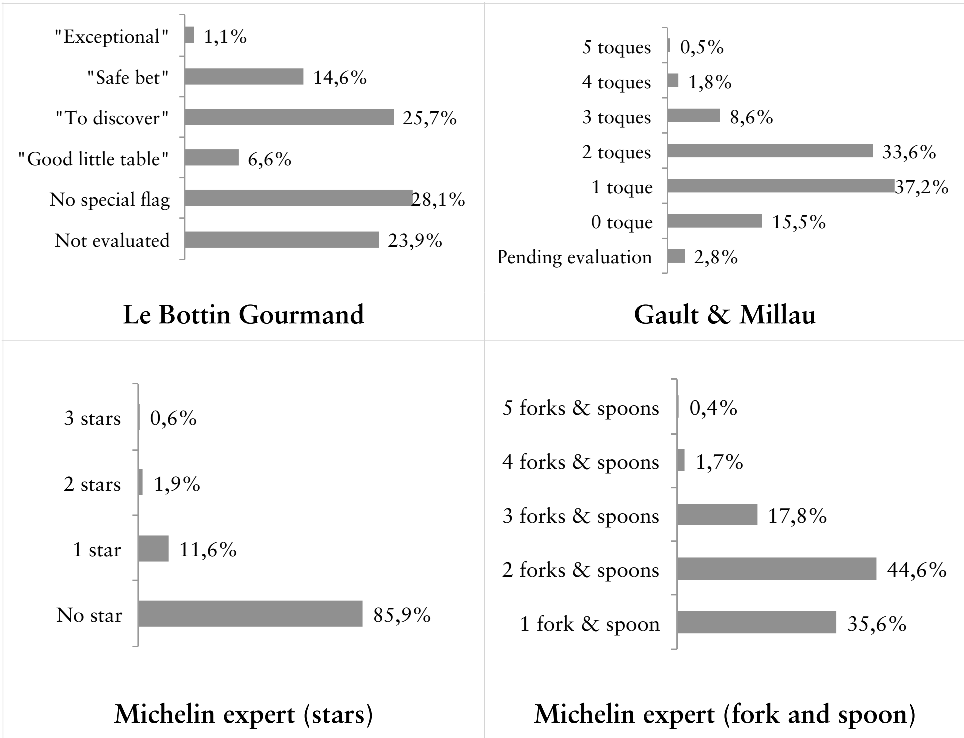


Figure 2. Distribution of restaurant ratings by traditional gourmet guides.

OCR websites build upon the raw material of (individual) ratings and transformations imposed on this material (mean score, rounding) in order to rank restaurants. Due to the particular character of participation, platforms must perform double duty: on the one hand, they must constantly stimulate the production of ratings on as many restaurants as possible; on the other hand, they must manage a very homogeneous—and thus undifferentiating—scoring, which is at odds with the traditional pyramidal grading of guidebooks. This double work implies editorial choices that vary according to the sites. It refers to different normative horizons of the democratic ideal.

Websites' Strategies and Their Democratic Achievements

As an activity performed by Internet users, rating has two main features: it is highly concentrated and unequally distributed, and average scores are very homogeneous. Website managers who were interviewed often indicated that they regard these features as a brake on business development: the heterogeneity of the distribution of ratings limits the audience (and revenue) of platforms by pointing out a small subset of restaurants; the homogeneity of the scores limits their ability to direct consumers by effectively recommending restaurants. Two different strategies are implemented by website managers to resolve these tensions. Both are shifts away from the democratization movement that we have described above. The first strategy aims at enhancing restaurant selection and differentiation, using various techniques that compensate for the homogeneity of ratings: equality among contributors is maintained at the expense of a decline in democratization-as-inclusion. The second entails differentiating contributors by bringing forth the "connoisseurs" (Blank 2007) among them, at the expense of equality within the group.

OCR websites offer a combination of two operations: commensuration (through an average score from one to five) and singularization (through the free expression of customers in reviews). By comparing the sites and their strategies to address the excessive homogeneity of ratings, we observe how the material apparatuses tend to favour one or the other of these two dimensions. If OCR websites may be considered as "algorithmic apparatuses" (Orlikowski and Scott, forthcoming), their algorithmic policies differ from one site to another, even from one moment to another in the site's history, producing different equilibria between commensuration and singularization.

Equality Among Contributors, Selection of Restaurants

The first strategy used to create differentiation in a homogeneous universe is to maintain equality among the (many) participants while emphasizing the selection and visibility of the best restaurants in the manner of a traditional guidebook. However it is based on the ability to aggregate a large number of reviews for each item evaluated.

The homogeneity of scores encourages platforms to implement effective cognitive tools to orientate consumers in a world where most restaurants are rated between 3.5 and 4.5. Apparently innocuous algorithmic and editorial choices take on great importance. We can identify several decisions shaping the websites that are made in response to the overall homogeneity of ratings.

One seemingly innocuous aspect is the choice of whether or not to round the average scores of restaurants. On a scale of 1 to 5, or 10 for LaFourchette, such a decision leads either to giving the same score to a large number of restaurants, or conversely, allowing a full hierarchical grading of them. Our observations show distinct policies in this area (Table 11).

Table 11. Rounding policy of OCR websites.

	Scale	Rounding
Michelin	5	none
LaFourchette	10	none
TripAdvisor	5	half
Cityvox	5	half
L’Internaute	5	integer
Qype	5	integer*

* Rounding changed to “half” in 2013 to enable the integration of ratings on Qype and Yelp after the former was purchased by the latter.

More generally, platforms develop editorial strategies in their selection and ranking of restaurants, in order to guide consumers effectively. LaFourchette explicitly positions itself as a new kind of “gastronomic guide” and not just another directory. Several times in the interview, the manager insisted on the role of the website as a reliable and effective source of information for consumers by providing, in the manner of a guide, a ranking of restaurants. If quality emerges from consumer reviews rather than from inspectors, the service provided to the consumer should be similar:

> We want to provide a service; we do not want to be just another search engine . . . I used the popular term of “curation.” We want to offer a limited sample of quality restaurants. The goal is not to suggest all restaurants, it is not to allow booking in any restaurant, it is to allow you to book in the best restaurants. That’s really our goal. Our service needs to be perfect, and so does the supply, so the quality of partner restaurants must be excellent. So yes, LaFourchette is positioned like a guide. (LaFourchette)

This positioning is reflected in several strategic choices made by website managers. The most important is the removal of poorly rated restaurants from the list: businesses whose average rating is stabilized below 6.5 (out of 10) are not listed on the site.

> In any case, when a restaurant starts having an average score below 6.5 to 7 over a significant period, we do quality control: we call up the restaurant, we might send someone for lunch or dinner and if we see that the rating is deserved, the partnership is broken. People often say, “But there are only good ratings in LaFourchette”—yes, and so much the better, since it means that we’re doing well our job of curation, because our goal is not to list the bad restaurants. (LaFourchette)

This strategy is also consistent with the policy, as we have seen, not to round the average score, so as to provide a complete ranking of the restaurants within the narrow range of visible ratings (from 6.5 to 10). This policy is possible thanks to the high number of ratings per restaurant (153 on average). In the case of LaFourchette, the goal of providing useful recommendations requires a relaxing of the constraint of exhaustiveness, and a partial renunciation of democratization-as-inclusion. LaFourchette promotes a model of excellence focused, of course, on selectivity; it relies on the website team to select the restaurants and these are, in turn, endorsed by the website's contributors through their ratings and reviews.

By combining the explicit recommendation of the guide and the pursuit of completeness peculiar to OCR websites, TripAdvisor offers another kind of algorithmic compromise between exhaustiveness and recommendation. On the one hand, the logic of the widest possible evaluation prevails, with restaurants rated between 3.5 and 4—scores being rounded to a half point. On the other hand, the site has also built for each city a complete ranking of all the restaurants. The ranking algorithm, which is kept secret, seems to combine the average score and the number of reviews. This ranking thus provides, beyond a relatively homogeneous scoring, a simple and explicit recommendation tool. In addition, TripAdvisor has recently developed "cityguide" applications for smartphones: for different cities, the site offers a selection of the best-rated places, thus approaching the traditional travel guide model.

The selective logic—embraced by LaFourchette and moderately employed by TripAdvisor with its smartphone applications—is not implemented by the other sites. The latter, which are sometimes local directories including various local businesses, must mobilize other resources to produce differentiation and provide effective recommendations.

Pushing "Connoisseurs" to the Forefront

A key symbol of democracy is the vote, which on the OCR websites is emulated by calculating the unweighted average of individual ratings. Its full realization assumes that each review is anonymous (like a ballot paper), and that all ratings weigh the same. However, we observe that personal details about the evaluator are often attached to the review. Thus, all platforms except Nomao associate each review with its author, who is described at least by a pseudonym and a profile that can include a photograph, personal information, statistics on ratings and reviews already published, etc.

In providing author details, platforms let the figure of the individual reviewer emerge. Website visitors will identify with this person or not, according to the displayed information and the credibility given to it. Another sign of this customization is when various compliments, badges, and statutes are distributed to contributors. These devices are very common, and they tend to single out contributors and give greater weight to the reviews of the most prolific authors.

Special consideration is often given to those who contribute the most, but also to those who receive the most evidence of confidence from the community: Amazon's "x people found the following review helpful" feature pioneered this field. Chen et al. (2001) showed that reviews deemed useful had a significantly stronger (positive) impact on the sales of books on Amazon. This button is now displayed on most OCR websites, including TripAdvisor, Qype, Michelin, Cityvox, and especially Yelp, where users are allowed to tag reviews as "useful," "funny," or "cool."

Provided that confidence in a review depends on the credibility of its author, the websites that display these metrics contribute implicitly to weighing contributors' importance. At Yelp, gaining entry to the "elite squads" of contributors certainly requires intensive assessment activity, but it also requires participation in discussions and events with the community which, in turn, influences opinions about the relevance and usefulness of specific contributors' reviews. These votes allow platforms to identify contributors acclaimed by the community and possibly to highlight them. They can become identified as among the top-ranking contributors (e.g., "Top-thousand" reviewers at Amazon; "Club 300" at Allociné) or otherwise singled out as important members, as with the "elite squad" at Yelp.

To reward the most active contributors and manage the community, a special relationship develops sometimes between platforms and their contributors, through the organization of cocktail parties (Dismoiou), meetings, and dinners (LaFourchette) or regular events, such as here at Yelp:

> In fact, every month we organize an elite event, which is an event dedicated to the most active and influential members. And the event is a way to move from the virtual to the real, from online to offline, to make the community lively outside the site, and create a special moment for those contributors who are, in a sense, ambassadors for Yelp. (Yelp)

The elitist approach of giving greater visibility and weight to the reviews produced by a subset of amateur experts—"connoisseurs" to use Blank's term—is carried to completion by iTaste. This website, unlike the others, includes a "follow" button, in the manner of Twitter. As they contribute and are followed by more users, active contributors increase their "reputation score." This score is directly used to moderate each contributor's rating when computing the restaurants' average scores, as explained by the website: "In order to make the grades more reliable, the ratings are influenced by one's reputation. For instance, if Peter (138 Reputation Points) gives 19/20 to a restaurant and Paul (1579 Reputation Points) gives 12/20, the average grade will be 13/20, not 16/20 as it was before."[17] In other words, iTaste implements the Matthew effect by giving greater importance to the voices of the most active contributors.

In an interview, a manager from LaFourchette declared his interest in this model and his intention to move closer to it—although this project was finally abandoned:

> Roughly, the idea of establishing a form of hierarchy between users where I can very well follow you without you following me, where I can very well be your scout without you being mine, where I can very well be your guide without you guiding me, all that is very important to us, this is part of a real dynamics . . . We could implement it, even if it is kind of unhealthy but so human, this form of competition between our users. We could give them badges, awards, and have a sort of leaderboard of the most reliable LaFourchette customers. And when you see a customer who left a good review or has a good score or a good badge: "I am the couscous pro," etc., you'll want to follow him and you'll be able to follow him and you will be alerted every time he posts a review—exactly like what happens on SensCritique or on Twitter. (LaFourchette)

This elitist approach involves introducing distinctions between various worlds of tastes and thus reconciles the principles of diversity of tastes and equality of participants. It permits the definition of specialists, Internet users who are more knowledgeable about certain types of foods or places and whose opinions will be considered most useful by those who share their tastes. In this respect, the Michelin website also has badges that attach regular contributors to a certificate of expertise in specific areas—these include not only "Gourmet food," and "French cuisine" but also "Pizza," "Creperie," and "Chinese cuisine."

Intensive contributors, motivated by badges certifying their expertise on some foods, are likely to become authors, such as those professionals found in Ferguson's (2008) work under the label of "judge"—embodied traditionally in the United States by culinary editors of prominent magazines and newspaper sections, or personified in France by Gilles Pudlowski. Some contributors on OCR sites are getting close to becoming such figures, although they are not distinguished or rewarded by the websites:

> [Among contributors] there is a very small portion of people who write huge essays. Since reviews are relatively unlimited in size, sometimes you will have an article, we have writers, people who like to think they are food critics and are indeed not far from the quality of a food critic. (LaFourchette)

Contrasting with this logic of acknowledgement, and sometimes of grading, of contributors is Nomao's approach. This platform aims to produce effective recommendations by aggregating and summarizing

[17] iTaste FAQ, http://www.itaste.com/en/application/faq.php (visited: 01/2014).

"all" the information from the web, not just assessments produced by the contributors on the platform. The platform seeks to index the maximum textual content about the places it rates (name + address). Textual content is broken down into semantic units (words or noun phrases) to which are attached positive or negative qualifications; they are also linked to a category (environment, atmosphere, service, menu, prices, etc). The overall rating of the place is derived from the synthesis of all of these positive and negative evaluations. A written review will have even more weight in the final score if its semantic content is rich and it comes from a source with a good reputation score. In other words, Nomao does not synthesize votes, but semantic evaluations and final assessments are not attached to its authors, who have no visible presence on the platform. Thus, with Nomao, the democratic principle is embodied in a radical computational and algorithmic form:

> The more advanced site when Nomao was launched was Yelp. It was interesting because, at the time, Yelp was already cluttered with content. When I say cluttered, it's the places that are a little trendy where you found yourself with hundreds and hundreds of reviews. And finally, as a user, either you know some other users that you trust, you'll read what they say as you would read a blog . . . Or you're searching and you are completely drowning in all that information. And finally, when you look at user behaviour, you realize that people end up looking at counters. They say "this restaurant, 500 reviews, a score of 4.5 out of 5, it looks good to me." And finally, the uploaded reviews, they are just useless . . . . So yes, the goal was to index everything you can find, everything that is produced by Internet users and to find a way to deal with it in order to, on the one hand, describe a database correctly, with depth, so we can say "This address is a restaurant, it is a gourmet restaurant, the terrace is nice, reception is friendly, etc."; and on the other hand, to rebuild the social graph of users connecting them to local places through shared affinities. The aim is to get to determine that a user—whose identity we do not seek to know—likes this restaurant and this restaurant because he has given good ratings to certain restaurants. . . . Indeed, we are today the only ones to make this work." (Nomao, founder and CEO)

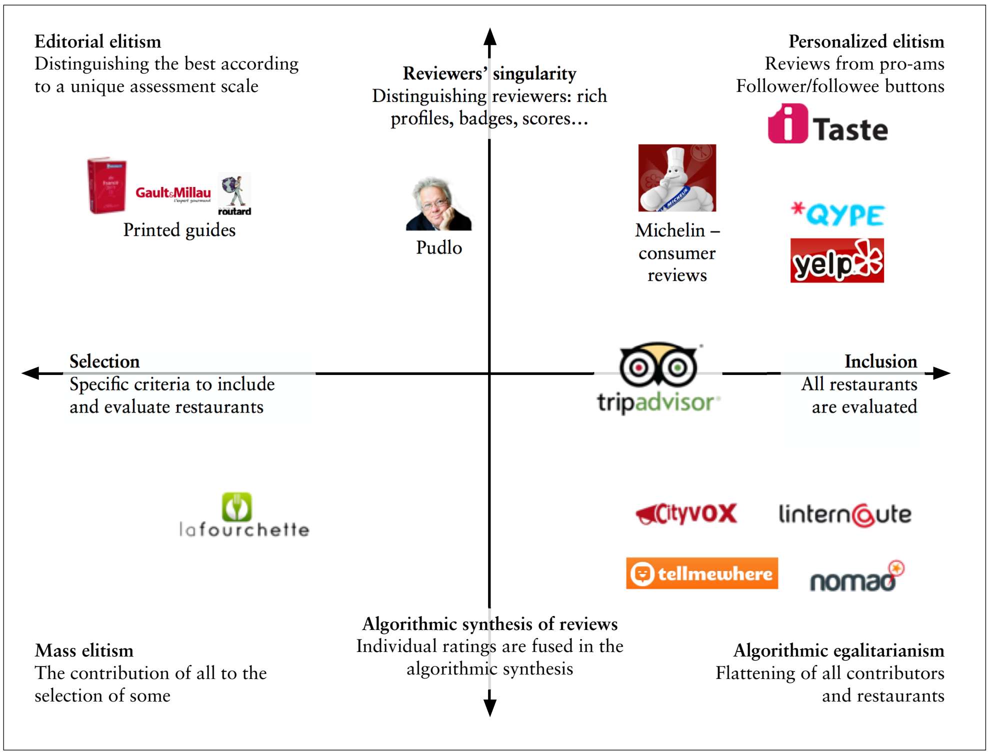
Democratic Achievements of OCR Sites: A Typology

At the beginning of this article we defined two democratization lines: democracy-as-inclusion and democracy-as-participation. We have seen that OCR sites have begun to achieve the first movement by broadening the base of reviewed restaurants, and they have fully opened the second one. However, along with this democratic movement comes tension between the ultimate need to distinguish restaurants from one another and the need to provide useful and effective recommendations to consumers.

The first source of tension is linked to the inclusive dynamics and refers to the relative equality among restaurants. While including all of the supply side, review sites are mixing different worlds of tastes and budgets, and they are flattening the differences between trades. The average kindness of lay rating reinforces this movement: it follows from these ratings that all restaurants are so close to each other that initial differences are denied. This flattening is treated variably by review sites: for some, the solution is to reintroduce selectivity, either by excluding a subset of the supply (LaFourchette) or by offering selections in the mass of restaurants ("Cityguide" mobile applications and "Travelers' Choice" selections of TripAdvisor); for others, it consists instead in including as many establishments as possible and then provide filters to categorize and manage the information (L'Internaute, Cityvox, TripAdvisor). On one side, the selective action is taken by the guide, on the other by consumers.

A second tension exists between participation and equality. If sites postulate *a priori* that all judgments are equal, the overall judgment on restaurants is formulated primarily by the most active participants because of the structure of online participation. Review sites need these intensive contributors, and they do not offer devices to temper their weight. Instead, they adopt a variety of postures, ranging from concealing individual contributors in the aggregate assessment, to highlighting each contributor's personality, tastes, and rating history.

Based on these two axes, we can outline a typology of OCR sites that distinguishes four embodiments of the democratic ideal under the constraint of effective recommendations:



Let us detail these four models:

1. *Editorial elitism* or "editism" is primarily chosen by printed guides and online editorial spaces; they set up in the same movement evaluation and inclusion criteria in the final list of restaurants.
2. *Mass elitism* keeps editorial elitism's selective logic while showing only well reviewed restaurants, but lets the assessment rely on the mass of lay reviewers. Cohabitation between different worlds of taste and judgment scales is counterbalanced by the decision to list only "good" restaurants. LaFourchette is emblematic of this logic. Democratization of food criticism implies here the removal of the individual reviewer for the sake of improving the quality of the recommendation.
3. *Personalized elitism* stages the diversity of tastes by personifying them through the profiles of intensive contributors. Democratization of food criticism is understood here not as the disappearance of the figure of the critic, but as the opening up of the role to amateurs: each contributor can adopt a critical posture and build and share his or her personal selection of restaurants within a comprehensive list. This logic is associated with rich profile pages of intensive contributors and internal social network tools (iTaste, Yelp).
4. *Algorithmic egalitarianism* is embodied by sites like L'Internaute, Dismoiou, and particularly Nomao. This democratic approach is inclusive and melts subjectivities in the algorithmic synthesis. It produces maximalist lists of restaurants whose average rating is only one characteristic among several others. What is important here is that the Internet users can choose, by using filters (such as location, price, type of food etc. as well as score), the restaurant that will best match their preferences.

Summary and Conclusion

Our aim in this article was to clarify, and qualify, the common claim that online consumer reviews contribute to the empowerment of consumers and the democratization of markets. The article produces evidence that contributes to the following argumentation. We identify two movements of democratization within culinary criticism prior to the Internet: inclusion and participation (section 1). We show that the development of OCR websites extends these two movements, by achieving inclusion and by systematizing participation (section 2). Then, we highlight two features that come with online democracy "in action": the unequal distribution of contributions, and the homogeneity and high level of average scores (section 3). These characteristics of rating are constraints that platforms have to deal with, because they are in conflict with the aim of building fair and meaningful recommendations. The fourth section describes the different types of compromise set up by websites in order to articulate the participation of all users and the effectiveness of the recommendations. This requires a weakening of the goal of democratization.

In our view, the article is a relevant contribution to the emerging field of valuation studies for two main reasons. First, it examines a valuation device that as yet has been little studied as such (online consumer reviews), examining both its construction and operation. We paid special attention to the framings made by OCR websites and to the material aspects of the production of valuation. If the articulation of the two opposite operations of commensuration and singularization is taken to traditional valuation devices—that is, reviews produced by experts (Blank 2007; Karpik 2010)—it is propelled to a new, algorithmic, scale in order to produce valuation from multiple disparate lay judgments. We also spotted the many subtle variations between websites and we stressed the importance of these variations.

Second, the article contributes to the study of the "politicization" of markets and valuation (by analogy with "economization," see Çalıskan and Callon 2009) by investigating the movement of democratization supposedly carried out by the Internet. We observed the (imperfect) making of a new subject in market agencement: a kind of citizen-consumer who is equipped with, and empowered by, new capabilities to access market information and to voice opinions. Meanwhile, in our case, the ideological discourse of the promoters of OCR has to be qualified. It also results in an *ex post* rationalization by Internet entrepreneurs who were searching for economic opportunities and improvised a lot. While there was an ideological and rhetorical commitment by the founders and developers of some of the online review sites, there were other motivations involved, such as the need for competitive positioning vis-à-vis what at the time were very well established and powerful incumbent publishers, and the ready availability of certain easy-to-integrate ranking and reviewing tools. In a sense, review sites had to use what they could to gain a foothold in the markets that were dominated by the incumbent players, and this involved democratization of both access and participation.[18]

The development of online consumer rating and review websites still marks a real democratization of calculative capacities in the restaurant industry, in terms of inclusion as well as participation. Is it a democratization of the market itself, understood as the pluralization of valuation devices and their openness to all stakeholders? Pluralization of valuation devices, if it does increase the power of consumers, does not, however, automatically imply a greater demand (more customers) for restaurant owners, but rather imposes new intermediaries whose

[18] We are grateful to an anonymous reviewer for bringing this issue to our attention.

strength they are discovering (Bessy and Chauvin 2013). Future work could therefore study how the restaurant owners and managers themselves—and producers and distributors in other industries—welcome and appropriate these new evaluation devices. Public statements by professional actors suggest that online consumer reviews are largely perceived as an illegitimate constraint rather than as an opening of the competitive game and a gain in transparency.

Acknowledgements. We are very grateful for the insightful comments and suggestions made by the two anonymous reviewers of our paper. In addition, we thank Liz McFall for her careful review of an intermediate version of this article. Usual caveats apply.

References

Anderson, Michael, and Jeremy Magruder. 2012. "Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database." *The Economic Journal* 122 (563): 957–989.

Barrère, Christian, Quentin Bonnard, and Véronique Chossat. 2010. "Democratization in the Gastronomic Market: From Michelin Stars to Michelin 'Bibs'." The 16th ACEI Conference, June, Copenhagen.

Beauvisage, Thomas, Jean-Samuel Beuscart, Vincent Cardon, Kevin Mellet, and Marie Trespeuch. 2013. "Notes et avis des consommateurs sur le web. Les marchés à l'épreuve de l'évaluation profane." *Réseaux* 177: 131–161.

Benkler, Yochai. 2006. *The Wealth of Networks. How Social Production Transforms Markets and Freedom*. New Haven and London: Yale University Press.

Bessy, Christian, and Pierre-Marie Chauvin. 2013. "The Power of Market Intermediaries: From Information to Valuation Process." *Valuation Studies* 1 (1): 83–117.

Blank, Grant. 2007. *Critics, Ratings, and Society. The Sociology of Reviews*. Lanham: Rowman & Littlefield.

Bonnet, Estelle. 2004. "Les critiques gastronomiques: quelques caractéristiques d'une activité experte." *Sociétés contemporaines* 53: 135–155.

Bourdieu, Pierre. 1984. *Distinction. A Social Critique of the Judgement of Taste*. Cambridge: Harvard University Press.

Çalıskan, Koray, and Michel Callon. 2009. "Economization, Part 1: Shifting Attention from the Economy Towards Processes of Economization." *Economy and Society* 38 (3): 369–398.

Callon, Michel, and Fabian Muniesa. 2005. "Economic Markets as Calculative Collective Devices." *Organization Studies* 26 (8): 1229–1250.
Cardon, Vincent. 2014. "Des chiffres et des lettres. Evaluation, expression du jugement de qualité et hiérarchies sur le marché de l'hôtellerie." *Réseaux* 143: 205–245.
Chen, Pei-Yu, Samita Dhanasobhon, and Michael D. Smith. 2001. "All Reviews Are Not Created Equal: The Disaggregate Impact of Reviews and Reviewers at Amazon.com." Heinz Research Paper 55.
Davis, Mitchell. 2009. *A Taste for New York: Restaurant Reviews, Food Discourse, and the Field of Gastronomy in America*. PhD Dissertation, Steinhardt School of Culture and Human Development, NYU.
Ferguson, Priscilla Parkhurst. 1998. "A Cultural Field in the Making: Gastronomy in 19th-Century France." *American Journal of Sociology* 104 (3): 597–641.
———. 2008. "Michelin débarque à New York: la loi des guides ?" In *Voyages en gastronomies*, edited by Julian Csergo and Jean-Pierre Lemasson, 116–125. Paris: Editions Autrement—mutations.
Gould, Carol. 1989. *Rethinking Democracy: Freedom and Social Co-operation in Politics, Economy and Society*. Cambridge: Cambridge University Press.
———. 2004. *Globalizing Democracy and Human Rights*. Cambridge: Cambridge University Press.
Jeacle, Ingrid, and Chris Carter. 2011. "In TripAdvisor We Trust: Rankings, Calculative Regimes and Abstract Systems." *Accounting, Organizations and Society* 36 (4-5): 293–309.
Johnston, Josee, and Shyon Baumann. 2007. "Democracy versus Distinction: A Study of Omnivorousness in Gourmet Food Writing." *American Journal of Sociology* 113 (1): 165–204.
Karpik, Lucien. 2000. "Le *Guide Rouge Michelin*." *Revue Française de Sociologie* 42 (3): 369–389.
———. 2010. *Valuing the Unique. The Economics of Singularities*. Princeton: Princeton University Press.
Lane, Christel. 2013. "Taste Makers in the 'Fine-Dining' Restaurant Industry: The Attribution of Aesthetic and Economic Value by Gastronomic Guides." *Poetics* 41 (4): 342–365.
Luca, Michael. 2011. "Reviews, Reputation, and Revenue: The Case of Yelp.com." Harvard Business School Working Paper 12-016, September.
Luca, Michael, and Georgios Zervas. 2013. "Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud." Harvard Business School Working Paper n° 14-006, September.
Mallard, Alexandre. 2000. "La presse de consommation et le marché. Enquête sur le tiers consumériste." *Sociologie du travail* 42 (3): 391–409.
Naulin, Sidolie. 2012. *Le journalisme gastronomique. Sociologie d'un dispositif de médiation marchande*. PhD Dissertation, Université Paris IV.
Nomao. 2012. Les avis utilisateurs sur les commerces locaux en Europe. Report, July.

Orlikowski, Wanda J., and Susan V. Scott. Forthcoming. "What Happens When Evaluation Goes Online? Exploring Apparatuses of Valuation in the Travel Sector." *Organization Science*.

Rao, Hayagreeva, Philippe Monin, and Rodolphe Durand. 2003. "Institutional Change in Toque Ville: Nouvelle Cuisine as an Identity Movement in French Gastronomy." *American Journal of Sociology* 108 (4): 795–843.

Scott, Susan V., and Wanda J. Orlikowski. 2012. "Reconfiguring Relations of Accountability: Materialization of Social Media in the Travel Sector." *Accounting, Organizations and Society* 37 (1): 26–40.

Shaw, Steven. 2000. "The Zagat Effect." Commentary, November.

Weygand, Félix. 2009. "Le développement d'un modèle d'affaires hybride: le cas de Cityvox." *Terrains & Travaux* 15 (1): 105–126.

Zukin, Sharon. 2004. *Point of Purchase: How Shopping Changed American Culture*. New York: Routledge.

Thomas Beauvisage is a researcher at the Social Sciences Department of Orange Labs (Sense). His research focus on social media, online advertising and market studies.

Jean-Samuel Beuscart is a researcher at the Social Sciences Department of Orange Labs (Sense) and associate researcher at the Université Paris-Est Marne-la-Vallée. His research focuses on the digitization of cultural markets. Recent research subjects include online advertising, self-promotion and online visibility of cultural producers.

Kevin Mellet is a researcher at the Social Sciences Department of Orange Labs (Sense) and associate researcher at the Centre de Sociologie de l'Innovation (Mines ParisTech). His research focuses on socio-economic consequences of the Internet on market intermediation and valuation. Recent research subjects include online advertising, social media marketing and online consumer reviews.

Marie Trespeuch is a researcher at the Social Sciences Department of Orange Labs (Sense) and associate researcher at the IDHE (Ecole Normale Supérieure de Cachan). Her PhD dissertation focused on the transformation of the online gambling market in France from 2007. Recent research subjects include online consumer reviews and the sharing economy.

Valuation Studies 2(1) 2014: 43–64

Uncertainty and the Development of Evidence-Based Guidelines

Esther van Loon* and Roland Bal

Abstract

This article explores how developers address uncertainty in the creation of an evidence-based guideline (EBG). As the aim of an EBG is to assist healthcare practitioners in situations of doubt, it is easy to assume that uncertainty has no place in guidelines. However, as we discovered, guideline development does not ignore uncertainty but seeks to accept it while establishing credible recommendations for healthcare. Dealing with omissions in knowledge, ignorance, or challenges in valuating different sorts of knowledge form the core of the work of guideline developers. Interviewing guideline developers, we found three types of valuation work: classifying studies, grading types of knowledge, and involving expertise and clinical practice. These methods have consequences for the credibility, and amount and kind of uncertainty EBGs can include.

Key words: valuation; uncertainty; evidence-based guidelines; development

Introduction

> With a background in science, you are used to thinking you know it all. Ask me something about a disease and I'll tell you all about it. But I can't tell you what I don't know. I think we need to make that more transparent, that we also don't know a lot. (Guideline developer involved in guidelines for infectious diseases)

Developing an evidence-based guideline (EBG) is a process of valuating and bringing order into a plethora of knowledge. Guideline making is collective work in which core issues are related to what knowledge is available, how this knowledge should be valued, which actors should be involved in the process, and how recommendations can be justified

Esther van Loon, Erasmus University Rotterdam, Institute of Health Policy and Management, the Netherlands, vanloon@bmg.eur.nl

© 2014 Esther van Loon and Roland Bal.
LiU Electronic Press, DOI 10.3384/vs.2001-5992.142143
http://valuationstudies.liu.se

(Moreira 2005; Moreira, May, and Bond 2009; van Loon, Zuiderent-Jerak, and Bal 2013). This valuation inevitably meets uncertainty. Yet, at face value, uncertainty contradicts the EBG movement. EBGs are developed to provide recommendations that assist healthcare workers make the right decisions about patient care. These recommendations are based upon “a systematic review of the evidence and an assessment of the benefits and harms of alternative care options.” (Graham et al. 2011, 4). The rhetoric of EBG is that guidelines provide *certainty* for healthcare workers who are faced with patients with ambiguous complaints and treatment choices with unpredictable outcomes. That such strong rhetoric works is understandable, as healthcare workers are increasingly held accountable for their decisions. Decision-making in healthcare has become more complex due to increased options for treatment and increased awareness of diseases. Yet, the idea that guidelines are free of uncertainty or the solution to clinical uncertainty is not realistic. Timmermans and Angell, for example, have shown that using EGBs in the socialisation of doctors sometimes helps to solve clinical uncertainty, but it also reproduces new kinds of uncertainty that need to be dealt with accordingly (Timmermans and Angell 2001). Uncertainty thus remains an aspect of clinical work, despite EBGs.

Rather than focusing on uncertainties in clinical work, in this paper we focus on uncertainties inherent in creating EBGs. We are interested in finding out how uncertainty manifests itself in this process, and what kind of valuation work is undertaken to engage with uncertainty. Valuation work is the social practice of bringing order into all kinds of information and signifying or giving worth to this information (Kjellberg and Mallard 2013; Helgesson and Muniesa 2013). It involves both the assessment of values (i.e., literature, opinion, expertise) and the reproduction of values into recommendations for EBGs. Uncertainty is an inevitable element in this process. We suggest that uncertainty in guidelines is not always detrimental. Uncertainty invokes reflection, and as we have discussed elsewhere, reflexivity in healthcare standards help practitioners to achieve good care (van Loon and Zuiderent-Jerak 2011). Yet expressing uncertainty makes one vulnerable. Therefore, as Gross puts it, “the challenge is how to knowingly and increasingly also publicly deal with what is not known without losing one’s credibility or ‘scientific authority’.” (Gross 2010, 3). The focus in this article is on how the EBG can balance between recognising and accepting uncertainty while producing reliable and credible recommendations to guide healthcare practitioners. Our research question is: *How is valuation work done to balance between acknowledging uncertainty and remaining credible in guideline development?*

To answer this question, we held semi-structured interviews with Dutch guideline developers from a wide range of healthcare organisations. The interviews focused on the struggles, debates, and valuation work of guideline developers in striving to create reliable and realistic recommendations and engage with uncertainty.

The remainder of this article is structured as follows: first, we define uncertainty and distinguish three ways in which it manifests itself in guideline development. After elaborating on the methods, we provide an analysis of our empirical findings. We discuss three valuation practices in creating EBGs: classifying studies, grading different types of knowledge and those involving expertise and clinical practice. We conclude by showing that different valuation practices have different consequences for acknowledging uncertainty.

On Uncertainty

Uncertainty gains a great deal of attention in the social science literature. Studies of decision-making on environmental issues, the practice of futurists, public involvement in science and health care are some examples (Wynne 1996; Callon, Lascoumes, and Barthe 2009; van Asselt, Mesman, and van 't Klooster 2007; Shackley and Wynne 1996; Mesman 2008). This section clarifies our approach towards uncertainty and discusses three forms in which it manifests itself in relation to the EBG. Further, we pay attention to the relation between uncertainty and ignorance.

Uncertainty is everywhere. It is part of scientific work, decision-making, and everyday life. As the opening quote of the introduction highlights, there is a general tendency to focus on certainty, rather than uncertainty. This makes uncertainty invisible to an extent (Star 1985; Shackley and Wynne 1996; Mesman 2008). The term also tends to have a negative connotation. Melse argues that it is an un-word, indicating that something is absent or missing (Melse 2003; van Asselt 2005).

In searching for a definition of uncertainty, we follow the work of Moreira who defined it as "the non-determinate or unsettled quality of a statement or knowledge claim" (Moreira 2011, 1335). Moreira's definition is highly suitable for us as his study investigated uncertainty in healthcare rationing. The reference to "unsettled" addresses the collective character of uncertainty. Uncertainty gets meaning in collaboration and discussion within a collective. However, "unsettled" also implies that work is needed to reveal uncertainties (or keep them hidden). Hence, "quality" in Moreira's definition underlines that knowledge valuation is not just the application of comparative techniques, but involves collective work. This combination, at the heart of valuation work, is what we want to study in connection with uncertainty in guidelines.

As uncertainty is often invisible and valued negatively, people are likely to avoid it, work around it or to try to overcome it. However, several authors who study ways of dealing with uncertainty point out that we should try to accept uncertainty. Jerak-Zuiderent studied patient safety and argues that healthcare practitioners must engage with uncertainty to deal with all kinds of demands. She refers to this as "living with uncertainty" (Jerak-Zuiderent 2012). Living with uncertainty has to do with the acceptance of a given degree of uncertainty in medical work, but also points to a healthcare practitioner's mind-set, to always be aware of the uncertain aspects in their work. The challenge is how to do this, and keep doing it when collective decisions must be made. For example, studying the work of futurists, van Asselt et al. refer to "certainification"—uncertainties initially acknowledged in the decision-making process eventually vanished from the definitive documents (van Asselt, Mesman, and van 't Klooster 2007). Whereas decision-makers may recognise uncertainties, these do not have to be included in the final decisions, and thus disappear into the background.

We follow Jerak-Zuiderent in considering that ignoring or banning uncertainty is not productive. To a great extent, however, it is still unknown how we can include uncertainty in EBGs so that coherent and clear recommendations that support healthcare decision-making are made. Certainification is not the solution, but the question we explore is how guideline developers balance between uncertainty and credibility. In the following section we will discuss credibility in relation to uncertainty.

Credibility Needs Uncertainty

Credibility is a key issue in developing EBG. Expressing uncertainty seems to have a great impact on becoming or remaining credible. Wynne's eminent work on Cumbrian sheep farmers shows how distrust can grow when uncertainties are ignored. Wynne's study deals with environmental hazards for farmers after the Chernobyl disaster. Accustomed to all the uncertainties of farming, those farmers had a flexible and adaptable way of life. The environmental experts advised them on how to deal with the possible hazards with a putative high degree of certainty, ignoring the farmers' knowledge, whereby the farmers' trust in the experts' expertise declined (Wynne 2000). In contrast, Gross discusses a redevelopment project for a former coal mining area in Germany. In this case uncertainties were seen as a normal part of the decision-making process and so it was easy for the experts to acknowledge them without losing credibility. This generated lots of space for finding the right solution for the issues involved (Gross 2010). By studying the interaction between scientists and policy makers in debates about the future of climate change, Shackley and Wynne (1996) argue that boundary work helps to establish the authority of science, despite expressing uncertain knowledge, and that it helps to create a common ground for discussing uncertainties in the science-policy domain (Shackley and Wynne 1996).

When creating trustworthy EBGs, it seems essential to accept a degree of uncertainty. Knaapen speaks of evidence-searched guidelines, as she shows how the essence of guideline development is to deal with absences of evidence (Knaapen 2013). In accepting uncertainty, the credibility of a guideline is ensured. In another study, Knaapen et al. observed a guideline development programme and concluded that strong evidence and deployed methods do not ensure the credibility of a guideline. Instead they argue: "[The guideline's] legitimacy rests on the articulation of heterogeneous types of expert knowledge and judgements, both within the guideline development group, and vis-a-vis an external world of textual documents." (Knaapen et al. 2010, 691). As we now go on to show, valuation work, or the work of giving meaning to several types of knowledge, is essential in guideline development.

Uncertainty in Evidence-Based Guidelines

Uncertainty manifests itself in three ways in an EBG. First, there is uncertainty that is inherent in knowledge. Generally, scientific articles and reports are concerned with presenting the facts and omitting all the struggles, insecurities and adaptations that were necessary to create these facts (Latour and Woolgar 1986; Star 1985; Shackley and Wynne 1996). New knowledge brings new insights, but it also brings new areas of ignorance and uncertainty to the forefront (Gross 2010; Jasanoff 2007). Guideline developers must find ways to deal with these (hidden) uncertainties and gaps in knowledge.

A second way in which EBGs are confronted with uncertainty is that they make use of heterogeneous knowledge, such as (cost) effectiveness studies, clinical trials, clinical expertise, patient experiences, often completed with ethical considerations and more. The various types of knowledge have different strengths and weaknesses. All these "knowledges" should be combined, assessed and weighed to be explicitly included or left out of the guidelines (Moreira 2005; Knaapen et al. 2010). As stated above, decision-making processes are full of uncertainties (van Asselt 2005; Wynne 1996; Jasanoff 2007). Many uncertainties must be resolved as guideline development constantly involves decisions on which practical problem to attend to, how to address the problem, which knowledge to leave in or out, and which experts to consult.

The final way in which uncertainty manifest itself in guideline development is in the translation of evidence into recommendations. Knowledge does not arrange a specific action by itself. Instead, knowledge must be actively translated to be of practical use. This work is done in guideline development, and has consequences for uncertainty.

Guideline Development as Valuation

Guideline development is a process of valuation. According to Kjellberg and Mallard (2013) valuation is a process of ordering. Guideline developers bring order into different knowledge sources and types of information. Guideline development is collective work. It is the work of classifying knowledge and giving value to this knowledge. This signification of knowledge is what happens in guideline collectives (Knaapen 2013). A multidisciplinary group of actors is involved in establishing the content of the guideline, supported by methodologists experienced in selecting evidence and writing guideline texts. The whole process of selecting a guideline development group, determining the focus, selecting and weighing the evidence, and deciding how to formulate recommendations has crucial consequences for the outcome of the guideline. Developing guidelines can take years.

Moreira observed these negotiations in a guideline development group and, based on Boltanski and Thevenot's work on justifications, distinguished four repertoires of evaluation in guideline development decision-making (Moreira 2005). These are science, practice, politics, and process. Science involves choices based upon the technical robustness of evidence, practice is about the usability of a recommendation for health care delivery, politics deals with the acceptability of recommendations for stakeholders, and process is about the way in which discussions in the guideline group are adequately represented (Moreira 2005). Moreira's work shows that these considerations engage with each other in the development of guidelines. Although it is not the aim of this article, it is likely that uncertainties play a role in such valuation work, and influence the choice of a repertoire. Knaapen argues that the core struggle of guideline development groups is to find ways to deal with the absence of knowledge. A central question that needs answering is what counts as evidence and what does not (Knaapen 2013). This discussion is the core of valuation work that emphasises signification (Kjellberg and Mallard 2013).

One way to do valuation work is to follow specific procedures for weighing and selecting knowledge. Such methods are important to give meaning to uncertainties (Knaapen et al. 2010). This article analyses some of these methods and explore how they deal with uncertainty. Specifically, we focus on the kinds of valuation work guideline developers engage in to create credible guidelines.

Research Methods

For this article, we interviewed fourteen medical guideline developers from eleven Dutch national organisations. Interviewing guideline developers gave us the opportunity to reflect on their methods and make their experiences central in the analysis. In the Netherlands, various groups and organisations, such as governmental organisations, associations for specific professionals or disease groups, and research institutes all make EBGs. The wide range of organisations involved in guideline development results in a broad variety of guidelines, both for single professional groups and multidisciplinary groups. There is no specific education for becoming a guideline developer in the Netherlands. Instead, guideline developers have different backgrounds. There are epidemiologists, healthcare practitioners with degrees in education, health scientists, and quality managers. Combinations are possible, such as medical doctor/epidemiologist. All the interviewed guideline developers have at least ten years' experience in developing guidelines. One guideline developer has been in the field for over 20 years. For some in this group, developing guidelines is their core task, whereas others combine it with other part-time work, such as being a practicing physician. We chose this wide selection of respondents as we believed the breadth would bring deeper insights into what happens to uncertainty in the development of EBGs.

The respondents were asked how they developed guidelines, which problems and uncertainties they encountered, and how they dealt with these situations. Colleagues from the institute of Health Policy and Management conducted half of the interviews, in relation to another project on guideline development (Zuiderent-Jerak et al. 2011). All interviews were recorded and transcribed verbatim. The results were analysed both inductively and deductively, in the latter case with a focus on ways of dealing with uncertainty. The empirical section starts with an explanation of guideline development, and then discusses the relation between classification systems and alternative methods for guideline development. We go on to explore the relation between ignorance and guideline development. This empirical section ends with an analysis on how credibility is accomplished in guidelines.

Guideline Development

In this section, we outline the guideline development process, as described by the guideline developers we spoke with. According to our respondents, their procedures are very similar to what is known from the literature (Knaapen et al. 2010; Moreira 2005), although there are differences between different Dutch guideline organisations.

Guideline development starts when there is a reason to develop a guideline for a certain problem. Reasons vary. At the start of the evidence-based medicine movement in the Netherlands, resolving uncertainty in medical practice was the reason to develop a guideline. A guideline developer involved in the field for some twenty years, provides an example:

> The guideline on oral contraception, the pill, was about abolishing check-ups for the pill. In those days, we still had pill check-ups and all women on the pill had to see the doctor twice a year for a smear test. The pill was first perceived as a

> risky thing, which needed to be examined regularly. Over time people started doubting the effectiveness of these check-ups, but how do you organise a stop to this? (Guideline developer/general practitioner involved in guidelines for general practitioners)

These first guidelines were developed to solve uncertainties in medical practice and/or reduce ignorance, according to our respondents. Over time, when the most striking problems had been addressed, the reasons for developing guidelines changed. Gradually guidelines became repositories of how medical work should be done. The same guideline developer remarks:

> Then the question for developing a guideline changed into 'What do guidelines lack? What common problem should we tackle next?' So that raises the question of what we want to achieve with these guidelines. Do we want to describe the entire medical terrain? Then it becomes a sort of handbook. Or do we focus on situations where something is going on, where doctors don't know what to do? (Guideline developer/general practitioner involved in guidelines for general practitioners)

Notably, most guideline developers criticise the idea of making guidelines for situations *without* uncertainties. This does not always mean that no guidelines are made. Interestingly, though, "good" guidelines, according to guideline developers, seem to include some degree of uncertainty; otherwise, the need for a guideline is questioned.

Reasons for developing guidelines change over time, according to our respondents. Sometimes, any new situation determines the need for a guideline. In infectious diseases, every new possible outbreak of a disease is a reason to develop a guideline. A consistent approach towards infectious diseases is essential to tackle the situation and guidelines are the way to reach the healthcare workers involved. Other guideline developers noted that the need for a guideline is determined on the basis of explicit criteria, including the prevalence of the problem, potentially achievable health benefits, solving controversies in practice, satisfying demands from professionals or patient groups and the availability of (at least some) evidence for the problem. These criteria help guideline developers to select relevant topics or to justify to others that such a topic is suitable for a guideline. In contrast, justifying that a topic is not suitable also occurs:

> We must be able to say this is not a subject for a guideline. For example, the geriatric society consulted us for a guideline on medical care for frail elderly on psychiatric wards. This could be a guideline topic. But when we investigated the source of the problem, we discovered that those geriatric beds in many psychiatric hospitals were under pressure due to financial problems. This affected the position of the geriatric doctors. How the medical care was to be given was not the question. Then you should rethink if this is a guideline topic. (Guideline developer/epidemiologist involved in clinical guidelines)

After selecting the topic, guideline developers establish the starting questions of the guideline. These are generally based upon the struggles, uncertainties, or bottlenecks in healthcare practice that are identified by consulting actors in the healthcare field. Who is consulted differs. Most often healthcare workers directly involved in the issue are asked, but for more complex or controversial issues, some guideline development organisations ask a broader range of stakeholders:

> In the guideline we made for intensive care we not only included practitioners, but also health insurers, academic hospitals, the local hospitals, the health inspectorate, health care spokespersons for political parties. We consulted everyone prior to developing the guideline, and asked what we should include, so that we knew what subjects to address and why. (Guideline developer/ epidemiologist involved in clinical guidelines)

Such an approach aims to ensure that most of the relevant issues are known up front, so that further development does not meet too many surprises.

After defining the starting questions, the core of the work of guideline development starts. This includes systematic searching, assessing, and selecting relevant knowledge, and translating various "knowledges" into guideline recommendations. Knowledge comes from scientific publications, reports and documents, international guidelines on the topic, experiences, and expertise, and also often from systematic reviews made, for example, by the Cochrane collaboration or the National Health Institute. The latter type helps translate large amounts of literature and makes it easier to apply in decision-making (Chalmers 1993). However, reviews still need valuation processes to be applicable in guideline development:

> Most of the Dutch guidelines are developed from scratch. We call it "de novo." Of course, we make use of international guidelines and reviews by, for example, the IHI or National Health Institute. They make good evidence reviews, which are also published in the literature. But this knowledge is not always applicable for the guideline we intend to make. So this kind of knowledge has limited use. (Guideline developer involved in GRADE working group)

Any kind of knowledge needs to be assessed for a guideline. This is done in guideline development groups and by guideline methodologists. Guideline development groups, consisting of various representatives with specific expertise and involvement in the issue, discuss the selected knowledge, judge its relevance, check its robustness, and deal with and (at times) resolve any omissions in the knowledge. This valuation work can take months or even years. The guideline drafts are the main focus of the debate. When the guideline is eventually finalised, it is introduced in healthcare practice. Often guideline development organisations have an infrastructure for implementation, such as websites, periodical publication of a book containing all guidelines, and a network of healthcare practitioners.

Classification Systems: A Curse or a Blessing for Accepting Uncertainty?

The core of the work of guideline developers is classification or ordering of knowledge, often done with classification systems or levels of evidence tables (Gugiu and Ristei Gugiu 2010; Knaapen et al. 2010). These frequently used methods are often criticised by guideline developers. Evidence tables have different levels but their hierarchy is predominantly based on study designs, with level 1 on top and level 4 or 5 on bottom. In such tables, meta-analysis of randomised clinical trials (RCTs) are on top and patients' and practitioners' experiences are considered the least form of evidence. Classification systems help demarcate between "stronger" evidence and more "anecdotal" evidence, as they enable guideline developers to indicate with how much certainty a claim is made. The strength of evidence is made transparent. The levels are a means to accept uncertainty, as they allow demarcating between more and less certain claims. However, levels of evidence tables only help deal with the uncertainty inherent in knowledge (i.e. the first kind of uncertainty discussed earlier). Uncertainties in knowledge valuation and uncertainties in knowledge translation are not resolved with levels of evidence tables. The following two examples clarify our point.

First, classification systems are based upon study design. Strong study designs such as meta-analysis or RCTs tell something about the robustness of the evidence supporting a claim. However, they do not say anything about the *quality* of knowledge for making recommendations in a particular guideline. One guideline developer expressed this as follows:

> If you want to compare two pills, then you use a RCT, if you want to know how to best organise care for a specific group of patients then you might use a qualitative research design. Depending on the purpose of the guideline different knowledge is seen as hard evidence. If you use the same classification schemes for both kinds of research, then the qualitative research is valued less and you might make recommendations that are less firm. Well, as guideline developers we need to pay more attention to these things. (Guideline developer/epidemiologist involved in clinical guidelines)

What knowledge should be rated higher or lower in the hierarchy depends on which question the guideline aims to answer. Levels of evidence tables do not allow for such specificity. The valuation of the quality of knowledge remains the work of the guideline development group.

Secondly, classification systems cannot deal with omissions in knowledge. They can only categorise available knowledge; unknowns cannot be valued in the levels. One guideline developer referred to this problem in terms of "thoughtless empiricism":

> On the one hand, if there is no trial, then you can say there is no evidence, nothing has been proven. If you are really strict, this means that you can no longer treat numerous patient groups. For example, if you look at osteoporosis, you see that almost all the trials have been done amongst women. So, what to do with men? Well, you could argue that it would work somewhat similar with men, and you could just give them the same pills. You could also argue that nothing has been proven for men, so you stop [prescribing]… But on the other hand, we sometimes face this [situation] if you are too restrictive. For example with heart attacks, certain medications are recommended, especially for the first six months. There are about six pills on the market but only two have been studied in decent trials. Should we then say use only those two, and not the others? What complicates the matter is that this treatment is prescribed for both diabetes and heart failure, and maybe other pills are being studied. So, you see, it's always a diffuse thing. It's what I call thoughtless empiricism; it depends strongly on what study has been done. We definitely need to find compromises. (Guideline developer/general practitioner involved in guidelines for general practitioners)

This quote shows, that omissions in knowledge need to be dealt with. Ignoring these unknowns leads to all sorts of partial recommendations, while the question is how to include the omissions. Classification systems focus only on available knowledge, so as the above guideline developer remarked, compromises are needed to solve these situations.

To sum up, classification systems are an aspect of valuation as they assist guideline developers to classify knowledge based upon study design and source of knowledge. While they rate knowledge, they do not tell anything about its quality. One guideline developer noted:

> Levels of evidence are like the star rating of a restaurant, but you only find out what a restaurant is like when you go and eat there. So the justification is more important than the rating. (Guideline developer at Dutch College of General Practitioners)

In short, classification systems can help to categorise knowledge in more or less proven claims. But that is all they do. They can neither deal with unknowns nor take the relevance of the knowledge to a particular context into account. Classification systems need other valuation practices, such as consensus-making amongst experts, to interpret the meaning of the classification. Classification systems therefore provide only modest assistance in dealing with uncertainty.

Grading Types of Knowledge

An alternative method that many of the interviewed guideline developers mentioned is GRADE, the Grading of Recommendations Assessment, Development, and Evaluation. Responding to some of the criticism of classification systems, the international GRADE working group has come up with a systematic approach to rate heterogeneous types of knowledge, which is based on more criteria than study design alone. In terms of Moreira's repertoires, GRADE offers a legitimate way to include more of the repertoires of practice, politics, and process, instead of only science (Moreira 2005). A guideline developer with experience in using GRADE explains:

> The advantage is that you can select on subjects that are clinically relevant. You look at results and not the study design. (Guideline developer/policy adviser involved in guidelines for elderly care physicians)

The GRADE method involves five factors that downgrade and three factors that upgrade the quality of evidence (Guyatt et al. 2011). Such factors as "inconsistency" and "indirectness" lower the quality and "large effects" increases the quality. By including more relevant factors in the decision-making process, GRADE tries to suit the valuation processes better. One guideline developer involved in the international GRADE working group explains:

> GRADE is a real step forward, but one of the consequences is that the strength of recommendations generally decreases. There are more considerations to take into account, and they generally turn out to give a lower recommendation. (Guideline developer involved in GRADE working group)

GRADE tries to give more space for valuating "other" (i.e. not considered hard evidence) knowledge, and for expressing uncertainties. By taking more aspects into account, GRADE offers more opportunities to deal with unknowns and uncertainties in guideline development. Guideline developers involved in guidelines for the frail elderly explained that especially in the case of ignorance and uncertainty, this method had advantages:

> Well, the point is that relatively little research is done on the frail elderly. Often there are no RCTs available. So you search for alternatives to find evidence that is clinically relevant for this group. (Guideline developer/policy adviser involved in guidelines for elderly care physicians)

Generally, if studies match only partially with the focus of the guideline, the strength of the recommendations decreases with GRADE. However, in some cases, when a lot of risk is involved, the strength of recommendations can increase. For example:

> One of the best is the WHO guideline on avian flu. It's good as it specifies the considerations and choices. But, if you look at the proof for the advice you can see many unknowns. One factor that influenced their decisions was the considerable risk of disaster, with high mortality and morbidity. This risk and probable low side-effects made the recommendations strong, although there was only indirect evidence. (Guideline developer in GRADE working group)

By including other and more criteria for weighing knowledge than just study design, GRADE brings a broader ground for valuating knowledge. GRADE seems to support decision-making involving uncertainty in valuation and uncertainty in the translation of knowledge into recommendations, while allowing for the uncertainties inherent in knowledge to be addressed. However, at the time of the interviews, most guideline developers had no or only limited experience in using GRADE. Some guideline developers expected GRADE to make their work more complex, as the more formal valuation procedures would make decision-making more technical and time consuming. We have yet to see what these reservations mean to the use of GRADE and its credibility in healthcare practice.

Involving Expertise from the Healthcare Field

> One issue in guideline development is that you can't solve every question with evidence. If we are too strict, there will be hardly anything left in the guideline, especially since we focus on nursing care for the elderly. (Guideline developer involved in guidelines for nurses and geriatric assistants)

Guidelines cannot be made without experiential knowledge; i.e. the knowledge of healthcare practitioners and patients in the healthcare field. However, as we have discussed above, this most anecdotal kind of knowledge forms the bottom level of the evidence system. It risks being seen as individualised information, which is difficult to make relevant to the guideline. This section explores how such knowledge is used and what happens with uncertainty.

Including the expertise of healthcare professionals and patients is assumed to have several benefits, as it brings different information about healthcare delivery to the fore. For example, one of the epidemiologists developing clinical guidelines remarks:

> Surgeons and orthopaedists have different policies on anti-coagulants for some conditions. They argue that the guidelines don't need to mention this, as they agree to disagree on this point. Yet, a focus group revealed that patients in a shared room find it troublesome to be getting different treatment for the same complication. (Guideline developer/epidemiologist involved in clinical guidelines)

Such experiences are important to include in a guideline. Patients' and healthcare practitioners' knowledge not only fills in important unknowns, it also explores whether guideline recommendations are feasible and accepted.

But how should this knowledge be included in guidelines? Guideline developers have little experience with methods for including experiential knowledge. Some guideline developers have used Delphi-like methods, but regard them as time consuming and expensive. One guideline developer refers to experiential knowledge as "impressionistic":

> It's like you say something, I say something and we put it together, but it's not systematic. (Guideline developer at Dutch Institute for Healthcare Improvement)

Interestingly, while guideline developers are highly systematic when it comes to knowledge assessment in general, they tend to be less systematic when it involves including more experiential knowledge (Zuiderent-Jerak, Forland, and Macbeth 2012). So how do guideline developers ensure that experiential knowledge is not too anecdotal? Generally, they rely on a large number of (patient) representatives:

> If there is a good patient-representing association we will contact it. They have investigated their members' demands and know what they want. Otherwise we often use focus groups of patients. If, for example, I make a guideline for emergency surgery, well there isn't a patient association for that, so then we'd consult a focus group. But we should evaluate if this is the best approach although I don't know how we could do it differently. (Guideline developer/epidemiologist involved in clinical guidelines)

The interviews revealed numerous cases of a request for a guideline, despite the absence of knowledge. As discussed above, uncertainties are often the reason to start developing a guideline. One example comes from the guidelines developed in youth health care:

> Very often there is no literature on our subjects, since we work in preventive care. It's on a different level. For example, we deal with screening programmes, how to screen for children that fall behind or don't function well. Well, you don't find this directly in the literature. […] So a huge part of our guidelines is practice- or expert-based. That's justified by grey literature, handbooks, expert opinions, focus groups etcetera. (Guideline developer/physician involved in guidelines for youth health care)

Another telling example is guidelines for new infectious diseases, made by a governmental organisation for infection prevention. With an outbreak of a new infectious disease (or the threat of one), such as the swine flu pandemic or SARS, there is a lot of uncertainty due to both ignorance and public reactions. A developer of the swine flu guideline explains:

> In the beginning we knew nothing. Something started in Mexico, but if and how it would affect us in the Netherlands was unknown. Our boss explained that it was severe in Mexico. The Spanish flu used to be severe as well, and that was our only frame of reference. (Guideline developer involved in guidelines for infectious diseases)

In the absence of knowledge and in the presence of the risk of an outbreak, guideline development becomes a delicate situation. The public is highly involved in this situation, and may reaction with fear, indifference, and criticism:

> We got a lot of flak, as if we were taking it [i.e. reaction to a possible swine flu pandemic] out of proportion out of our own interests, since people suspected us of having stakes in the vaccine industry. Based on this criticism, you'd think that people would refuse the vaccine because, they argued, we made a problem out of nothing. But, people did take the vaccine, despite the fact that they also thought we made a big fuss about it. (Guideline developer involved in guidelines for infectious diseases)

The study by Gross point out that in situations of ignorance, communication of uncertainty is accepted (Gross 2010), but here the expertise of the governmental organisation was questioned and criticised. The guideline developer explained that they felt that acknowledging uncertainty was not an option, as there was a lot of pressure on them to come up with "an answer." She reflects:

> We concluded that maybe we should say explicitly that we don't know either. But people assume they'll get an answer from us. So we're almost forced to say something. And if we don't know either, then what should we do? Then we say "take all possible measures." It is actually impossible if you think about it. (Guideline developer involved in guidelines for infectious diseases)

In the absence of knowledge on the infectious agent or possible remedies, the governmental organisation for infectious diseases follows another approach to develop their guidelines. Especially with novel infectious diseases there is often a lack of knowledge on the disease as it is too new. Therefore guideline developers include the literature on viruses that look similar and—until more knowledge becomes available—they adapt the interventions suggested to deal with similar viruses. Of course, for "older" infectious diseases, such as hepatitis, rabies, or measles, specific literature is more widely available. Besides this literature search, experts and healthcare professionals in the Community Health Services[1] are intensively involved in guideline development. An external expert (a medical specialist, biologist, or virologist) is consulted to write the text and the texts are subsequently discussed with fifty representatives, one from each Community Health Service. The group reflects on all the comments and the result is the definitive guideline recommendations.

This organisation of guideline development ensures that experiential knowledge becomes known and can be included at a relatively early stage. After the guideline is finished and published, the governmental organisation encourages feedback. Guideline users can report all their new knowledge and experiences of using the guideline

[1] Dutch Public health policy is executed by regional Community Health Services serving a varying number of municipalities.

on a special 24/7 telephone service. This feedback not only enables the guideline developers to adjust their advice, but at the same time informs them about new knowledge and the practical usefulness of their recommendations. If they conclude, from this information, that the guideline should be changed, then this is done immediately. Acknowledging uncertainty thereby becomes an open and collective effort between guideline developers and practitioners. It is achieved by creating feedback moments, when the comments and experiences of users can be inserted in the guideline, even after its publication. The highly interactive process shows how uncertainty is fully integrated into the process of making guidelines. This approach not only improves development and the fine-tuning after publication, it also deals with uncertainties involving the implementation and use of guidelines. Feedback brings important insights into how the guideline is used and interpreted.

To sum up, guideline developers are very aware that they need experiential knowledge from healthcare practitioners, patients and specialised experts to create guidelines. There are, however, still great challenges in including this knowledge in ways that move beyond the overly "impressionistic." The feedback system used in guidelines for infectious diseases is a promising example of how uncertainties can be addressed collectively.

Ensuring Credibility of Guidelines

A core concern of developers is how their guidelines are received and used in healthcare practice. How can guidelines remain credible and express uncertainty at the same time? The "evidence-based" label gives the impression that evidence makes guidelines credible. However, as Knaapen argues, evidence-based medicine is more often about how to deal with the absence of evidence (Knaapen 2013). When asking guideline developers what "evidence-based" means, they answered that it deals congruently with working *systematically* and *transparently*:

> For me, a guideline is evidence-based when we have followed the process. So, when you define the focus and the limits at the start, and then you search the literature systematically, in all the databases. Evidence-based is when you select and assess the literature systematically, so that you come to a systematic conclusion. (Guideline developer/epidemiologist involved in clinical guidelines)

And another guideline developer explains:

> An evidence-based guideline is one where you can see if each recommendation is based on consensus or the literature. You can see that the literature has been searched in depth, so you can repeat a search. And you can see the justification for the recommendation, like "Jansen says this, Pietersen says that, and we chose this because..." (Guideline developer involved in guidelines for infectious diseases)

Evidence-based does not refer to the strength of the evidence found, but to the process of making guidelines (see also Knaapen 2013). In terms of uncertainty, the procedures for making EBGs involve systematic searches to ensure that there is indeed evidence, and if not found, that there is "truly" no evidence (Knaapen 2013). In other words, doing things systematically and transparently ensures the "evidence-basedness." As earlier work on guideline development has shown, guideline credibility is determined by the inclusion of a diversity of knowledge sources and comparisons to similar reports and documents. At times, therefore, we found that strong evidence is presented with softening nuances, otherwise it would reduce the credibility of the guideline. For example:

> I was involved in a guideline on sedation policy. There was very strong evidence that it's good to have an extra professional monitoring a patient during sedation. But we don't have these professionals and it involves training. It's unclear who should pay and how many of these professionals are needed. So it's worthwhile knowing this all, but to keep the actual recommendations a bit loose. Otherwise it leads to all kinds of problems in acceptance of the guideline. This then affects the trust in the whole guideline, not just this recommendation alone. (Guideline developer/epidemiologist involved in clinical guidelines)

Despite the strong evidence, the guideline developers chose to soften the recommendation a bit, since recommending unfeasible things can affect the acceptance of the whole guideline. In contrast to Wynne's sheep farmers, the situation here shows that uncertainties in practice are not ignored, but form a part of the rationale for deciding which evidence to include and how to include and present it.

One way to ensure credibility is to use a systematic evidence-based working method:

> We often get attacked for the recommendations we make. As a governmental organisation, we're under attack anyhow. That's why we need to make evidence-based guidelines. If we can't make well-founded statements, based on good knowledge, we're in trouble. We are very conscious of that. (Guideline developer involved in guidelines for infectious diseases)

A systematic evidence-based working method legitimises the credibility of the governmental organisation in making their guidelines. False certainty or certainification does not take place, according to our respondents.

Trust and credibility affect decision-making in the development of guidelines and are one of the many considerations that must be taken into account. Guideline development seems a practice that inherently addresses uncertainty, and therefore does not run into credibility issues, as Wynne describes. Instead, as we have tried to show in the empirical sections, guideline making is reflexive work that seek optimal ways to reflect what is known and what is uncertain, and to do this in such way that it can retain credibility and guide healthcare practice.

Conclusion

This article explored the valuation work that guideline developers undertake to develop EBGs and how uncertainty is addressed in the process. We distinguished three valuation practices, based on empirical findings: classification of studies, grading types of knowledge, and those involving expertise and clinical practice. These three valuation practices differed in the types and amount of uncertainty they could endorse. Classification studies seem helpful for guideline developers in dealing with uncertainties inherent in knowledge, but cannot deal with ignorance and do not help to relate knowledge to a particular context. Thus, guideline developers need other valuation practices to interpret and include knowledge than solely classification systems. Grading different types of knowledge is, in the guideline developers' view, slightly better equipped to assist in valuation practices and to live with uncertainties. GRADE seems to better allow one to include various kinds of uncertainty and provides a ground for legitimising the choices made in the guideline development process. Involving expertise and practice endorses all three types of uncertainty, but risks being too anecdotal.

The type of valuation practice has consequences for the outcome; some types are better capable of accepting uncertainty than others. What seems essential is that the valuation practices that work better seem better capable of including various kinds of uncertainty and provide the grounds to legitimately justify the choices made in the decision-making process. This combination—allowing for uncertainty and yet being able to justify choices made through some form of systematic way of working—enabled guideline developers to deal with uncertainty.

The reflexive aspects of valuation work are particularly interesting. Valuation work in guideline development not only involves input (assessment of knowledge) but also the output (how users perceive the result). A telling example is the case of guideline making for infectious diseases. Feedback from users helped the developers improve the guideline and gain insight into how the guideline was used. A feedback system is likely to prevent some of the uncertainties that tend to occur in guideline implementation, such as uncertainties in the uptake of recommendations and the spread of the guideline.

The question how to remain credible can be solved by including heterogeneous types of knowledge (Knaapen et al. 2010). Wynne's study showed that ignoring fundamental aspects of knowledge (sheep farmers' local knowledge) leads to distrust and unrest (Wynne 2000). The challenge for guideline developers is thus to include relevant knowledge from various sources and of different strengths, and doing this systematically and transparently. Justifying choices is essential and guideline development methods seem to offer a formal way to do this justification.

Our data showed that guideline development seems to be most systematic with knowledge that is more certain, and least systematic when knowledge is less certain. That is, knowledge stemming from patient experiences and expertise of professionals is generally not collected and included following a systematic approach, but as one guideline developer argued, it is "impressionistic." The knowledge that is most uncertain, in relation to its external validity, is included least systematically. How to approach this situation is one of the challenges for the future of guideline development.

For this article we interviewed guideline developers. We selected guideline developers working for different organisations and with different personal backgrounds. The benefit of this choice is that we could explore a broad range of valuation practices, and also see which elements of the evidence-based approach were common in all the different places. Dutch guideline development is likely to be done differently than in other countries, and this should be taken into consideration interpreting these results. We relied on the interviews as our main research method. Observation of guideline-making practices might produce different findings.

In studying valuation practices in guideline development we found that uncertainty is in many ways inherent and is essential to create EBGs. We conclude that guideline developers use different valuation practices to deal with this inherent tension in their work and these practices have different consequences for the types of uncertainties that can be taken on board. Studying guideline development as valuation work enabled us to move beyond a more rational investigation of classification of knowledge. Instead valuation serves as a valuable notion to study how heterogeneous and divergent knowledge can be connected, and how and where uncertainties are acknowledged.

Acknowledgements. We would like to thank the respondents for the open conversations and the valuable insights into their work. Thanks to Teun Zuiderent-Jerak, Sonja Jerak-Zuiderent and Hester van de Bovenkamp for sharing some of their data on guideline development with us. Finally, we would like to thank Marjolein van Asselt, Charles Bosk, Alena Bleichner and colleagues of the Health Care Governance group, especially Teun Zuiderent-Jerak, for comments on earlier version of this article.

References

Callon, Michel, Pierre Lascoumes, and Yannick Barthe. 2009. *Acting in an Uncertain World*, translated by Graham Burchell. Cambridge, Mass.: The MIT Press.

Chalmers, Iain. 1993. "The Cochrane Collaboration: Preparing, Maintaining, and Disseminating Systematic Reviews of the Effects of Health Care." *Annals of the New York Academy of Sciences* 703 (1): 156–165.

Graham, Robin, Michelle Mancher, Dianne Miller Wolman, Sheldon Greenfield, and Earl Steinberg. 2011. *Clinical Practice Guidelines We Can Trust*. Institute of Medicine. Washington, D.C.: The National Academies Press.

Gross, Matthias. 2010. *Ignorance and Surprise*. Cambridge, Mass.: The MIT Press.

Gugiu, Cristian P., and Mihaiela Ristei Gugiu. 2010. "A Critical Appraisal of Standard Guidelines for Grading Levels of Evidence." *Evaluation & the Health Professions* 33 (3): 233–255.

Guyatt, Gordon, Andrew D. Oxman, Elie A. Akl, Regina Kunz, Gunn Vist, Jan Brozek, Susan Norris, Yngve Falck-Ytter, Paul Glasziou, Hans deBeer, Roman Jaeschke, David Rind, Joerg Meerpohl, Philipp Dahm, and Holger J. Schünemann. 2011. "GRADE Guidelines: 1. Introduction-GRADE Evidence Profiles and Summary of Findings Tables." *Journal of Clinical Epidemiology* 64 (4): 383–394.

Helgesson, Claes-Fredrik, and Fabian Muniesa. 2013. "For What It's Worth: An Introduction to Valuation Studies." *Valuation Studies* 1 (1): 1–10.

Jasanoff, Sheila. 2007. "Technologies of Humility." *Nature* 450 (1): 33–33.

Jerak-Zuiderent, Sonja. 2012. "Certain Uncertainties: Modes of Patient Safety in Healthcare." *Social Studies of Science* 42 (5): 732–752.

Kjellberg, Hans, and Alexandre Mallard. 2013. "Valuation Studies? Our Collective Two Cents." *Valuation Studies* 1 (1): 11–30.

Knaapen, Loes. 2013. "Being Evidence-Based in the Absence of Evidence: The Management of Non-Evidence in Guideline Development." *Social Studies of Science* 43 (5): 681–706.

Knaapen, Loes, Hervé Cazeneuve, Alberto Cambrosio, Patrick Castel, and Beatrice Fervers. 2010. "Pragmatic Evidence and Textual Arrangements: A Case Study of French Clinical Cancer Guidelines." *Social Science & Medicine* 71 (4): 685–692.

Latour, Bruno, and Steve Woolgar. 1986. *Laboratory Life: The Construction of Scientific Facts*. Princeton: Princeton University Press.

Melse, Johan. 2003. "Van Onzekerheid Weg Denken." In *Niet Bang Voor Onzekerheid*, edited by Marjolein van Asselt and Arthur Petersen, 61–72. Utrecht: Lemma.

Mesman, Jessica. 2008. *Uncertainty in Medical Innovation: Experienced Pioneers in Neonatal Care*. Basingstoke: Palgrave Macmillan.

Moreira, Tiago. 2005. "Diversity in Clinical Guidelines: The Role of Repertoires of Evaluation." *Social Science & Medicine* 60 (9): 1975–1985.

———. 2011. "Health Care Rationing in an Age of Uncertainty: A Conceptual Model." *Social Science & Medicine* 72 (8): 1333–1341.

Moreira, Tiago, Carl May, and John Bond. 2009. "Regulatory Objectivity in Action: Mild Cognitive Impairment and the Collective Production of Uncertainty." *Social Studies of Science* 39 (5): 665–690.

Shackley, Simon, and Brian Wynne. 1996. "Representing Uncertainty in Global Climate Change Science and Policy: Boundary-Ordering Devices and Authority." *Science, Technology & Human Values* 21 (3): 275–302.

Star, Susan Leigh. 1985. "Scientific Work and Uncertainty." *Social Studies of Science* 15 (3): 391–427.

Timmermans, Stefan, and Alison Angell. 2001. "Evidence-Based Medicine, Clinical Uncertainty, and Learning to Doctor." *Journal of Health and Social Behaviour* 42 (4): 342–359.

van Asselt, Marjolein B.A. 2005. "The Complex Significance of Uncertainty in a Risk Era: Logics, Manners and Strategies in Use." *International Journal of Risk Assessment and Management* 5 (2–4): 125–158.

van Asselt, Marjolein B.A., Jessica Mesman, and Susan A. van 't Klooster. 2007. "Dealing with Prognostic Uncertainty." *Futures* 39 (6): 669–684.

van Loon, Esther, and Teun Zuiderent-Jerak. 2011. "Framing Reflexivity in Quality Improvement Devices in the Care for Older People." *Health Care Analysis* 20 (2): 119–138.

van Loon, Esther, Teun Zuiderent-Jerak, and Roland Bal. 2013. "Diagnostic Work Through Evidence-Based Guidelines: Avoiding Gaps Between Development and Implementation of a Guideline for Problem Behaviour in Elderly Care." *Science as Culture* 23 (2): 153–176.

Wynne, Brian. 1996. "May the Sheep Safely Graze? A Reflexive View of the Expert-Lay Knowledge Divide." In *Risk, Environment and Modernity: Towards a New Ecology*, edited by Scott Lash, Bronislaw Szerszynski, and Brian Wynne, 44–83. London: Sage Publications.

———. 2000. "Misunderstood Misunderstanding: Social Identities and Public Uptake of Science." *Public Understanding of Science* 1 (3): 281–304.

Zuiderent-Jerak, Teun, Frode Forland, and Fergus Macbeth. 2012. "Guidelines Should Reflect All Knowledge, Not Just Clinical Trials." *BMJ* Vol. 345, Article Number e6702.

Zuiderent-Jerak, Teun, Sonja Jerak-Zuiderent, Hester van de Bovenkamp, Siok Swan Tan, Leona Hakkaart-van Roijen, Werner Brouwer, and Roland Bal. 2011. *Variatie in Richtlijnen: Wat Is Het Probleem?*. Instituut Beleid en Management Gezondheidszorg. Erasmus Universiteit: Rotterdam.

Esther van Loon is a PhD candidate in the Healthcare Governance group at Erasmus University Rotterdam, the Netherlands. She has a background in nursing and health sciences. Her research focuses on innovation practices in long-term healthcare and the development and implementation of standards and guidelines. She is especially interested in understanding the effects of the interaction between development and use of healthcare innovations and technologies. She has previously published in *Health Care Analysis* and *Science as Culture*.

Roland Bal is Professor of Healthcare Governance at Erasmus University Rotterdam. His research interests include science, policy, and the law and governance infrastructures in health care. With Wiebe Bijker and Ruud Hendriks, he published *Paradox of Scientific Authority* (MIT Press), on the role of science advisory councils in (health) policy. More recently, he has studied the creation of public accountability in healthcare, studying ways in which public services 'organise for transparency'. Recent publications have appeared in, amongst other journals *Social Science & Medicine*; *Science, Technology & Human Values*; and *Public Administration*.

Valuation Studies 2(1) 2014: 65–70

Review note

Extending the Cosmopolitical Right to Non-Humans

Helen Verran

Abstract

This short essay is a review of Bruno Latour's *An Inquiry into Modes of Existence: An Anthropology of the Moderns* (Harvard University Press, 2013) and a commentary on the wider move that accompanies the book.

Should readers of *Valuation Studies* bother themselves with the latest intellectual fad emerging from the Left Bank in Paris? I will suggest there is at least one good reason to do so. *An Inquiry into Modes of Existence*, the latest modernizing intervention to emerge from Paris' 6th and 7th arrondissements, a hotspot of modernizing activity in the past, has valuation in its sights. The very place that gave us the modern valuation regime of 'rational metrication' is at this very moment planning to set a few depth charges; planning how to blow things apart in a significant section of the world of valuation, in order to start again. This time the valuation regimes in question must be designed to avoid the epistemological mistake of believing and acting as if valuations associated with economization can be made referential in the way say cartographic valuations can (with great difficulty) be made and maintained as referential. Before I consider this foreshadowed intervention however, I attempt to introduce *An Inquiry into Modes of Existence*—a sprawling endeavour of an interactive website understood as a multimedia platform, an impressive social marketing campaign, and a book (a fat advertising pamphlet?).[1] As you might gather from my title, I set Latour's injunction that we should re-

Helen Verran, School of Historical and Philosophical Sciences, University of Melbourne, Australia, hrv@unimelb.edu.au

© 2014 Helen Verran
LiU Electronic Press, DOI 10.3384/vs.2001-5992.142165
http://valuationstudies.liu.se

institute modernity as cognizant of its core values (which is why valuations associated with economization need to be blown apart), alongside a proposal on modernity promulgated from Prussia's Königsberg in 1795, in the form of Kant's cosmopolitical right. I will develop that analogy in concluding my review.

It is a mistake to think of *An Inquiry into Modes of Existence* as just Latour's latest and most fantastic intervention. It is that nevertheless, a continuation of the trajectory of Latour's idiosyncratic attempt to characterize modernity and thus diagnose the roots of the planet's ecological crisis. This might be seen as beginning with the polemic *We Have Never Been Modern* (Latour 1993), continued through *The Politics of Nature* (Latour 2004). However, since in common with *Inquiry into Modes of Existence*, these books are characterized by Latour's resolute refusal of modernity's reductionism, it is useful to understand "Irreductions" (Latour 1988) as the source. Seen from this starting point the trajectory incorporates Latour's inveighing against the 'iconoclashes' of modern critique (the upshot of competing reductions vying for dominance) in *Iconoclash* (Latour and Weibel 2002), and *The Modern Cult of the Factish Gods* (Latour 2010). Latour has stayed loyal to the brand of metaphysics he began with. But *An Inquiry into Modes of Existence* is not only a book: the interactive website (www.modesofexistence.org), which a cynic might claim is just a means of crowd sourcing content, and a travelling roadshow with Bruno as its star, are just as integral to the project. The book has no index, but the multimedia platform (the "extended book"), which encourages contributions from fellow "anthropologists of modernity," incorporates (a perhaps overwhelming, and often rather slow) searchability. I found that neither of these elements were satisfying to use on their own, but that they work together well. Together the elements of *An Inquiry into Modes of Existence* exemplify two important modes of existence—the networking mode of existence, and the position-taking mode (which here is called the prepositional mode), and also offer some sort of culmination of Latour's analytic journey.

So how to characterize this *Inquiry*? Here's how Latour explains its purpose:

> [Through this enquiry] I am exploring . . . a series of contrasts to distinguish the values that [Modern] people are seeking to defend from the account that has been given of them throughout history, so as to attempt to establish these values, or better yet to install them, in institutions that might finally be designed for them. (Latour 2013, 7)

How is the "series of contrasts" presented? I read *An Inquiry into Modes of Existence* as a blue-print for a working a machine of the sort loved by semioticians; a machine that in its workings reveals "useful contrasts." In developing this machine the hope is to sensitise Moderns who engage with it to clashes of values that, over several hundred years, they have learned to simultaneously experience and not experience. It is hoped that through taking to this machine many Moderns will, perhaps for the first time, articulate their disconcertment at such clashes of values. Thus the website is an attempt to crowd source stories of disconcertment. The grand end-point of this cycle of reform will perhaps set about the task of re-establishing institutions, and begin the next cycle. This is an ambitious and long-term project.

The machine, which both seeks to sensitise Moderns to clashes of values, and to support the articulation of stories of disconcertment, takes some time getting into however; much to and fro-ing from the book to the website is needed. Like all good semiotic machines it's clunky and exaggerated. But then its ancestors were perhaps even more preposterous. I speculate that its parents are Donna Haraway's "Four Square Cyborg," and A.J. Greimas' "infamous semiotic square . . . a clackety, structuralist meaning making machine." Here is how Haraway introduces her line of "artificial devices that generate meanings very noisily"—Haraway could be understood as foreshadowing the semiotic machine of the *Inquiry*:

> To get through the artifactual to elsewhere, it . . . help[s] to have a little travel machine that also functions as a map . . . [,] a structuralist engine put to amodern purposes . . . [,] a view of the history of science as a culture that insists on the absence of beginnings, enlightenments, and endings: the world has always been in the middle of things in unruly and practical conversation, full of action and structured by a startling array of actants and of networking and unequal collectives . . . [.] [This] amodern history will have a different geometry, not of progress, but of permanent and multi-patterned interaction through which lives and worlds get built, human and unhuman. (Haraway 1992, 304)

The trajectory of Latour's analytic journey, and a semiotic machine of the sort favoured by Haraway? All this probably seems quite distant from the interests of those studying modernity's valuation regimes. Let's take a closer look at how the machine of the *Inquiry* is currently working in the academy.

Calculemus!

Part of wonder of *An Inquiry into Modes of Existence* is its generosity in inviting (and funding) engagement. A series of workshops and lectures have been organized across 2013 and 2014 and, at the time I write, preparations are underway for a workshop that readers of *Valuation Studies* would do well to pay attention to: "Let's Calculate: Reinventing Accounting with Bruno Latour."[2] This is how the call for contributions starts:

> Calculemus! (Let's Calculate!) Thus finishes Bruno Latour's last book . . . where he takes on accounting frontally and calls for its reinvention. Modern

> accounting, Latour argues, is indeed the cause of major evils, and therefore also the place where reform should start if one wants to close the modernist parenthesis and finally settle the debt of the West, the Whites, or yet again the Moderns with the rest of the world. This workshop aims at understanding, discussing and eventually furthering this new, radical critique of accounting.

It is not just the profession of accounting and its academic arm that is to be confronted here. To understand what is under attack one needs to grasp the term 'economization' as it is used in the *Inquiry*, following the approach of Latour's former colleague Michel Callon: "When one is dealing with economic matter, one has to be prepared to pile performatives on top of performatives, like tortoises in the fable, 'all the way down'!" (Latour 2013, 405). It comes down to devices of economization and the duplicitous masquerade they wage in presenting themselves as devices of referentiality when in actuality they are devices of control—devices of bad faith used with good will, when what we need, as any good contract lawyer will tell you, is devices of bad will, used in good faith.

"The [e]conomy fuses with organization to obtain idealized matter it uses inappropriately to shut down its calculations of interests and passions rather too quickly" (Latour 2013, 411). Thus the call: Calculemus! If Latour and his team have their way, the days of economy's/economics' duplicity are numbered. Recognizing that economics "produces measuring measures [methods of control] and not measured measures [assessments of extent]" (Latour 2013, 408), economic valuation will be changed forever. Of course this recognition will involve considerable re-organization; significant re-institutionalization. Before that can be done, we need to collectively—and this time wittingly—invent a few devices that will, in bringing publics into existence, initiate, maintain, regulate, and publicize flows of organization proceeding in the opposite direction.[3] You can see why I think readers of *Valuation Studies* need to collectively prick up their ears.

Extending the Right of Cosmopolitics to Non-Humans

So why set the *Inquiry* alongside a proposal made by Kant foreshadowing a cosmopolitical right (for humans)? In 1795, writing at the cusp of a waning feudal Europe and the dawn of capitalist modes of production, Kant argued that international commerce was a historical condition of the cosmopolitical community because commerce was incompatible with war and the self-interest of states. Appalled by the vision of "perpetual war" as the seeming implacable consequence of Hobbes' account of modernity, Kant's vision of a cosmopolitical right asserted in the name of a common humanity, attempted to provide an ideal institutional framework for "perpetual peace" through regulating the anarchic behaviour of states (Kant 1999).

Latour's *Inquiry* comes to life in Europe at a time when the cosmopolitanism of the European Union begins to fade under the assaults of nationalism (again), when the might of corporations who dominate international trade far exceeds the strength of many nations, and when modernity has progressed so far that in the face of ecological crisis, now economization (in the form of financialization) of the very planetary stuff that feeds modernity and its capitalizations, seems an appropriate response. Appalled at the continuing blindness of moderns to the values that *should have been institutionalized* in Kant's times and places, but which because of an absurd collective bedazzlement by ideals were unrecognized, indeed unrecognizable, Latour and his team set out to act.

Not attracted to ideal institutions at all—that's what's caused the problems since they brought with them ideal epistemic practices—the *Inquiry* has cosmopolitics proceeding in the opposite direction, from the ground up so to speak, and this is how non-humans will get into the collective acting this time around, if the team have their way. But how are we to have a chance of knowing what we're doing as, collectively, we blindly feel around the elephant of modernity, with no agreement on what this phenomenon *is*? This is where the machine comes in. It both helps "read the surface of the elephant that is modernity," and helps connect groups and individuals feeling the bumps and crevices in various sites—Latour has started us off by identifying fifteen sites; fifteen modes of modern existence. But, in his opinion, turning around the direction of economization is the most urgent. This does seem to be something to which those of us with interests in the workings of valuation regimes need pay attention.

Notes

1. The book (Latour 2013) is accompanied by a project ("AIME" for "An Inquiry into Modes of Existence") hosted by Science Po in Paris, which counts on a team of collaborators and an internet platform available at: www.modesofexistence.org (accessed 25 March 2014).
2. "Let's Calculate: Reinventing Accounting with Bruno Latour?" AIME Workshop, Paris, 5 May 2014, organized by Martin Giraudeau and Vincent Lépinay. See: www.modesofexistence.org (accessed 25 March 2014).
3. Helpfully recognizing that examples are important in this task, one contributor to the "extended book" points out this is the direction that accounting/economic/organizational valuation proceeded in ancient Egypt, referring us to the paper by Mahmoud Ezzamel (2009).

References

Ezzamel, Mahmoud.2009. "Order and Accounting as a Performative Ritual: Evidence from Ancient Egypt." *Accounting, Organizations and Society*, 34 (3–4): 348–380.

Haraway, Donna. 1992. "The Promises of Monsters: A Regenerative Politics for Inappropriate/d Others." In *Cultural Studies*, edited by Lawrence Grossberg, Cary Nelson, and Paula A. Treichler, 295–337. New York: Routledge.

Kant, Immanuel. 1999. "Perpetual Peace: A Philosophical Sketch." In *Political Writings*, edited by Hans Reiss, 98–99. Cambridge: Cambridge University Press.

Latour, Bruno. 1988. *The Pasteurization of France*. Cambridge, MA: Harvard University Press.

———. 1993. *We Have Never Been Modern*. Cambridge, MA: Harvard University Press.

———. 2004. *The Politics of Nature: How to Bring the Sciences into Democracy*. Cambridge, MA: Harvard University Press.

———. 2010. *On the Modern Cult of the Factish Gods*. Durham, NC: Duke University Press.

———. 2013. *An Inquiry into Modes of Existence: An Anthropology of the Moderns*. Cambridge, MA: Harvard University Press.

Latour, Bruno, and Peter Weibel, eds. 2002. *Iconoclash: Beyond the Image Wars in Science, Religion and Art*. Cambridge, MA: The MIT Press.

Helen Verran researches in anthropology, philosophy, and sociology of science and technology. She is reader (emeritus) in History and Philosophy of Science at the University of Melbourne where she taught for twenty five years, and adjunct professor at the Northern Institute, Charles Darwin University in Australia. The author of prize-winning *Science and an African Logic* (Chicago University Press 2001), Helen Verran studies, among other topics, how practitioners of disparate knowledge traditions produce numbers and work with them.

Valuation Studies 2(1) 2014: 71–85

Discussion note

Valuation Studies: A Collaborative Valuation in Practice

Gordon Haywood*, Johan Nilsson*, Michael Franklin, Paul Gilbert, Linus Johansson Krafve, Lisa Lindén, Mark MacGillivray, Robert Meckin

Abstract

This discussion note provides a perspective on valuation studies by a group of PhD students. Based on impressions from the Valuation as Practice workshop at The University of Edinburgh in early 2014 we were inspired by the example of Kjellberg et al. (2013) to debate how we see, understand, and are inspired by the field of valuation studies. It is the hope of the editors that sharing the concerns of early-stage researchers starting out in a field in flux, may be of use to, and perhaps spur, senior contributors to further develop this emerging research landscape. Using the workshop experience as a springboard, we argue that the domain of valuation studies still relies heavily on influences from the study of economics, with a strong emphasis on processes of quantification and calculation. With apparent pragmatism within the field, concern as to what might be lost by this narrower perspective is raised. Additionally, we call for the exploration of the possibility of a common language of valuation, to better define shared features, and identify as well as manage conflicts within the field.

Key words: values; valuation practices; discussion; pragmatism

Introduction

In the middle of February 2014 over thirty scholars and PhD students gathered for a workshop on valuation practice at The University of Edinburgh. The workshop brought together a number of academic

Gordon Haywood, The University of Edinburgh Business School, gordon.haywood@ed.ac.uk

Johan Nilsson, Department of Thematic Studies—Technology and Social Change, Linköping University, johan.nilsson@liu.se

© 2014 The authors
LiU Electronic Press, DOI 10.3384/vs.2001-5992.142171
http://valuationstudies.liu.se

fields including science and technology studies (STS), critical accounting studies and the sociology of finance, to explore the development of practices and processes of valuation. This discussion note stems from the reflections of participating early-stage researchers on their experience engaging with the emerging field of valuation studies. The editors of this text, Gordon Haywood and Johan Nilsson, seek to distil, in the manner of Kjellberg, Mallard, et al. (2013), the PhD students' "collective two cents": through asking for contributions from participating students in response to a simple questionnaire we work our way from the intimate surrounding of a workshop to what the field of valuation studies can encompass. The use of the first person in this text primarily reflects the perspectives of editors, in conversation with the larger field of valuation research (surely, dear reader, we can consider ourselves in the same boat?). At times we attempt to formulate what the collected body of co-authors have put forward, but mostly we have aimed to be true to the different standpoints of the individual contributors. We trust that in this, we can present heterogeneous opinions in a format that is coherent and thought provoking.

This move from small places to large issues: from starting in the experiences of the contributors to their ideas of what they say about valuation studies at large owes much to perspectives from ethnographic traditions (see for instance Hylland Eriksen 2001). The text will briefly describe the workshop setting, describe our inspiration from Kjellberg et al.'s method of collective review, and then move on to the authors' shared thoughts on what the field of valuation studies currently appears to be, what our hopes for it are and also offering some critical comments. This area of research is still very much preoccupied with economics and its effects, there is often an emphasis on processes of quantification and calculation, and it takes a pragmatist standpoint. We would like to raise the question of whether something is lost by these preoccupations. Additionally, we call for an exploration of the language of valuation, to better define commonalities as well as conflicts, in this field.

Background: The Valuation as Practice Workshop

> How can an insurance claim for a damaged knee be resolved into a sum of compensation? What is the fair price of polluting the environment? On what grounds can a new therapeutic technique be allowed introduction into the health care system? Did the right singer win the talent show? (Excerpt from the workshop invitation, 2013)

The dust is freshly settled after the workshop. Established researchers and PhD students met to talk about the determination and comparison of values, be they price, quantifications of quality, or ethics, which affect many parts of life, including university performance (Espeland 2014), charity shopping (Magee et al. 2014), technology procurement (Campagnolo and Pollock 2014) and systems of taxation (Björklund Larsen 2014). The symposium brought together a variety of local and international keynote speakers including Wendy Espeland (Northwestern University), Claes-Fredrik Helgesson, Karin Thoresson, Lotta Björklund Larsen (Linköping University), Neil Pollock, Gian Marco Campagnolo, Siobhan Magee, Chris Speed,[1] and Paolo Quattrone (University of Edinburgh). The participants also enjoyed two tracks of PhD students talking on engaging contemporary topics ranging from smart grids, open scholarship, independent films, and market research, to the politics of HPV vaccination and the marketization of the welfare state. With this experience in mind we now hope to characterise a PhD perspective on valuation as a field of study.

Valuation studies is an emerging field of research that places emphasis on the means and processes of achieving values, as well as their comparison and use (Helgesson and Muniesa 2013). The study of value and valuation ties in with the study of the mediation of multiple regimes of value, and the establishment of commensurable values (Sauder and Espeland 2009; Styhre 2013). The perspective also opens up avenues for study of the socio-technical means of calculation and mediation (Callon 1998; Vatin 2013). In this respect valuation studies relates to the fields of STS and of critical accounting: following the construction of economic accounts, and the technologies of evaluation, is an important way to gain a full picture of valuation that connects with, but goes beyond, conventional economics. However, the notion of "valuation studies" is not without its critics.

As PhD students, and consequently targets for the workshop's format as a doctoral school as well as a symposium, we are left wondering: did we take part in the on-going articulation of a promising new academic field? If so, what did we learn from a meeting of notable researchers interested in this field? What is valuation anyway, and what relationship does it have with perfectly common-place words like value(s) and evaluation? Are we more, or less, confused than before the workshop?

Method: The Co-Written Review

Working according to the format outlined by Kjellberg, Mallard, et al. (2013) we engaged participants attending the workshop to co-write this paper with us. Specifically, we invited all PhD students who attended the Valuation as Practice symposium to participate in a follow-up exercise. Following Kjellberg, Mallard, et al. (2013), we urged those interested in taking part to write back to us with short

[1] Presented paper co-authored by Siobhan Magee, Eric Laurier, Chris Speed, Mark Hartswood, Andrew Hudson-Smith, Fionn Tynan-O'Mahony, and Martin De Jode.

statements (1) about the workshop itself, (2) to let us know what thoughts about the field it had elicited, and (3) what contributors felt important to thrive in the valuation field in terms of tools and methods. Answers to our questions about these three areas have gone through an editing process, where we have attempted to represent the responses thematically.

We asked our contributors to let us know if they would like to feature in name as co-authors, or rather remain anonymous, and to what extent they were interested in taking an active part in the writing. Like the authors of the paper we took inspiration from, we summarised responses and synthesised a group consensus where possible before re-circulating the material for comments from the contributors—turning material that would otherwise be one-way contributions in the form of responses, into a more interactive collaborative process. As editors we have assumed the role of provoking and representing responses and have limited our input to contributions in sections that reflect the group of authors as a whole.

Approaching Valuation

As a group of early-stage researchers, we outline our understanding of what valuation studies mean, having seen and heard it manifest in workshop action. From there we move on to how we understand the relevance of valuation studies in light of this research, as well as point to how we can turn our confusion and doubts into hopes for future developments, elaborations of methods, and some sense of direction. Many of us are new to valuation studies, which should not come as a surprise given that the term is rather recent, and we range from those already contributing to the field, to those trying to grasp what it is in the first place. Thus the question of the relevance of valuation studies became a question of how to approach this novel enterprise. One way of approaching “valuation” (and indeed “value”) is to assess a number of ostensive definitions in order to get a feeling of what the matter at hand might actually be. In light of this, the workshop demonstrated valuation by offering a range of different sites where the term applied. Michael Franklin, found that examining historical accounting practices and popular media metrics unpacked the social and material concerns of valuation in various eras and laid out the recurring influence of such processes in economic life. Lisa Lindén noted that the inclusion of accounting, marketing methods, economic valuations of natural resources and the studies on health economics into one field is highly relevant in a contemporary society that tends to further economize diverse welfare practices. Robert Meckin observed that to study the processes by which the translations of value(s) from group to group are enacted may be a way to interrogate his data in the future.

Fascination with Calculation and Quantification

The inclusiveness of the workshop did not fully overcome a tendency that was identified by several contributors: the focus on sites deemed economic, and potentially calculable, that is, on value in the singular. To Linus Johansson Krafve, the workshop as a distillate of the field of valuation studies prompts the unfortunate announcement that "valuation" seems to be, first and foremost, about money, and that it offers too narrow a scope. Paul Gilbert himself perhaps a little like the contributors to Kjellberg et al. (2013, 19), travelled to Edinburgh in equal parts "fascinated and bothered" by "the dominance of *economic* value in contemporary society"; and left equally fascinated and bothered by the dominance of economic value in the valuation studies field. His impression from listening to the papers presented at Edinburgh, is of a field that is for the most part being assembled around an interest in the calculative side of valuation—to the apparent neglect of the social practices through which transcendental, moral and plural values, judgements and justifications are enacted.

Gilbert's impression was that it is in part the methodology of the valuation studies field that pushes it towards an overwhelming concern with calculative and economic forms of value and valuation; and that value, or moral and ethical valuation, may only appear inasmuch as calculative valuation practices throw light upon them (rather than in their own right). Referring to recent attempts to delineate the field (Kjellberg et al. 2013, 22, 24–25), he notes that an overwhelming emphasis is placed on studying domains of finance, "hot" economies, and ranking technologies, as well as on apprehending valuation processes via socio-technical *agencements* made up of instruments, devices and routines. This, it seems, is clearly a language of calculation, and is not symmetrically focused on questions of worth, valorization, judgement or justification. In the two workshop papers that Gilbert found to most obviously engage with questions of worth, plural values, and the mediating role played by concepts of reasonableness and fairness (Björklund Larsen 2014; Magee et al. 2014), non-calculative, non-economic values only manifested themselves in *valuation practices* in so far as they competed with or were involved in the co-production of economic value or price. Gilbert therefore wonders if studying valuation-as-practice, through socio-technical agencements comprised of instruments, devices, routines, rankings and classifications, has an elective affinity with calculative and economic forms of valuation.

If, as many of the contributors to Kjellberg et al. (2013, 13–14) point out, economic value is increasingly pervasive, then adopting an approach to studying economic valuation-as-practice which problematises flawed dominant models in the economic sciences (Kjellberg et al. 2013, 17) certainly seems reasonable. But Gilbert wonders whether there might be a risk that in adopting a pragmatist approach, and deriving many tools from the social studies of markets, the dominance of economic and singular modes of valuation is amplified via the way we choose to study it? For lack of a consensus on this among the co-authors of this discussion note, we suggest that it is a point of relevant contention. To return to the discussion during the doctoral workshop on theories for studying valuation, Gilbert argues that one approach singled out for criticism by other attendees and the workshop leaders was that which could be glossed as the "Marxian approach," for assuming that values exist before, and guide, social practice. Further, he points out that the debt that valuation studies owes to Actor-Network Theory (ANT) was made explicit in this workshop. It is well documented that ANT practitioners reject the notion that there is something analytically significant called "capital" or "capitalism" (Callon et al. 2002; Latour and Callon 1997; Latour 2004) which can be used as an explanation for social practices. Instead, ANT-inspired studies of markets (see for example Callon, Millo and Muniesa 2007) have focused on the contingent assembly of concrete socio-technical assemblages—a commitment the valuation studies approach shares (Kjellberg et al. 2013, 22). Gibert asks, might this therefore become a form of social analysis that mimetically reproduces, rather than approaches head on, the spread of economic value? Are Bertrand Russell's comments (Russell 2000, lxiii), made in the course of a debate with Dewey, that "pragmatism is the philosophical expression" of "commercialism" worthy of consideration, given the explicit debt to Dewey that was acknowledged by some of the leading valuation studies scholars presenting in Edinburgh?

Continuing to wrestle with the relationship between the subject matter of valuation studies and its methods, Gilbert wonders: if the pragmatist, ANT-inspired approach of the valuation-as-practice programme is going to concern itself primarily with economic value and calculative valuations—and perhaps there is no reason why it should not—then might it not benefit from taking methodological account of the wider structural features (often framed as "capitalism" or "neoliberalism") over which some of the field's frontrunners do seem to express concern (Kjellberg et al. 2013, 13–14)? John Michael Roberts' (2011) recent work may offer a way forward in this respect. The language might not be universally appealing, but Roberts' criticism, drawing on Deleuze and Guattari, that "ANT tends to focus on planes of organization at the expense of planes of immanence" (Roberts 2011, 38) seems worthy of further consideration, Gilbert suggests. By focusing on organisation—the socio-technical assemblages and market devices mentioned above—ANT (and Gilbert would add, "valuation-as-practice") misses out on "planes of immanence," effectively the structural features of capitalism. Rather than rejecting ANT/valuation-as-practice, is it not possible to embrace the pragmatism of the valuation-as-practice approach, but to do so in a way that avoids practitioners becoming "descriptive recorders of the "hidden principles" of concrete-contingent relations" (Roberts 2011, 42) in specific valuation contexts? While the authors of this article may not have found a consensus on the specifics of the pragmatic tendency of valuation studies displayed during the workshop (indeed, some of us would rather defend it!), we observe that this is certainly a point for further debate that would enrich the field.

Are we letting the allure of calculation and economic processes get the better of us? Lisa Lindén notes in a similar manner to Gilbert that while she could see the relevance of the workshop, it would have been more so had it demonstrated greater diversity in topics and theoretical approaches. Lindén points to a serious danger if the field of valuation studies turns out to be little more than grappling with economical values as empirical cases like the majority of the presentations at the workshop were doing. To claim relevancy requires symmetrical attention to other kinds of valuations. As an STS scholar who is not from a business school and who is not first and foremost interested in the empirical study of economic matters, the workshop made Lindén wonder about the extent to which she wants to be a part of valuation studies and what the field can contribute to her research. To Lindén it is crucial that we care for the diversity of valuation studies, its trans-disciplinary roots and potential. If not, it will be seen as a branch of studies focusing on business, marketing and economics. And there is more potential than that in the field, she argues.

As co-authors, we wish to see a move beyond Parson's pact that does not end up with us losing sight of values as they have been described in ethics, or social studies. Johansson Krafve phrases this succinctly: the borders of valuation studies must be inclusive in the making of "values" (in general), that is, the making of desirable states and ethical concerns. It is not enough to settle with a definition of valuation as a practice to determine the value of *something* (in monetary terms). We (who attempt to contribute to the field of valuation studies) must be extremely cautious not to move to the other side of Parson's pact, leaving our "old" concerns behind. Valuation studies must never settle for only describing valuation in economic terms, and must take other, plural perspectives (moral, ethical, etc.) seriously as well. Gilbert hopes that the valuation studies field, with its focus on valuing as "an activity" (Heuts and Mol 2013, 129) can be reconciled with or brought into productive dialogue with parallel attempts to reinvigorate the study of value(s) in anthropology, for example where Michael Lambek (2013, 148) argues for an increased attention to the (ethical) values that are "generated in social acts" and activities.

"Valuation": What Does It Mean?

Several of us are concerned about the terminology we utilise to understand and explain phenomena that may be emically regarded as judgement, assessment, reactions, classification, rating, ranking, and evaluation. During the workshop a number of researchers spoke about valuation practices and studies, prompting the contributors to wonder what to make of this. Johansson Krafve noted that as a non-native English speaker the workshop served to elaborate what we mean by valuation (including evaluation, valuables, values/-s, worth etc.). He was not alone in noting that language was a point of concern. Finding shared terminology is important and remains a goal to be achieved. As Johansson Krafve puts it: we better bring to the fore what we really mean by values and valuation. Perhaps we have been a little too happy exploring what could be included under a valuation banner, at the expense of not having spent enough time arguing about what we mean by the term "valuation."

Michael Franklin suggests that one conflict that has permeated the discussion of valuation and market studies is the notion that in including everything (materials, humans, processes, institutions, individuals) the initiative explains nothing. What, exactly, would give the field of valuation studies its unique tenor, and what distinguishes it from other more diffuse attempts to grapple with "Everything That Matters" (cf. Demian 2003)? Responses to this issue were forthcoming at the workshop. By avoiding pure description, and shifting the focus to comparative studies of particular configurations of valuation there may be a way out of this critical problem. In exploring the fights that break out at the boundaries of different arrangements of evaluative practice, it might be possible to learn something about the interaction of values and evaluation. Robert Meckin took the opportunity to try and make sense of his fellow researchers' use of terminology: there appear to be two dominant forms of the use of case studies in valuation practices which relate to the discussion of value as a noun and as a verb (Kjellberg et al., 2013). The first might be understood as the "(re)production of values" or the "performance of values" and focuses on identifying values as outcomes of activity and interaction. The second strand considers how values are made—manifest in practices, tools and use—and seems to sit more readily with the title *valuation practices* since it focuses on the emergence and effects of *valuations* as opposed to the emergence of *values* as moral principles. Lisa Lindén saw the lack of theoretical conflicts as a reason to think otherwise. She found many of the presentations oriented around empirical issues with only implicit theoretical consideration. This is not a problem per se, but she felt explicit theoretical discussions were missing. It is important to continue discussing the tensions and possible convergences between theories used in valuation studies. As both Kjellberg et al. (2013) and the workshop indicated, there is great potential in casting a wide net, but we need to know how expansive it is, and how fine a mesh.

Where Does Valuation Studies Go from Here? Our Suggestions

The contributors and editors of this discussion piece seem to have arrived at largely the same level of confusion about what our seniors could possibly mean when they speak so confidently about valuation. And we are faced with understanding that valuation could be many things (though it often tends to include calculation and economic value in the singular as a recurring sounding board for other values). So, what is the direction of this field? What is this group of young researchers hoping for? First of all, valuation studies needs to deal with jargon. Franklin suggests that stable and cohesive terminology may add legitimacy and give a useful analytical toolset. Antecedents of valuation studies include market studies, which draws on ANT and STS, but suffers from an explosion of vocabulary. Terms like "market," "socio-technical agencements," "qualculation," etc. are often the result of useful explorative arguments but are too unwieldy to be adopted for any length of time, even by their creators. While not all contributors agree on the possibility of a common terminology we maintain that a surgical use of prose would be beneficial.

Related to the call for clarity of nomenclature, those of us who wish to participate in valuation studies need to remember what values mean more generally. Proponents of the valuation studies approach cannot forget that values have been about ethics as much as they have been about money. The workshop placed considerable emphasis on commensuration practices. But, as Johansson Krafve notes, we should not get stuck in understanding commensurability of values by describing processes that *make* something commensurable (as happens when everything is translated into the same metric). We also need to account for the differences in meanings of things that appear to be the same (e.g. a dollar bill could mean many things and stand as a proxy for a multitude of values). We should attempt to describe such events and phenomena as processes: after all, they seem to be about valuation practices in some way.

Paul Gilbert similarly calls for recognising that valuation can be about values in a broader sense by drawing on Paolo Quattrone's closing talk, "Valuation in the Age of Doubt," which ended with a reference to Grafton and Jardine's (1986) *From Humanism to the Humanities*. Quattrone argued that accounting (valuation) practices had only recently (in the 1980s) become purely financial and representative—having spent 500 years as an exercise for questioning the morality of spending, based on practices of invention and mediation. True enough, it might be a contemporary reality that economic value is primary (though Quattrone reminds us that "Parson's Pact," and the abandonment of "social" or "moral" plural values, only became a reality in accounting practice in the 1980s), but shouldn't the valuation studies field be a little more hesitant in its pragmatic commitments? Might such pragmatism be driving this newly constituted endeavour towards an overly economic, calculative focus, in effect "over-correcting" the wrongs caused by Parson's Pact? Gilbert urges a rethinking or confrontation of the reasons for the focus on economic value and calculative valuation practices. Is it because economic value is foremost in most contemporary times and places? If so, do we in the field run the risk of amplifying that state of affairs by not developing methodologies for approaching non-economic, plural, moral, values? Additionally, he calls for serious consideration or incorporation of attention to the structural conditions in which economic or calculative valuation practices take place—or at the very least, a thorough justification for not engaging with abstract and structural aspects of economic and calculative valuation. Finally a deep reflection on the pragmatism that underpins the valuation-as-practice approach is needed.

Lisa Lindén similarly misses valuation studies that deal with the tension between more pragmatic takes on values and valuations, and more normative ones. As a part of that, she believes it potentially fruitful for valuation studies to also discuss and include feminist, critical, and political approaches to valuation practices as a way of exploring the possible diversity of the field. Another important tension would be between the more STS pragmatic take on the topic and the more Foucauldian theories on governmentality. At the workshop, it was apparent that (implicitly or explicitly) understanding values as enacted in practices is different from considering values as the reason for individual actions or societal processes. Such tensions between the different ways in which values and valuation are spoken about need to be debated, not because they are problematic but because they make possible fruitful contrasts and potential connections.

What can those of us in the field do about the way we choose to study valuation? Mark MacGillivray wants to see more experimenting, possibly leading us to better notions of what sort of theories and methods we need for a given problem or phenomenon. Valuation is an act that occurs in a given situation, and the state of that situation is what is relevant to the act rather than the history or tradition (although those things may help characterise the given state), suggesting that ANT is probably quite relevant to valuation studies. Lindén wishes to see a greater diversity of approaches: it is crucial to value different kinds of methods. It would be problematic if valuation studies, for example, mainly draws upon STS studies dealing with practices as ethnographic findings, she notes. Since many interesting and important valuation practices happen in, and through digital media, she also thinks these spaces could be further explored as interesting sites for valuation studies. This would require us to be more creative when it comes to how we use methods. On a similar note, research dealing with visual and performance arts and design is interesting for further investigating the potential of valuation studies as both an academically and a politically relevant area of study.

Researchers in the field of valuation studies ought not to forget that we are dealing with actors involved in valuations, and that insiders may offer us rich local reflexivity beyond a more mechanistic perspective. Indeed this reflexivity can be taken as far as being about ourselves. Johansson Krafve notes that we should develop and maintain an emic perspective in presenting and demonstrating the various values we encounter in our research. Meanwhile, we should develop great skill in giving flesh—and be fair—to the reflexivity of the informants. Those working in practice with (for example) markets know that they are changing and performing values all the time. On the other hand, we must not shy away from engaging in moral issues. Johansson Krafve argues that we, in the valuation studies field, cannot suggest that categories of "enactment of value" are emic things, when they are not. We claim to see categories enacted all the time, be it responsibility, ontology or value, but the frame of reference for detecting those categories has been established long before they were encountered as "enacted."

Meckin also takes interest in more analyses of the affective dimension of valuation. During the workshop, both Espeland and Pollock mentioned actors' animosity to, or even devaluing of, the tools of valuation themselves. Yet the actors, often through coercion, still take part in that particular valuation practice. This "valuing of valuation" is a broader phenomenon and hopefully one which will be explored further. The call for reflexivity is also noted by Gilbert: little is heard of how writing about valuation practices can be a form of valuation in its own right. And, since the *Valuation Studies* journal is (commendably) open access, the manner in which academic "experts" intervene in valuation practices, recasting them as examples in their writing and theory building, should perhaps be subject to a touch more reflection?

Concluding Remarks

Standing back from our own reflections after a local valuation studies event, what have we the contributors learned about the field at large? Firstly, it seems that as a basis for empirical research, the study of value and valuation practices can be developed in a wide range of disciplines in new and interesting ways. Critically, this could be seen as evidence that the endeavour is too broad or even unbounded. There is certainly a challenge determining a common language and ground for valuation studies—indeed its integrity could be said to rely on a welcoming frontier spirit of subjects and studies coupled with the malleability of academic English to incorporate value as both noun and verb, outcome and process. Still, our experience is that there is hope in valuation as a social practice that transcends the at times artificial boundaries of academic study. This prompts us to deal with jargon, and to keep in mind several lessons from other disciplines and mature areas of research, when embarking in this field. While the emphasis on examining calculation and economic valuation may be a sign of the novelty of valuation studies outside of already established fields of research such as accounting, and economic sociology, it is important to go beyond areas of primarily financial and quantitative values. Scholars of valuation studies do well not to keep reinventing the wheel, and we ought to keep our options open in terms of methods. Finally, we suggest that with such a promising area of study it is also time to bring theoretical commonalities *as well as* conflicts to the surface in productive debate regarding the limits of valuation studies. Our discussion note is a modest attempt to contribute to such an endeavour.

References

Björklund Larsen, Lotta. 2014. ”The Shaping of a Tax Payer: Values and Valuations at the Swedish Tax Agency.” Workshop presentation. Valuation as Practice. 2014-02-13–2014-02-14. The Scottish Graduate School of Social Science, Edinburgh University.

Callon, Michel. 1998. *The Laws of the Markets*. Oxford: Blackwell.

Callon, Michel and Latour, Bruno. 1997. "'Tu ne calculeras pas!'- ou comment symétriser le don et le capital" Revue du MAUSS, no. 9.

Callon, Michel, Cecile Meadel, and Vololona Rabeharisoa. 2002. "The Economy of Qualities." *Economy and Society* 31 (2): 194–217.

Callon, Michel, Yuval Millo, and Fabian Muniesa. 2007. *Market Devices*. Oxford & Malden: Blackwell.

Demian, Melissa. 2003. "Review of *An Anthropological Theory of Value: The False Coin of Our Own Dreams* by David Graeber." *American Ethnologist* 30 (2): 316–317.

Espeland, Wendy. 2014. ”Valuing Universities: How Media Rankings Have Changed Education in the U.S. and Elsewhere.” Workshop presentation. Valuation as Practice. 2014-02-13–2014-02-14. The Scottish Graduate School of Social Science, Edinburgh University.

Latour, Bruno. 2004. "Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern." *Critical Inquiry* 30: 225–248.

Grafton, Anthony, and Lisa Jardine. 1986. *From Humanism to the Humanities: Education and the Liberal Arts in Fifteenth- and Sixteenth-Century Europe*. London: Gerald Duckworth & Co.

Helgesson, Claes-Fredrik, and Fabian Muniesa. 2013. "For What It's Worth: An Introduction to Valuation Studies." *Valuation Studies* 1 (1): 1–10.

Hylland Eriksen, Thomas. 2001 (1995). Small Places, Large Issues: An Introduction to Social and Cultural Anthropology. Second edition. London & Sterling: Pluto Press.

Kjellberg, Hans, Alexandre Mallard, Diane-Laure Arjaliès, Patrik Aspers, Stefan Beljean, Alexandra Bidet, Alberto Corsin, Emmanuel Didier, Marion Fourcade, Susi Geiger, Klaus Hoeyer, Michèle Lamont, Donald MacKenzie, Bill Maurer, Jan Mouritsen, Ebba Sjögren, Kjell Tryggestad, François Vatin, and Steve Woolgar. 2013. "Valuation Studies? Our Collective Two Cents." in *Valuation Studies* 1(1) 2013: 11–30.

Lambek, Michael. 2013. "The Value of (Performative) Acts." *HAU: Journal of Ethnographic Theory* 3 (2): 141–160.

Magee, Siobhan, Chris Speed, Eric Laurier, Mark Hartswood, Andrew Hudson-Smith, Fionn Tynan-O'Mahony, and Martin De Jode. 2014. "Haggling Teapots and Valuation Practices in Charity Shops". Workshop presentation. Valuation as Practice. 2014-02-13–2014-02-14. The Scottish Graduate School of Social Science, Edinburgh University.

Pollock, Neil, and Gian Marco Campagnolo. 2014. "Subitizing." Workshop presentation. Valuation as Practice. 2014-02-13–2014-02-14. The Scottish Graduate School of Social Science, Edinburgh University.

Roberts, John Michael. 2011. "Poststructuralism Against Poststructuralism: Actor-Network Theory, Organizations and Economic Markets." *European Journal of Social Theory* 15 (1): 35–53.

Russell, Bertrand. 2000 [1922]. "As a European Radical Sees It." In *The Collected Papers of Bertrand Russell, Volume 15: Uncertain Paths to Freedom: Russia and China, 1919–22*, edited by Richard A. Rempel, Beryl Haslam, Andrew Bone and Albert C. Lewis, 326–333, London: Routledge.

Sauder, Michael, and Wendy Nelson Espeland. 2009. "The Discipline of Rankings: Tight Coupling and Organizational Change." *American Sociological Review* 74(1): 63–82.

Styhre, Alexander. "The Economic Valuation and Commensuration of Cultural Resources: Financing and Monitoring the Swedish Culture Sector." *Valuation Studies* 1(1): 51–81.

Quattrone, Paulo. 2014. "Valuation in the Age of Doubt". Workshop presentation. Valuation as Practice. 2014-02-13–2014-02-14. The Scottish Graduate School of Social Science, Edinburgh University.

Vatin, François. 2013. "Valuation as Evaluating and Valorizing." *Valuation Studies* 1(1): 31–50.

Gordon Haywood is a PhD candidate at The University of Edinburgh Business School, and affiliated with the Institute for the Study of Science, Technology and Innovation (ISSTI). He is interested in social and managerial aspects of information infrastructures, specifically the emergence of markets for Enterprise technologies, with a focus on futures, fashions and expectations. His doctoral study is a longitudinal ethnographic account of technology sales and marketing in the corporate sector. Gordon holds degrees in Computer Science and Business Administration both from The University of Edinburgh.

Johan Nilsson is a PhD candidate in Technology and Social Change at Linköping University. He is trained as an anthropologist through the CREOLE master program in Stockholm University and the National University of Ireland, Maynooth. During these studies Johan researched refined wine consumption. His research interests have come to orient around questions of local knowledges and cosmologies, in particular the making of knowledge, markets, and values in marketing, by ethnographically studying market researchers. Theoretically Johan relates to economic anthropology and sociology, the anthropology of markets and consumption, as well as contemporary science and technology studies.

Michael Franklin is a PhD candidate at the Institute for Capitalising on Creativity, University of St Andrews and is affiliated with Creative Scotland: the national leader for Scotland's arts, screen and creative industries. Michael's doctoral research focuses on market devices for managing uncertainty in the independent film business and addresses issues of valuation, performativity and digital materiality. This work has been presented and developed at the Oxford Internet Institute SDP, and the Medici Management Program of NYU Stern, HEC Paris & UNIBO. Michael is a graduate of the universities of St Andrews, Edinburgh and Cass Business School.

Paul Robert Gilbert is a Ph.D. candidate in social anthropology at the University of Sussex. He has previously published on compensation payments and mine closure preparation in Durham Anthropology Journal, and on the marketisation of biodiversity in Agriculture & Human Values. He is currently conducting doctoral research focused on Bangladesh and the anthropology and socioeconomics of extractive industry investment in a 'frontier' market, and working as a fellow at the UK's Parliamentary Office of Science and Technology as part of a project looking into alternative currency schemes.

Linus Johansson Krafve is a PhD Candidate at the Department of Thematic Studies—Technology and Social Change, Linköping University. He is interested in economic valuation practices in the public sector and its effects on the relations in and between politics, administration and citizens. His dissertation is about the design of a primary care market in a Swedish county council.

Lisa Lindén is a PhD candidate in Technology and Social Change at Linköping University. Her main research interests lie in the intersection of feminist theory, science and technology studies (STS) and medical sociology. Her PhD project is focused on how practices of public health campaigns for HPV vaccination assemble subjectivities, technologies and care. In the thesis, she discusses questions of the making of goods and bads in HPV vaccination practice. In doing so, she is concerned with how certain bodies, lives and futures are enacted, valued, and managed.

Mark MacGillivray is a PhD candidate in Social Informatics at the University of Edinburgh. He is trained as both a philosopher and software developer through prior MA and MSc programs at Edinburgh. From personal experience running a software development company building open source software for research and education he has often encountered situations where technical solutions alone are insufficient, and where a socio-technical perspective is of benefit. Mark finds the application of the concept of Valuation Studies to be an excellent method for extending software engineering into a socio-technical reconstruction of value practices.

Robert Meckin is a PhD student on the Science and Technology Studies stream of the White Rose Doctoral Training Centre. He is based at the University of Sheffield. First he trained and worked as a biomedical scientist researching antipsychotic pharmaceuticals, then worked as a secondary school science teacher. He is currently interested in the field of synthetic biology and the incorporation of various futures in objects during the process of innovation.