Valuation Studies vol. 11(1) 2024

Contents:

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

The journal IT-infrastructure is provided and maintained by the *Linköping University Electronic Press* (LiU e-press).

Web:	http://valuationstudies.liu.se
E-mail, editors:	editors.valuationstudies@gmail.com
Twitter:	@Val_Studies

Valuation Studies 11(1) 2024: 1–8

Theme issue editorial

Experiences of Digitized Valuation

Francis Lee, Andrea Mennicken, Jacob Reilley, and Malte Ziewitz[1]

In the editorial to the first part of this themed issue (Lee et al. 2022), we suggested that digitization is not simply a process of turning existing valuation instruments and practices into code. Rather, digitizing valuations can have unique implications for how social order is established, challenged, and maintained. To help us think about the dynamics of digitized valuation, we outlined six initial themes: digitization, infrastructure, power and agency, automation and judgment, accountability and fairness, as well as generativity and performativity. Each of these themes raised a number of questions, some of which have been addressed by papers in this double issue and some of which will be addressed in future work. Instead of adding further to the list, this closing editorial attempts to shift perspectives and explore an aspect of digitized valuation that has not yet been given much attention in the context of this journal, namely the relationship

[1] Alphabetical order, authors contributed equally.

Francis Lee is Associate Professor of Technology and Social Change at Chalmers University of Technology.
Andrea Mennicken is Associate Professor of Accounting at the London School of Economics and Political Science.
Jacob Reilley is Assistant Professor of Accounting at the University of Groningen.
Malte Ziewitz is Associate Professor at the Department of Science & Technology Studies at Cornell University.

Hosted by Linköping University Electronic press
https://doi.org/10.3384/VS.2001-5992.2024.11.1.1-8
http://valuationstudies.liu.se
Valuation Studies logo by Rebecca Elfast

between practices and *experiences* of valuation. We suggest that it can be fruitful to revisit and examine more explicitly how experience is mediated, challenged, and constituted in contexts of digitized valuation practices (see also Ziewitz forthcoming).

A focus on the lived experiences of people as a way of understanding social life is of course not new. Philosophers, sociologists, anthropologists, and many others have long wondered how we, as human beings, make sense of and comprehend the world (see, e.g., Berger and Luckman 1967; Schutz 1967; Jay 2005). In the Modern American and European tradition, experience tends to be seen as an inward-looking awareness of the world – or as the anthropologist Robert Desjarlais (1994: 888) put it, a state that "readily equates with a person's inner life of consciousness and is often synonymous with subjectivity." We propose a different approach to the study of experience, namely one that is aligned with developments in science and technology studies (STS) and work associated with the so-called postphenomenological turn. Postphenomenology, as Ihde (2009) and Verbeek (2005) explain, substitutes embodiment for subjectivity. It is an attempt to overcome the modernist dichotomy between subject and object, human and world, by replacing it with a mutual interrelation (Verbeek 2005: 110). The subjectivity and objectivity of experience are constituted in relation to each other (see also Vindenes and Wasson 2021). At the same time, there has been a growing interest in the "sciences of subjectivity" as a form of world-making in STS and related fields (Shapin 2012: 179), focusing on subjectivity as a practical accomplishment and challenging the subjective-objective divide (e.g., Stenner 2008; Liberman 2014).

In other words, rather than taking experience or the existence of experiencing subjects for granted, we would like to ask how self-concepts are constituted through repeated encounters with digitized evaluation. Although researchers have already begun to broaden the scope of their investigations from a focus on data-driven technologies to the experiences of those who are subjected to these technologies, we suggest that stronger connections can be made to the notion of experience. While there is, by now, a rich literature on different valuation practices and devices (see the articles published in this journal over the years), less is known about the subjects of evaluation and their experiences of being valued, especially if they are not actively involved in valuation or digitization processes. As Ziewitz (forthcoming) points out:

> This area of concern … has become particularly salient with the rise of computational and other automated forms of valuation that tend to track and trace their subjects often without them being aware of their predicament, as in the case of predictive policing, credit scoring, and workplace monitoring.

Furthermore, as examples of "user experience" and "patient experience" have shown, experience has become a commodity and object of evaluation in its own right (Ziewitz 2017). It is this duality of experiencing valuing and evaluating experience that deserves more attention. How have (e)valuations of user experience changed with the rise of new digital technologies? How are such (e)valuations experienced by those subjected to them?

Exploring experience in the context of digitized valuation in this way can help us address a number of important issues. For one, as we already observed, subjects of evaluation are often not aware that they are being tracked and measured, raising questions of transparency and agency in the shadow of these systems. But even when people are aware of digital surveillance, it is often not quite clear how exactly experiences are turned into ratings, scores, and rankings, making it difficult to challenge judgments after the fact. For the most part, data subjects are told to "be themselves" in order not to interfere with processes of measurement, keeping them "scientific" and "objective" (Ziewitz and Singh 2021: 2). While such behavioral imperatives make sense from the perspective of managers and engineers, they tend to take on different lives in practice. As scholars have shown, people subject to evaluation engage in a range of reactive practices, developing new forms of adjustment, contestation, and resistance (e.g. Espeland and Sauder 2007; Ziewitz 2019; Rahman 2021; Ossandón 2022).

A good illustration of the dissonances that can emerge between digitally controlled experiences and what users are actually looking for is Lury et al.'s article (this issue) on "Digital Valuation: Lessons in Relevance from the Prototyping of a Recommendation App." Studying how people experience digital music recommendations in a world in which machines cannot grasp a lot of social context, the authors show how music recommendation apps may include people in categories that may not match their social world. An evening of listening to Elvis Presley with your mum, for instance, does not make you an aficionado of 1960s rock'n roll. Systems may process the world differently, missing important clues about what is happening in the user's world.

The article thus highlights an interesting facet of being valued by and valuing through digital systems. When users' experiences are mediated through data, applications, and infrastructures, questions about the nature of experience gain new salience. Whose experience are we talking about? Where is experience located? Where and how is the interaction ordered? In these cases, users try to make sense of systems and their own experiences with them. Trying to understand how systems work will arguably lead to speculation about why a system recommends a particular item. Think of folktales about Facebook listening to your conversations and then starting to display "related" ads. Why is the system constantly inserting Kanye West into

your life? Such experiences provoke a shift in analytical perspective. Rather than asking how valuation is being digitized – focusing on the roles and relevance of automated systems or databases – we are prompted to explore how digitization is made meaningful by those who are being algorithmically categorized, rated, and ranked.

The question of meaning-making is foregrounded in Wagenknecht et al.'s article (this issue) on "Digitised Valuation in Videoconference Workshops: Attaching Online Data to Stakes, Selves, and Other Data." The authors highlight the subjective, situated dimension of making online data meaningful and relevant, and thereby valuable. Drawing on ethnographic observations of two virtual workshops on the scarcity of water, they investigate how workshop participants understood and valued different types of data through "attachments" (Hennion 2007, 2017), i.e., relations they created between themselves, their experiences, the stakes of their task, and various data types. The analysis shows that for data to become valuable to people, they have to be made relatable and manageable. Giving data value requires grappling with, and modifying, the situatedness of data in complex relations.

Statements like "this data is beautiful," "this data is ugly," and "this data is relevant" illustrate the processes whereby seemingly objective data are imbued with users' valuations and experiences. The work of cleaning, selecting, and relating data to different experiences, value registers, and tastes is always situated; data are made present, apprehended, and endowed with worth through everyday relations. This observation points to how digital systems are not only mediating human experience, but also are themselves being mediated through experience. Just as realities are performed for users through digital systems, realities are performed for digital systems by users.

Of course, the formation of attachments, concerns, and subjectivities in digitized valuation is likely to be experienced differently depending on one's position and role in these emerging domains. We ought to take seriously the politics of experience and the challenges of popular calls for "giving people a voice." Digitized infrastructures of classification and valuation affect whose experiences count and generate new forms of inclusion and exclusion (Fourcade and Healy 2013). We need to understand the political implications of exploring digitized valuation experiences. As Scott (1991: 797) reminds us, "what counts as experience is neither self-evident nor straightforward; it is always contested, and always therefore political."

The final article in this themed issue is Justesen and Plesner's study of "Angry Citizens and Black Belt Employees: Cascading Classifications of and around a Predictive Algorithm." The article traces the development, roles, and effects of a predictive algorithm in a debt-collecting public sector organization. Drawing on concepts of nominal and ordinal classification (Fourcade 2016), they examine how

intended non-hierarchical classifications glide into new hierarchical valuations of both citizens and employees. In their terminology, classifications were cascading. Classifications provided by the algorithm, such as classification of citizens in terms of their "readiness to pay," became entangled with other classifications. As Justesen and Plesner (this issue: 11) write:

> Organizational actors superimposed new and different classifications onto those provided by the algorithm. The latter became entangled with classification of citizens in terms of motivation or attitude (who is willing to pay), the potential trouble they might cause (who is a 'difficult' person), or their emotional state (who is an 'angry' person). At the same time, employees had to be recategorized to match the algorithm's proposed citizen categories.

The notion of "cascading classifications" thus draws attention not only to the dynamic and unstable relationship between algorithmic, organizational and individual valuation practices, but also to the political consequences of digitized valuation.

A focus on experience thus changes how we understand and approach the study of digitized valuation's political implications, how people are made (in)visible, how they can (or cannot) participate in processes that reclassify and evaluate them. A concern with experience can help us trace the unarticulated import of assumptions about social, political, and other differences. Since we are already confronted with digitized evaluative infrastructures (Kornberger et al. 2017) in our everyday lives, analyzing the lived experiences of data subjects — and data producers — can help us to reflect on how we understand the new categories, roles, and processes of defining value that are emerging, as well as how users' experience mediates systems' experience. It allows us to explore what it means to be measured in these situations or how people perceive their agency in digitized spaces. This focus might further help us understand how someone sees the concealment of human input or the embedding of biases in algorithmic systems; or whether there are new forms of intersubjective agency that may emerge as a result. Similarly, we could ask what it feels like to be involved in accountability dynamics generated by digitized valuation, or what kinds of experiences and subjectivities lead to the resistance, use, appropriation, and creation of different digital valuations.

Together, the three articles provide a wealth of inspiration for studying the intersection of experience and digitized valuation. Each article raises a unique set of questions about the methodological, theoretical, and political dimensions of experience as both a topic and a resource for inquiry. In doing so, they also contribute more generally to the study of valuation as a "problem" (Board of Editors 2020) – not in the normative sense of claiming there is something wrong with

forms of digitized valuation (although we can discuss that, too), but in the sense of studying the "problems of those who value and are subject to valuation" (p. 2). Just as "raw data" is an oxymoron (Gitelman 2013), "raw valuation" is a contradiction in terms. Digitization, experience, and valuation are always already folded into one another and should be studied accordingly.

References

Berger, Peter L., and Thomas Luckman. 1967. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. Garden City, NY: Doubleday.

Board of Editors. 2020 "Towards a Reformulation." *Valuation Studies* 7(1): 1–2.

Desjarlais, Robert. 1994. "Struggling Along: The Possibilities for Experience among the Homeless Mentally Ill." *American Anthropologist* 96(4): 886–901.

Espeland, Wendy N., and Michael Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." *American Journal of Sociology* 113(1): 1–40.

Fourcade, Marion. 2016. "Ordinalization: Lewis A. Coser Memorial Award for Theoretical Agenda Setting 2014." *Sociological Theory* 34(3): 175–195.

Fourcade, Marion, and Kieran Healy. 2013. "Classification Situations: Life-chances in the Neoliberal Era." *Accounting, Organizations and Society* 38(8): 559–572.

Gitelman, Lisa. 2013. *"Raw Data" Is an Oxymoron*. Boston, MA: MIT Press.

Hennion, Antoine. 2007. "Those Things That Hold Us Together: Taste and Sociology." *Cultural Sociology* 1(1): 97–114.

Hennion, Antoine. 2017. "Attachments, You Say? How a Concept Collectively Emerges in One Research Group." *Journal of Cultural Economy* 10(1): 112–121.

Ihde, Don. 2009. *Postphenomenology and Technoscience: The Peking University Lectures*. Albany, NY: SUNY Press.

Jay, Martin. 2005. *Songs of Experience: Modern American and European Variations on a Universal Theme*. Berkeley, CA: University of California Press.

Kornberger, Martin, Dane Pflueger, and Jan Mouritsen. 2017. "Evaluative Infrastructures: Accounting for Platform Organization." A*ccounting, Organizations and Society* 60: 79–95.

Lee, Francis, Andrea Mennicken, Jacob Reilley, and Malte Ziewitz. 2022. "Digitizing Valuation." *Valuation Studies* 9(1): 1–10.

Liberman Kenneth. 2014. "The Phenomenology of Coffee Tasting: Lessons on Practical Objectivity." In *More Studies in Ethnomethodology*, 215–166. Albany, NY: State University of New York Press.

Ossandón, José. 2022. "(University) Management after Valuation Studies: Carving a Practice between the Offended Native, the Anxious Scholar, and the Useless Practitioner." *Valuation Studies* 8(2): 61–72.

Rahman, Hatim A. 2021. "The Invisible Cage: Workers' Reactivity to Opaque Algorithmic Evaluations." *Administrative Science Quarterly* 66(4): 945–988.

Schutz, Alfred. 1967. *The Phenomenology of the Social World*. Evanston, IL: Northwestern University Press.

Scott, Joan W. 1991. "The Evidence of Experience." *Critical Inquiry* 17(4): 773–797.

Shapin, Steven. 2012. "The Sciences of Subjectivity." *Social Studies of Science* 42(2): 170–184.

Stenner, Paul. 2008. "A.N. Whitehead and Subjectivity." *Subjectivity* 22(1): 90–109.

Verbeek, Peter-Paul. 2005. *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. University Park, PA: Pennsylvania State University Press.

Vindenes, Joakim, and Barbara Wasson. 2021. "A Postphenomenological Framework for Studying User Experience of Immersive Virtual Reality." *Frontiers in Virtual Reality* 2: 1–15.

Ziewitz, Malte. 2017. "Experience in Action: Moderating Care in Web-Based Patient Feedback." *Social Science & Medicine* 175: 99–108.

Ziewitz, Malte. 2019. "Rethinking Gaming: The Ethical Work of Optimization in Web Search Engines." *Social Studies of Science* 49(5): 707–731.

Ziewitz, Malte. Forthcoming. "On STS and Valuation." In *The Routledge International Handbook of Valuation and Society*, edited by Anne K. Krüger, Thorsten Peetz, and Hilmar Schäfer. Abingdon; New York, NY: Routledge.

Ziewitz, Malte, and Ranjit Singh. 2021. "Critical Companionship: Some Sensibilities for Studying the Lived Experience of Data Subjects." *Big Data & Society* 8(2): 1–13.

Francis Lee is Associate Professor of Technology and Social Change and senior researcher at the unit for Science, Technology and Society at Chalmers University of Technology in Sweden. His research focus lies on the politics of information infrastructures and knowledge production. Of particular concern for Lee is how information infrastructures – such as AI, algorithms, or big data – shape science and society. For instance he is interested in how disease outbreaks are constructed with algorithms, how AI is reshaping research practices of

the biosciences, or how risky people are constructed and tracked using different types of information infrastructures. His webpage is available at https://francislee.org.

Andrea Mennicken is Associate Professor of Accounting at the London School of Economics and Political Science, and Co-Director of the Centre for Analysis of Risk and Regulation. Her work has been published in A*ccounting, Organizations and Society*, *Financial Accountability and Management*, *Foucault Studies*, *Organization Studies*, *Sociologie du Travail*, amongst others. Her research interests include social studies of quantification and accounting, auditing, governance and regulation, processes of marketization, standardization, and public sector reforms. Lately, she completed an international research project exploring the changing relationships between quantification, administrative capacity and democracy in healthcare, correctional services, and higher education. Recent publications include "What's New with Numbers: Sociological Approaches to the Study of Quantification" co-authored with Wendy Espeland, *Annual Review of Sociology* 2019, and *The New Politics of Numbers: Utopia, Evidence and Democracy*, co-edited with Robert Salais, 2022, Palgrave Macmillan.

Jacob Reilley is an Assistant Professor of Accounting at the University of Groningen. His research focuses on processes of quantification and digitization in non-profit organizations. He is currently researching how risk and quality are digitized in hospital management systems; more recently, he has also examined how impact is accounted for in NGOs. His work has been published in *Research in the Sociology of Organizations*, *Critical Perspectives on Accounting*, *Financial Accountability and Management*, and *Accounting, Auditing and Accountability Journal*.

Malte Ziewitz is Associate Professor at the Department of Science & Technology Studies at Cornell University. An ethnographer and sociologist, he studies the changing role of governance and regulation in, of, and through digitally networked environments. His recent work has looked at the lived experience of credit scoring subjects, the search engine optimization (SEO) industry, and attempts at algorithmic regulation. At Cornell, he directs the Digital Due Process Clinic, a clinical research program that helps ordinary people cope with, understand, and challenge automated decision systems.

Valuation Studies 11(1) 2024: 9–37

Theme issue contribution

Angry Citizens and Black Belt Employees: Cascading Classifications of and around a Predictive Algorithm

Lise Justesen and Ursula Plesner

Abstract
Over past decades, predictive algorithms have been used extensively as profiling tools in the private sector, but today they are also increasingly entering public sector domains. This article builds on an ethnographic study of the development of a predictive algorithm in a debt collecting public sector organization. The algorithm was designed to profile citizens on the basis of their calculated 'readiness to pay' their debt and to guide employees' case handling according to 'type' of citizen. The article examines how the classification of citizens produced by the algorithm was mediated by different visualizations and by organizational actors who superimposed new and different classifications (moral and emotional) onto those provided by the algorithm. The article draws on the concepts of nominal and ordinal classification to identify how intended non-hierarchical classification glides into new hierarchical valuations of both citizens and employees. Classifications were 'cascading' – a concept the article develops to account for how classification of and around the algorithm multiplied and had organizational ripple effects. Based on empirical insights, the study advocates an agnostic approach to how algorithmic predictions impact work, organizations, and the situation of profiled individuals. It emphasizes a dynamic and unstable relationship between algorithms and organizational practices.

Keywords: predictive algorithms; valuation; classification; cascading classifications; profiling of citizens

Lise Justesen is Associate Professor at the Department of Organization, Copenhagen Business School.
Ursula Plesner is Associate Professor at the Department of Organization, Copenhagen Business School.

https://doi.org/10.3384/VS.2001-5992.2024.11.1.9-37
Hosted by Linköping University Electronic press
http://valuationstudies.liu.se

Introduction

Over past decades, predictive algorithms have been used extensively as profiling tools in the private sector, especially in marketing, banking, and finance (e.g., Fourcade and Healy 2013). Today, predictive algorithms are also increasingly entering public sector domains such as health care (Amelang and Bauer 2019), social work (Eubanks 2018), policing (Brayne 2017; Benbouzid 2019), and education (Jarke and Macgilchrist 2021). Predictive algorithms score, classify, and 'profile' people or organizations based on data sets, generating statistical estimations of their likely future behaviour. The present-day explosion in digital traces enhances the possibilities for the algorithmic sorting of people into classificatory schemes (Jürgenmeyer and Krenn 2016: 178) and these new 'classification situations' (Fourcade and Healy 2013, 2017) have significant consequences for individuals as well as for organizations.

Scholars have examined how predictive algorithms affect organizations when they become part of everyday work practices and how algorithmic classifications shape employees' behaviour and decisions. For instance, clients' personal risk profile may be calculated to set the price of their insurance (Cevolini and Esposito 2020, 2022), or a social worker may decide to intervene in a family situation because the children are classified by the algorithm as being 'at risk' (Eubanks 2018). Much research on predictive algorithms has suggested that they reduce employee agency and lead to a bypassing of the heuristics that employees otherwise usually apply when making decisions (Kellogg et al. 2020: 373). As such, algorithms are often portrayed as highly agential, leaving employees increasingly disempowered. Other studies focus on the embedded values and biases of algorithms and how these biases reproduce inequality and discriminatory practices (Friedman and Nissenbaum 1996; O'Neil 2017) that lead to 'algorithmic oppression' (Noble 2018: 4), to the stigmatization of profiled individuals, and to toxic feedback loops with performative effects (O'Neil 2017).

While much research has demonstrated the strong agency and often discriminatory effects of predictive algorithms, some scholars have begun to bring in more nuance to studies of algorithms and challenge what can be perceived as an almost deterministic, or at least too linear, account of algorithmic agency. These studies devote more attention to 'algorithmic assemblages' (Lee 2021) where human agency also plays an important role and they argue for a more agnostic, symmetrical, and empirically attuned approach to how algorithms work in organizational practices (Seaver 2017; Dudhwala and Björklund Larsen 2019; Lee and Björklund Larsen 2019; Lee and Helgesson 2020; Lee 2021). In this vein, Lee and Björklund Larsen (2019: 2) posed a note of caution: 'might we risk losing sight of the practices, negotiations, and human action that algorithms always are intertwined

with? Might we become so seduced by the algorithms that we forget the many social practices that surround them?'.

The point is not to deny that algorithmic classification and valuation are often very powerful and do structure 'life chances' in ways that reinforce inequality (Fourcade and Healy 2013). Yet, this literature highlights that predictive algorithms are many different things and that organizational contexts and human agency make a difference in relation to the algorithms' functioning and classifications, just like the specific design of the algorithms does. Taking inspiration from Lee and Björklund Larsen's question, this article advances a view attuned to the multiple classification and valuation practices that algorithms become entangled with in practice as people interact with them.

This article builds on an ethnographic study of a public sector organization ('the Center') that collects public debt. The Center developed and implemented a predictive algorithm designed to profile and sort citizens on the basis of their calculated 'readiness to pay' their debt and to guide employees' case handling according to the 'type' of citizen (cf. Deville 2012). Based on this study, we examine how classifications of the algorithms were moulded, reinterpreted, and modified in different ways to shape the organizational practices of which they became part. We develop the concept of 'cascading classifications' to account for how classifications of and around the algorithm multiplied. In this way, we theorize how classifications may condition each other and lead to new, sometimes surprising, or indirect classifications.

The concept of cascading helps us shed light on how classification of citizens in terms of their 'readiness to pay' became entangled with other classifications. Organizational actors superimposed new and different classifications onto those provided by the algorithm. The latter became entangled with classification of citizens in terms of motivation or attitude (who is willing to pay), the potential trouble they might cause (who is a 'difficult' person), or their emotional state (who is an 'angry' person). At the same time, employees had to be recategorized to match the algorithm's proposed citizen categories. Such an indirect organizational effect can also be captured by the concept of cascading classifications. To understand and qualify the many layers as well as the ambivalence of the cascading classifications, we draw on Fourcade's (2016, 2021) concepts of nominal and ordinal classifications, which particularly help us identify how intended non-hierarchical classification glides into hierarchical valuations of both citizens and employees.

Our analytical approach is inspired by Science and Technology Studies (STS) and Seaver's (2017) perspective on algorithms as 'sociomaterial tangles', which implies that "algorithms are not singular technical objects that enter into many different cultural interactions,

but are rather unstable objects, culturally enacted by the practices people use to engage with them" (Seaver 2017: 5). Along these lines, we treat both human and non-human actors as mediators that never merely transport effects but transform them along the way (Latour 2005: 128). While the basic premise of STS is that the technological and the social mutually shape each other, the technological side has sometimes been given too much weight in accounts that ascribe strong agency to algorithms while reducing the role of human actors. In our analysis, we pay attention to how algorithmic design, visual cues, and other material aspects became entangled with organizational practices and were recalibrated (Dudhwala and Björklund Larsen 2019), reinterpreted, and sometimes even ignored (Plesner and Justesen 2023) by human actors in the organization.

Since there is increased interest in using predictive algorithms, knowledge about how employees work with and around them is important – including how the classifications and valuations they imply affect the relationship between public employees and the citizens they are supposed to serve. The article contributes by theorizing the cascades of classification which surround the development and implementation of a predictive algorithm. With this concept, we are able to analyse how an algorithm's classifications – and hence the values inscribed in it – are mediated by actors within organizational practices and have organizational ripple effects.

Valuations and algorithms in organizational life

Algorithmic society

Recent work on valuation has given us vivid descriptions of how big data and algorithmic tools allow for new ways of tracing, sorting, assessing, and ranking individuals and organizations. In the words of Jürgenmeyer and Krenn (2016: 178), we are witnessing the "emergence of a valuation regime which exploits the ever more abundant digital traces of our everyday lives to algorithmically sort and slot people into classificatory schemes". According to Fourcade and Healy (2013, 2017), this has significant consequences for individuals as the emergence of new 'classification situations' shapes individuals' life chances by the proliferation of algorithmic scoring and decision making.

Fourcade and Healy portray classification (in their case, credit scores) as "an active, independent force that structures people's life chances" (Fourcade and Healy 2013: 569), thus depicting the algorithmic classifications as agents per se. Several studies have shown how algorithmic classifications are also implemented in settings they were not intended for, such as when credit scores are used to assess job candidates (Jürgenmeyer and Krenn 2016: 179; O'Neil 2017). This

kind of algorithmic creep may have serious consequences for the classified individuals and can lead to unequal service and treatment (Eubanks 2018).

Predictive algorithms and analytics differ from pre-programmed, deterministic ones that operate on the basis of a simple 'if … then' logic (Bucher 2018: 23). They build on often opaque processes using machine learning to analyse data, identify patterns (Burrell 2016), and make predictions about a likely future situation based on these data sets. In Bucher's terms (2018: 28), "machine learning is about strengthening the probability of some event happening, based on evolving information". Hence, predictive algorithms are concerned with possibilities and probabilities often expressed through calculation of risk scores and the profiling of individuals based on such quantification. These output scores are supposed to predict individuals' likely future behaviour, such as their risk of dropping out of school (Jarke and Macgilchrist 2021), their ability to repay a loan (Fourcade and Healy 2013), or the risk that they will commit a crime (O'Neil 2017).

Hopes are high regarding the usefulness of analysing big data and generating profiles of individuals' expected behaviour as a basis for decision making (Cevolini and Esposito 2020). However, the literature shows that organizational uses in practice are fraught with uncertainties. Predictive algorithms are often based upon limited and biased data (Jarke and Macgilchrist 2021) or dubious proxies (O'Neil 2017) that reinforce inequalities and discriminatory practices. While a critical literature highlights the problematic nature of an increasingly 'algocratic society' (Aneesh 2009), this approach leaves little room for human agency.

Enactment of algorithms in organizational everyday practices

Other studies have shown how it is precisely the uncertainties that allow for human agency and for mediation (Latour 2005) or even mitigation of some of the potential discriminatory effects. An increasing number of ethnographic studies demonstrate how algorithms are intertwined with everyday organizational life (e.g., Amelang and Bauer 2019; Dudhwala and Björklund Larsen 2019; Lee et al. 2019; Lee and Helgesson 2020; Plesner and Justesen 2023) and how they may have very different consequences. For instance, in their comprehensive review of the literature on everyday uses of algorithms in organizations, Kellogg and colleagues (2020) examined how algorithms produce new conditions for control in organizations. Drawing on labour process theory, they argued that employees are prompted to follow the recommendations of algorithms and act accordingly, even when, in principle, they have the autonomy to make a different decision.

Other organizational studies of algorithms portray them as more open to translation and interpretation, emphasizing the indeterminate outcome of using algorithms to solve particular tasks. These studies grant more agency to employees. Many of them are inspired by insights from STS, where sociomaterial sensibilities lead to a focus on how algorithms 'fold' heterogeneous data, methods, and objects with ethical and political effects (Lee et al. 2019), or where algorithms are viewed as enacted and as "the manifold consequences of a variety of human practices" (Seaver 2017: 4). Following the everyday lives of algorithms, as suggested by Neyland (2018), opens the way for analysing not only the ordered sets of instructions which comprise the algorithm, but also the various actions that the algorithm inspires.

Taking this approach, Dudhwala and Björklund Larsen (2019) showed how employees recalibrated the output suggested by algorithms when the output conflicted with their own knowledge, intuition, and judgement. They found that users often experienced a 'technological dissonance', i.e. a mismatch between their own expectations and the algorithm's output. This led employees to question the output and to 'recalibrate' it. Employees simply acted differently than the algorithm suggested. The recalibration of output could be based on different numbers from those provided by the algorithms, or on the employees' own experiences, or on their own contextual knowledge from elsewhere (Dudhwala and Björklund Larsen 2019: 11).

Amelang and Bauer (2019) demonstrated how a risk-predicting algorithm was embedded in everyday medical practices and gave rise to several translations and reactions. Staff members embraced some of the algorithmically based practices and resisted others. Here, the algorithm became both an external mediator *and* a source of authority. It was both used to reinforce arguments and was contested when it interfered with employees' intuitive grasp of the situation. Similarly, in their study of algorithms used in a laboratory for generating instructions for sample handling robots, Lee and Helgesson (2020) showed how employees made varying assessments of the procedures and outcomes of the algorithms. Algorithms were appreciated for their role in reducing human subjectivity in selection processes, but they were also criticized for destroying 'raw data'. Lee and Helgesson concluded that different 'styles of valuation' can coexist in the same organization around the same algorithms.

In a study of predictive algorithms in police work, Brayne (2017) described how massive amounts of heterogeneous data were constantly amassed in large databases, and based on the patterns identified, alerts were generated by algorithms. Brayne found that predictive algorithms did guide behaviour in some instances, while in other cases police officers would claim to have intimate knowledge that overruled the algorithm, e.g., about specific zones where they knew that crimes were

likely to take place. In such situations, the police officers considered the algorithmic recommendations to be superfluous or unreliable. Eubanks's (2018) study of predictive algorithms that were supposed to identify child neglect highlighted how employees were encouraged to be sceptical of the scores and rely on their own experience.

Both the literature on organizational uses of decision support algorithms generally and predictive algorithms specifically alert us to their multiplicity and different effects in practice. Importantly, they illustrate various ways in which the agency of algorithms is curbed by humans' pushbacks, recalibration, overruling, or deliberate neglect. Algorithmic calculations and outputs have disparate effects, depending on how they become entangled with human actors' interpretations and calibrations, and as such, their outcomes may be enacted in various ways. One important aspect of predictive algorithms, which has effects in organizations, is their entanglement with classification and valuation practices.

Classification and valuation practices

Classification and valuation are at the core of the design and functioning of predictive algorithms (Fourcade and Healy 2013; Fourcade 2016, 2021; Bucher 2018) and one approach to studying the entanglements of algorithms and organizational life is to focus on how algorithms classify. Predictive algorithms extend and transform classification practices in several ways. In their seminal work on infrastructure and classification practices, Bowker and Star (1999) argued that "classification schemes always have the central task of providing access to the past" (255). Predictive algorithms expand the temporality of classification schemes. Based on past data, they attempt to provide access not only to the past, but to the present and the future. Besides this, algorithms are not only based on prior organizationally produced categorizations. Building on designer input, algorithms are designed to produce 'their own' classifications whose logic sometimes escape even the designers of the algorithm, such as in machine learning (Burrell 2016). On the one hand, algorithms make classification more explicit because the algorithms formalize and standardize much of the tacit categorization work embodied in professionals' everyday heuristics and work practices. On the other hand, algorithms also make many of the specific choices and values invisible because functions of the algorithms tend to be opaque for their users (Burrell 2016). Some scholars have emphasized that rather than viewing complex algorithms as 'black boxes' (Pasquale 2015) that need to be 'opened up' for scrutiny, algorithms are never completely opaque or transparent (Lee 2021: 78). Instead, opaqueness and transparency are situated, enacted, and dispersed as part of specific assemblages (Lee 2021: 78).

Classification orders the world by dividing and grouping people or things in particular ways according to certain principles depending on purpose and context, and classification practices are always entangled with valuation (Kjellberg et al. 2013: 17). Valuation is a process (Kornberger et al. 2015), and valuation practices depend on classification as the basis for comparing and assigning worth to different people or objects (Lamont 2012). At the same time, it is well-established that classification practices as such are imbued with values and norms. In Fourcade and Healy's (2017: 287) terms, scores and classifications are "dual to one another", and scores are "categories all the way down".

For analytical purposes, however, it makes sense to distinguish between different ways of connecting classification and valuation. Taking inspiration from mathematics, Fourcade (2016; 2021) distinguishes between different principles of classification, which she refers to as nominal, cardinal, and ordinal. These classification practices are ideal types. In practice, they always overlap and intersect, implying that "much of social life around the world takes place at the intersection between judgements of kind and judgment of worth" (Fourcade 2016: 179).

Nominal categories are judgements of kind. Linnaeus's classification of plants in the eighteenth century is an example of classification based on kinds (Fourcade 2016: 176). Other examples could be when gender is described in binary terms or other identity categories are essentialized. Nominal classification establishes knowledge about essences by grouping together people or things with perceived resemblances and differentiating them from other kinds. As ideal types, nominal classifications are flat and horizontal and often appear as if they were natural and neutral differences, but practice looks different. Nominal classifications have often been imbued with inequality and discrimination, as in racism, sexism, etc.

Whereas nominal classification builds on a qualitative ontology, cardinal judgements are quantitative, aggregative, and compare different elements. Ordinal judgements are oriented towards commensuration based on relative ranks. In contrast to the horizontal ontology of nominal classification, ordinal classification is, by definition, vertical and tends towards scoring and quantitative commensuration (Fourcade 2016: 178). Fourcade elaborates:

> Unlike mere nominal difference, ordinal relations imply different valuations, a distinction of (at a minimum) two levels, highest and lowest, above and below. In the old Parsonian vocabulary, they are 'evaluative'. Unlike cardinal judgments, which are focused on magnitudes, ordinal judgments are interested in relative ranks, no matter the size of the difference. (Fourcade 2016: 178).

Hence, ordinalization involves rankings and tends to be competitive and fluid. Historically, ordinal classification has been tied to ideals of political liberalism and meritocracy because it judges individuals based on their performance instead of on their belonging to a certain kind (Fourcade 2016, 2021). Sorensen and Roberson (2020) illustrated this reorientation in their study of OECD education governance where they demonstrated how modes of comparison have shifted from nominal classification of countries (related to a 'modern' versus 'traditional' distinction) to ordinalization where countries are compared and ranked based on output indicators. This change marks a reorientation from 'being' to 'behaviour' as the foundation for judgement. However, as Fourcade (2021: 163) remarks, "ordinal citizenship often reproduces those very categorial inequalities it was meant to circumvent, albeit through different means". Digital technologies support the shift towards ordinalization as "computers are by nature oriented to sorting: they 'order' the world by spewing out priorities and queues" (Fourcade 2021: 162). In that sense, "digital citizenship […] dwells in ordinality" (Fourcade 2021: 162). Credit scoring is an example of this.

Increased interest in using predictive algorithms raises important questions about their uses in practice, including how employees work with and around them and how the classifications and valuations they imply affect the relationship between public employees and the citizens they are supposed to serve. The literature discussed in the two sections above on recalibration of algorithms and on different modes of classification provides a foundation for understanding situations where predictive algorithms become entangled with classification and valuation as well as with human agency in practice.

Empirical context and methodology

This article builds on an ethnographic study conducted in a Scandinavian public organization, pseudonymized as 'the Center', whose main function is to collect public debt (unpaid parking tickets, day-care bills, nursing home services, etc.) from citizens and to give advice on debt repayment options (e.g., dividing citizens' debt into monthly instalments). The Center was a small unit in a larger department within a big public organization. It was led by a team of managers and project managers and employed around 30 caseworkers. A significant part of employees' daily work consists of taking phone calls from citizens, clarifying their queries about their bills and debts, and advising them about repayment options. In some ways, the caseworkers' job resembles call centre work since, equipped with headsets at their desks, they take calls in sequence and are monitored in terms of processing time, client waiting time, and other performance measures well- known from other call centre contexts (Winiecki 2009).

However, employees and managers repeatedly emphasized that the centre should not be considered a call centre, and that efficiency and reduction of case processing time were never goals in themselves. In contrast to many call centre employees, who receive minimal training, the Center caseworkers are trained and skilled clerks with a broad knowledge of finance, legal regulations, IT systems, communication, etc. Rendering high quality casework and treating citizens fairly and equally were viewed as essential by both the caseworkers and their managers, who emphasized in interviews and at meetings that good casework often requires time and attention because the specific nature of each individual case demands careful consideration of the citizens' entitlements.

A goal of the Center is to encourage citizens to set up debt repayment agreements. Such agreements benefit the financial situation of the public organization and is supposed to make it easier for citizens to repay their debt. Sometimes citizens refuse to repay their debt (typically an unpaid parking ticket) because they think it is unfair. In other instances, the citizen's financial situation makes it difficult for them to repay the full amount at once. Therefore, the Center's employees and managers see it as a success if citizens can repay in monthly instalments. Such agreements can be concluded during the phone calls between staff and citizens. The high complexity and variation in the calls (regarding the types and amount of debt, individuals' financial and personal situations, attitudes, etc.) require good caseworker skills, including strong communication and people skills.

The Center had already digitalized many work processes. For instance, all cases were digitized and personal data, such as age, address, unpaid bills, photos of their car in the case of a parking ticket, memos from previous encounters with the Center etc., were readily available on employees' screens when a citizen contacted the Center. During our fieldwork, we witnessed how the Center management continuously sought to advance the digitalization agenda and to implement new digitalization initiatives, even in areas where technologies were untested and success uncertain, such as it was the case for the predictive algorithm, which is the focus of this article. A member of the management team with a background in the financial sector had first-hand experience with credit scoring and when a funding opportunity occurred within the overall organization, the Center applied for funds to develop a similar algorithmic tool for profiling citizens and matching them with the right employee. The Center relied on IT support from a unit in the larger department, so it turned to this IT unit for help in developing the algorithm. The operation of the Center as well as the municipal services in general obviously already relied on various algorithms (understood simply as computers' procedures for problem solving), but this project was

talked about and promoted as the first experiment with a *predictive* algorithm. The stated goal of the project was to improve the efficiency and quality of the casework by providing a better and faster service. Another motivation for developing the algorithm, however, was the wish to experiment with emerging technologies so as not to fall behind in the race towards more digitalized public organizations (Plesner and Justesen 2023).

Data collection and analysis

The ethnographic fieldwork took place over a period of 12 months in 2018–2019. This allowed us to follow the development and subsequent implementation of the algorithmic profiling project. Data were collected through participant observation, qualitative interviews, and review of relevant documents. We attended staff meetings and meetings of the management team. All caseworkers in the Center were supposed to be affected by the introduction of the algorithm, and we observed a selection of different caseworkers' everyday work, both before and after the introduction of the algorithm. We sat next to them with headsets, listening in real-time to their phone conversations with citizens and observing their screens during and between calls. In addition to numerous informal conversations, we twice conducted semi-structured interviews with eight caseworkers (before and after implementation of the algorithm except that one caseworker was interviewed only once as he resigned before the implementation) as well as with managers, project managers, and the IT staff responsible for the development of the algorithm. These interviews were recorded and transcribed, as were several of the meetings. In sum, we conducted 60+ hours of observation, including 278 phone conversations between caseworkers and citizens, conducted 22 interviews, and collected key policy and procedure documents as well as PowerPoint presentations.

Data were analysed by reading the entire material in several rounds, thereby familiarizing ourselves with the data before entering a thematic coding process (Braun and Clarke 2006). As this was an ethnographic study conducted over several months, and because it included different kinds of data, we ended up identifying many different themes that pointed in quite different directions. One cluster of themes related to categorization and this cluster was related to the algorithmic profiling project. In this cluster, we grouped data related to categorization of both citizens and employees, matching of citizens and employees, design choices such as algorithmic variables deployed to construct the categories, visualization of the different categories, etc. We analysed the data by focusing specifically on 'classification' and 'valuation', striving to be open towards many different types of both, and organizing the analysis to display the multiplicity of classification and valuation practices emerging throughout the Center's work with

the predictive algorithm. In the following analysis, we show the classification work involved in designing the algorithm and we outline how algorithmic outputs were recalibrated and led to multiple new classifications and valuations. In doing so, we highlight how valuation had cascading effects throughout the organization.

'It's about making the right match': Designing an algorithmic profiling tool

Developing the rationale for a predictive algorithm

The Center developed their new predictive algorithm to be used as a profiling tool for increasing knowledge about citizens and qualifying employee intervention based on this knowledge. The idea was that this would provide instant knowledge about callers and their likely future behaviour, whereby they could be matched with the caseworker who possessed the most suitable skills for handling precisely this 'type' of citizen. Based on the algorithmic profiling of each citizen who called the Center to clarify their debt situation, the algorithm would match citizens and caseworkers by automatically directing the call from a particular category of citizen to a particular category of employee. A Center manager described the purpose of the project in the following way:

> It's about making the right match. It's about giving the citizens the right service – the right matching of citizen and caseworker, thereby actually supporting the caseworker's job in an intelligent way. Whereas previously, they had to make their own judgment, like, 'hmm, this is the type of citizen I'm talking to' […]. Now, the majority of citizens who call will be sorted for you [by the algorithm].

The manager presented the algorithm as a way of providing knowledge about citizens by classifying them in a new way. It was intended to sort citizens for the employees, thereby supposedly replacing 'their own judgment'. In this manager's view, the previous, 'analogue', mode of knowing the citizen was a less valid source of knowledge. The different citizen types represented by the algorithmic profiling tool were considered to be pre-existent, but invisible or unclear, and the idea was that the algorithm could make these types visible, as would subsequent management intervention encouraging staff to utilize or adapt to the algorithm's profile categories.

Knowledge provided by classification of the algorithm were meant to become the basis for two types of intervention. One type was concerned directly with citizen service. It was assumed that knowledge of 'type' would enable better service because it would be the 'right

service' targeted precisely to this type of citizen. Another intervention was more indirect and organizational. To ensure the right match, the employees also needed to be categorized in a new way – the assumption being that employees had different skills, and some would be better than others at 'handling' certain types of citizens. It was also assumed that such knowledge would enable employees to apply particular skills and communication strategies matching a specific profile. Even if the employee already had access to ample digitized information about the individual caller, this new tool was supposed to provide them with an extra layer of information about the *type* of caller. The extra layer required a notification – a visual cue on the screen – that would signal a categorization instantly.

In interviews with the Center managers and project developers, they repeatedly stressed that the algorithm's classification of citizens were not intended to be hierarchical. The categories were not supposed to be ratings or rankings that assigned different worth to citizens. They were, it was emphasized, merely categories that would provide valuable information about how best to handle a particular citizen call. In that sense, they were presented as nominal rather than ordinal categories (Fourcade 2016). The algorithm was supposed to signal different citizen *types* in a *non-hierarchical* manner. Often explicitly invoking key public values, both employees and managers stressed that all citizens were entitled to a fair and equal service no matter how they were classified by the algorithm. No citizens were more worthy of receiving good service than others. In that sense, strong egalitarian values and a public sector ethos were combined with the design of the algorithmic project from the beginning.

Such values of equality and fairness were important not only for ethical and professional reasons, but also for legal ones. The organization's legal department was active in ensuring the legality of the predictive algorithmic project and prepared a document stating that in their interpretation, the project was compliant with legislation. If the project entailed discrimination or unequal service it would be illegal, but according to the assessment, the purpose of the project was to ensure that every citizen obtained the best service by matching them with that caseworker who was best suited to handle their specific case. It can be argued that there was a tension built into the algorithm from the beginning: citizens are classified but must be treated equally based on their individual case. As we will see, this ambivalence paved the way for recalibration as did the specific technological, communicative, and organizational design choices that had to be made, which will be illustrated in the following sections. It turned out that the initial classifications were followed by new, and sometimes surprising, classifications and valuations.

Entangled classifications: Ready to pay?

Despite the ambition of designing a non-discriminatory algorithm that would produce purely nominal classifications, the design work was imbued with values and norms from its inception. This was reflected in the design choices of which variables would be used (cf. Friedman and Nissenbaum 1996; O'Neil 2017) and as the algorithm was designed to be a predictive tool, it also had to be decided what exactly should be predicted. After some negotiation this became the likelihood that the citizen would, as it ended up being phrased, be 'ready to pay' his or her bills, and thereby the likelihood that a payment agreement would result from the phone conversation. The management team and the designers had lengthy discussions about how to find proxies for 'readiness to pay'. Which indicators would point to a citizen's future behaviour in terms of their likelihood to be 'ready'? What would a 'ready' or 'not so ready' person look like statistically? Several potential variables were considered, such as age, gender, number of children, marital status, and residential district as well as debt and payment history. It turned out that some of these variables could not be used for legal reasons, while others required data that was unavailable. These discussions were part of a process whereby citizens were to be grouped nominally, i.e. based on resemblances and shared characteristics.

Eventually, the citizen's debt and payment history over the past two years became one of the variables, and the algorithm was designed to operate with three main categories of citizens based on what management decided to call the citizen's 'readiness to pay' – after having considered their 'willingness to pay' as an alternative. The categories were then termed 'high readiness to pay', 'medium readiness to pay', and 'low readiness to pay'. A category called 'unknown' was added, reserved for citizens who could not be recognized by the algorithm. This classification scheme was based on the assumption that citizens' phone calls to the Center had to do with payment of unpaid bills and not with efforts made by the citizen to clarify or contest their debt status as such. In fact, our data show that clarifying mistakes made by the authorities was a frequent topic of the calls. The 'readiness' variable was also based on the assumption that citizens already had the *ability* to pay their debt. Their likelihood of repaying was considered to be a function of their 'readiness', making the issue one of motivation, although the more explicit criterion of 'willingness' had been rejected in the naming. Furthermore, there was a built-in assumption that the more money citizens owed the authorities, the less they could be expected to be ready to pay.

While the designers sought to construct a set of categories that could inform the caseworkers about the citizens' assumed readiness to pay, the 'readiness' classification quickly intersected with other classifications that became superimposed on the profiles. Early in the

design process, the persons classified were ascribed value as 'easy' (the ones in the 'high readiness' category), 'difficult' (the 'low' category), and 'worth spending extra time on' (the 'medium' category, where it was hypothesized that people could be convinced to pay if an effort was made by caseworkers). Here, nominal classifications seemed to glide into ordinal classifications, which contradicted the egalitarian values that were otherwise inscribed in the project. The very terms high, medium, and low suggest a hierarchy, and hence an ordinalization, and when this vertical classification became linked with notions of 'easy' and 'difficult' citizens, a seemingly technical classification became entangled with a moral evaluation.

Distinctions regarding citizens' 'readiness' were not only about semantics. The categories were supposed to influence the behaviour of the caseworker in the sense that they should adjust their response to match the citizen profile. As a rule, caseworkers were supposed to spend less time on citizens categorized as having 'low readiness to pay' because they were considered less likely to repay anyway. A project manager explained this logic:

> Okay, maybe it's fine to say that we need to talk less with these people [those classified by the algorithm as 'low readiness to pay']. Maybe this is where we can become a bit more efficient because we shouldn't waste half an hour on a call when we know that they usually don't pay and will never pay. In that case, it is better just to forward their call to [name of another public authority]. We just test to see whether they have changed their minds, because of course, we have the Public Administration Act, the legal stuff, and the equal service stuff [the law on citizen social services], but we need to become quicker at detecting if they want to pay – if they have changed their attitude and perhaps become more ready to pay and more willing to pay.

The tension between different classifications and valuations are clear because the project manager recognized the importance both of equal service and of giving people a chance to show that they had 'changed their attitude'. Still, an implication was that on average the caseworkers would spend less time on 'these people' and that the 'low readiness to pay' citizens should be approached differently from citizens in the other categories. The classification of assumed readiness became the basis for a new politics of differentiated treatment. In the quote, the project manager also presented the predictive algorithm as if it were a simple and deterministic algorithm rather than a machine learning-based probability model with relatively high statistical uncertainty. This understanding was reflected in the manager's description of the project's efficiency potential: "We think we can save some time there, because *we know they will never pay*, no matter the amount of tools and good conversations we have with them, so we

need to save time there, not in the 'want to pay' or 'will maybe pay'" (emphasis added).

Talking in statistical terms, the project manager explained that 'the likelihood that they will ever make an instalment plan is zero'. He suggested not spending much time on these citizens, but rather, to 'spend our resources on the medium group, where you may say that if we do it right, we can actually make them pay, whereas if we don't, they won't pay'.

Here, the category 'low readiness to pay' was described by the project manager as predictive of a citizen who will 'never pay', such that 'readiness' becomes a matter of 'willingness'. In many ways, this statement was puzzling as the project manager clearly knew about the uncertainties and probabilities of the algorithm. It seemed that even if the algorithm had primarily categorized a citizen on the basis of their payment history, a pretty solid narrative followed, portraying the citizen as someone who would never pay. Already in the design phase, valuations of the different categories adopted a moralizing tone and an indication of the development of a particular approach to some citizens is also illustrated in the following quote from the project manager:

> Again, in the red boxes, we don't want to listen to all that whining – bam – we just need to get to the point where we know if they want to pay or not. Because if you want to, we are very happy to help, but otherwise, we cannot be bothered.

This section points to cascades of classification and valuation practices that predictive algorithms generate in an organization. It illustrates how moral categories became entangled with the readiness to pay classification produced by the algorithm. The next section demonstrates a further step in this trajectory by showing how yet another new classificatory logic was prompted by visual translation of the categories.

Visualization and valuation

Based on the algorithm's classification of the caller, the phone call would be directed to a caseworker who was considered a good match for this particular type of citizen. If the algorithm should operate efficiently in informing the caseworkers about the caller's profile, the task would be to figure out how exactly to convey the information on 'type' in practice. The caseworker only had a few seconds before answering the call, so communication needed to be instant. The solution was to install pop-up icons on the caseworker's computer screen. When citizens called and entered their personal ID number, icons would immediately appear on the caseworkers' screens,

signifying the category of the caller and thereby his or her assumed 'readiness to pay'. This raised the more specific design question for the project management team about how to visualize the three different citizen categories. A project manager told us that they considered several options before deciding on a set of yellow emoticons. At first, they had considered traffic lights, and even different kinds of animals, but in the end, the management team and project managers decided to launch a competition among employees, asking them to come up with ideas for icons.

The selection of icons for the three 'readiness to pay' categories ended up being a happy smiley with a thumbs-up gesture for 'high readiness to pay', a semi-happy smiley for 'medium readiness to pay', and a frowning, thumbs-down emoticon for 'low readiness to pay'. In addition to the three different emoticons, the design team added a ghost icon to signify 'unknown citizens' and an icon showing a monkey, which would indicate that the system had made an error. As we will show in the next section, visualizing the different categories with these emoticons led to new types of classification that became absorbed into those already established through the design phase.

Classification of moods and temper: 'If I see it's the angry one, I put up my defenses'

Once the algorithm was up and running, caseworkers and managers began to discuss citizens using the algorithm's three main groupings (low, medium, or high readiness to pay). This discussion was prompted by the emoticons that now appeared on their screens, although the implementation was still only partial, and many calls would go through without pop-up icons. It turned out that the categorized citizens were not so much discussed in terms of their 'readiness to pay' or their actual financial situation. Instead, employees as well as their managers repeatedly talked about the profiled citizens' expected moods and temperament. The emoticons with various facial expressions led to translations of the initial classification. This seemed to be supported by the fact that even though the traffic light symbol had been discarded and the smileys all had the same yellow colour, there continued to be talk about some citizens being 'red'. One employee explained, 'if the red smiley comes up and indicates that this is a difficult citizen, you think, "Oh", and you take a sip of water before answering, and you may already be gearing up towards a tough conversation'.

The yellow frowning thumb-down emoticon was sometimes referred to as a 'red smiley' and conceived as a warning sign, where the caseworker should be prepared for trouble. In this way, a new classification of the caller's emotional state was superimposed on the 'readiness to pay' classification. In line with this, many employees

referred to the frowning emoticon as 'the angry one', associating the expression of the icon with the mood of the caller. A caseworker reflected on how the emoticons impacted her work: "[The emoticons] are helpful … I think it will be nice to be prepared for someone who is really angry. Because they can be really angry." Being prepared was important, she said, because it was very hard to be "yelled and screamed at".

The idea that icons helped the employees prepare mentally for the call was repeated by several caseworkers. One of them explained:

> you just need a split second to prepare mentally. If I see it's the angry one, I put up my defenses, I pay attention on a different level, I don't handle emails at the same time, I am fully focused. Because I know I need to pay attention, not necessarily because it's an angry citizen [on the phone], but because it can be a difficult case.

Here, the knowledge-intervention nexus departs from the rational and efficiency oriented ideal presented by managers where the knowledge provided by the algorithm would be about the statistically calculated likelihood that the person on the line would be ready to set up a payment agreement and the intervention should fit the classification of this likelihood. Instead, caseworkers interpreted the knowledge provided by the algorithm as knowledge about mood and temper, and their intervention was calibrated to handle possible emotional outbursts: 'Knowing' that the caller might be an 'angry one' led to increased mental focus, preparedness, and 'putting up one's defences'. Here, the encounter is imagined and described in affective terms (cf. Deville 2012).

Classification practices relating to assumed emotions show that the classifications inscribed in the algorithm during the design phase were not simply duplicated in practice. Rather, they evolved and became entangled with new concerns and different classifications and valuations. Employees superimposed their own classifications onto those of the algorithm, and employees' classifications seemed to be based on the visualization of the categories (the emoticons) rather than on the 'readiness to pay' terminology. Like the 'readiness to pay' classifications, the emotion classifications are behavioural categories where the nominal and ordinal intersect because although the emotional categories were not an explicit ranking of citizens (ordinalization), the emotional descriptions had moral undertones. In this sense, employees recalibrated the algorithmic output by ascribing different values to citizens. However, as the next sections show, this way of classifying citizens in terms of their emotions did not stand alone among the employees. They intersected with other ways of classifying citizens.

Everyday classification in action

The employees' mood categorizations of citizens were in many ways puzzling. During our fieldwork, we listened to more than 200 phone conversations conducted by different caseworkers, both before and after implementation of the algorithm. During these conversations, citizens were, in fact, very rarely angry. Callers might at times express frustration with their personal financial situation, or because mistakes had been made. In a few instances they even cried.

However, we saw no connection between these expressions of emotion among callers and the specific icons displayed on the employees' screens. Overall, our observation data showed that the vast majority of citizens was calm and polite, very often calling to clarify a question, to resolve some misunderstanding, or to establish a repayment agreement on their own initiative. The caseworkers were also friendly and polite and, in general, the encounters showed little tension. Many of the calls included laughter on both sides and ended with the citizen explicitly thanking the caseworker for helping them resolve the issue. When we discussed our observations with caseworkers in-between calls, they agreed.

This relative lack of tension in the actual encounters suggests that caseworkers deployed much more fluid classification in practice, nuancing them based on their experience and knowledge of the specific case. Our observations of the phone encounters with citizens showed that caseworkers paid little attention to the emoticons in these situations. Instead, they quickly tuned in on the caller's question by listening and asking clarifying questions, while at the same time quickly navigating through various digital documents and payment regulations on their double screens to resolve the issue.

As part of the Center's efficiency ambitions, employees had been provided with a set of standardized phrases and questions that varied according to the citizen categories, the idea being that communication with a citizen classified as having a 'low readiness to pay' should be different from communication with a citizen whose readiness to pay was classified as 'high' or 'medium'. As we saw in the previous sections, the goal was for caseworkers to spend less time on citizens classified as having 'low readiness to pay'. However, we never observed an employee use the script cards or the standardized phrases in practice. Instead, their interventions were adapted to the specific cases at hand.

'Supermen' and 'black belts': Classification of employees

The algorithmic profiling project also led to other types of categorization processes in the Center. These were not inscribed into

the algorithm as such, but instead an indirect consequence of the profiling project with organizational effects, derived from the fact that the project was intended to match citizens with the appropriately skilled caseworker. Hence, employees had also to be formally categorized in a new way. From listening to conversations among both managers and employees, we could detect that before the introduction of the algorithm, some kind of informal hierarchy already existed among employees. This informal ranking was based on how quickly and competently staff members were able to conclude their conversations with citizens and set up the much-desired repayment agreements. For instance, both before the algorithm project and during its implementation, one particular employee was frequently praised as the quickest and best phone agent by both colleagues and managers. It was also known that statistics on employee productivity were available, and in the open office setting, employees and managers could easily overhear how calls were handled.

But now that employees were to be placed in categories in order to be matched with citizens, the more informal hierarchies among employees in the Center were affected by the work with categorizing citizens. There was some confusion about how to carry out employee classification in relation to the citizen categorizations. Even managers disagreed about whether the categorization should be about degrees of employee competence. For instance, one manager used the term 'lowest level' when she described how the employees were divided into the new 'match groups'. She explained that the goal was not to keep employees in their groups, but to allow them to work with their competencies so that everybody would be able to take all kinds of calls. When all staff members would one day reach that level, it would only make sense to use the match groups for new employees: 'They enter at the lowest level and work themselves up through the systems' she explained. Another manager expressed the opposite viewpoint, emphasizing that classification of employees did not operate within a hierarchy of some being more skilled than others:

> We quickly found out that it was really important that we repeatedly said out loud, 'This is not good-better-best'. And when that was established and people realized that's how it was, then it was accepted. But in the beginning, it was a bit unpleasant because people thought they were being categorized in that way.

Even if the idea was described as matching citizens with 'the right' caseworker, the process of placing employees in different groups meant that they attached value to themselves and their colleagues. Employees used terms such as 'being supermen' or 'having the black belt'. For instance, when we asked an employee how the process of categorizing employees took place, she was not quite sure. She guessed that two of

the managers had tried to assess employees' skills, placing most of them in the middle group. She described the matching process as a typical ordinal classification: 'Well, I know that some are placed up there at the top, but I think … but I can also hear that sometimes when you need their help, they can … actually, they say precisely the same as the rest of us.'

When prompted by the interviewer to elaborate what being 'at the top' implied, the employee explained that "those who are supermen at the phones, if that makes sense".

Another employee explained how calls from the 'difficult citizens' would be directed to the most efficient employee mentioned above, the one considered a brilliant handler of all kinds of calls: 'The idea is that the difficult citizens are directed over to him so he can deal with them and convince them that they should pay'. As described by management, and by the efficient employee himself, this was not how the match was supposed to be carried out. Employees who were good at convincing citizens to set up instalment plans should deal with those in the 'medium readiness to pay' group and not waste their time on the 'low readiness' callers. But this was not the understanding of this employee:

> Interviewer: 'Which group do you think you will be placed in?'
>
> Employee: 'The second [medium], the … you know, I haven't reached the taekwondo black belt yet, that calls for years of experience, and I'm not there yet.'

Another employee described his competencies and categorization as follows:

> Employee: 'Yes, I can handle all types of calls […].'
>
> Interviewer: 'So, you are in the … what do they call that group?'
>
> Employee: 'They just call it black belt.'
>
> Interviewer: 'They call it black belt – would you say it's a kind of elite group?'
>
> Employee: 'Yes, that just sounds a bit stupid […]. We just call it black belt.'

Like managers, employees could shift between different employee classifications where the nominal and ordinal intersected. In one interview, an employee used the metaphor of steps on a staircase to portray a hierarchy, while at the same time insisting that this was not a matter of good, better, and best. In the first part of the interview, the

employee described how he was 'placed on a stair step', where he could advance: 'If I want to move up a step, it's easy for me to identify, to say that I would like to move up to this category instead.' Later, however, he portrayed the categorization of employees as non-hierarchical, emphasizing that this is not about climbing stairs:

> This is not about good, better, best in that sense, it's more about where I have my strengths. We have another colleague who can handle 100 cases per day when we are busy. He can handle all types of calls, without doubt, he can handle the most difficult ones when you need to talk to the citizen for a long time, but he really shines when it comes to the middle category, where citizens need a gentle push to agree to set up an instalment plan. He is much better than me in that category, which doesn't mean that this category is easier, it's just another tool that is needed in that conversation, so it's not about good, better, best, in that staircase sense.

These examples show how both managers and employees made attempts to entertain the idea of purely nominal (non-hierarchical) classification, where each employee was a type with different, but equally valuable skills. At the same time, however, it was notable how the ordinal (hierarchical) classification of citizens in terms of their difficulty led to a corresponding ordinal classification of employees. In this sense, the hierarchical imagery entered into discussions of how employees were to be allocated to different types of citizens, especially the 'difficult' ones. 'Angry citizens' should be matched with 'blackbelt employees'. We understand this as a cascading effect: the valuations of citizens led to new valuations of employees, even though these valuations were neither inscribed into the predictive algorithm, nor a direct result of its operation.

Discussion

This article has explored how a predictive algorithm aimed at profiling citizens and matching them with appropriate caseworkers became entangled with different classification and valuation practices when it was implemented in a public sector organization. As is the case with any algorithm, the classifications were based upon values and normative assumptions related, e.g., to the choice of proxies for 'payment readiness'. As such, our findings are in line with other studies showing how algorithms have values inscribed into them by their designers (e.g., Friedman and Nissenbaum 1996; O'Neil 2017). However, the contribution of this article lies in elaborating how the algorithm's classifications – and hence the values inscribed in it – were mediated by actors in organizational practices. To describe how human agency and algorithmic classifications interact and become entangled

with one another, and constitute new classification situations, we develop the concept of cascading classifications.

Cascading classifications

The term 'cascade' has two key connotations. It may refer to large amounts that occur at once, as in 'cascades of water' (Oxford English Dictionary). At the same time, cascading also denotes a sequential movement conveyed by the image of a waterfall or a sequential linking of elements (Oxford English Dictionary). In an STS context, Latour (1986) talked about cascades of inscriptions, but without conceptualizing cascades as such. Ruppert and colleagues (2013: 31) refer to Latour's usage when they write that "It is through such cascades of inscriptions – for instance from reams of data to indices – that simpler and more mobile digital inscriptions are often generated". However, based on our findings, we can elaborate on the concept of cascading to better understand the opposite movement, whereby the 'simple' classifications are cascading and thereby become multiplied rather than simplified. In our case, cascading is the output of rather than the input to predictive algorithms.

Our analysis demonstrated how classification of citizens in terms of their 'readiness to pay' became entangled with a cascade of other classifications. Organizational actors superimposed new and different classifications onto the ones provided by the algorithm when citizens were classified in terms of motivation or attitude (who is willing to pay), the potential trouble they might cause (who is a 'difficult' person), or their emotional state (who is an 'angry' person). In these classifications the nominal and the ordinal intersected and they involved moral evaluations, albeit always in an ambivalent manner. The ambivalence was related to the fact that a strong public sector ethos and values of fair and equal treatment had a major influence on caseworkers' everyday interactions with citizens. Democratic values of equality and fairness intersected with undertones of moral judgements about unwilling, difficult, or angry citizens, and those were talked about as not always worth spending time on. The moral evaluations about indebted citizens resonate with research on debt: there is always a moral aspect to any debt relationship (Fourcade and Healy 2013; Fourcade 2021). Referring to Nietzsche, Fourcade (2021: 163) writes that the domain of credit and debt 'is one of the most potent sites for the social distribution of feelings of superiority, moral desert, shame and guilt'. However, while managers and employees in our study sometimes voiced their moral judgements, such valuations were also contested by caseworkers, who emphasized public sector values and the right of all citizens to fair and equal service as a principle of good casework practice. In such cases, the guiding valuation principle was the maintenance of a strong public sector ethos and in the concrete

situations and encounters with citizens' caseworkers often used their room for manoeuvre to ignore the 'readiness to pay' categories as well as the visual cues.

One type of cascading effect implied that citizens were routinely talked about as 'happy' or 'angry'. Another type of cascading effect led to employees being talked about as 'having the black belt' (or not). The latter was one of several indirect organizational effects of the introduction of the algorithm. Not only did it produce new ways of classifying citizens, it also led to new categorizations of employees as they became linked to their respective 'match group'. The numerous new classification activities revolving around the introduction of the algorithm can be thought of as processes of making the organization 'algorithm-ready'. These insights about classification and valuation practices related to predictive algorithm contributes to the literature that challenges tech determinist approaches (e.g., Dudhwala and Björklund Larsen 2019; Lee and Björklund Larsen 2019; Lee and Helgesson 2020; Lee 2021), helping us understand in more detail how human agencies become entangled with the digital, with organizational effects.

Human agency and enactment of the algorithm

Hype around the expansion of algorithmic society might lead us to overlook the uncertainties and grey zones that allow room for human judgement and recalibration (Dudhwala and Björklund Larsen 2019). While several studies have shown how some predictive algorithms structure people's life chances in credit institutions or other settings (Fourcade and Healy 2013), or how they can lead to algorithmic oppression (Noble 2018), many organizations still find themselves at the 'data frontier' (Beer 2019) where hopes are highly inflated compared to everyday organizational realities (see also Plesner and Justesen 2023).

Our study has illustrated that predictive algorithms in some contexts end up being less powerful than commonly assumed in the critical literature. Algorithmic output can be enacted in many ways (Seaver 2017) and thereby be subject to significant recalibration. Our analysis showed that sometimes the predictive algorithm sparked caseworkers' interpretation and activities, whereas at other times, the algorithm's classification was downplayed or completely ignored (cf. Plesner and Justesen 2023). Employees were not blindly prompted by the icons in casework practice, nor were they left disempowered or frustrated, as the literature emphasizing the close link between algorithms and organizational control has argued (Kellogg et al. 2020). Instead, employees would still deploy their own judgement of the situation, often overruling the algorithmic prompts and carefully designed icons. Hence, we found that the original classification

situation was less determinant than demonstrated in other case studies of algorithmic prediction (Fourcade and Healy 2013). While there can be no doubt that predictive algorithms can lead to discrimination, oppression, and increased inequality (Fourcade and Healy 2013; O'Neil 2017; Noble 2018), our study shows that there is more to algorithms than steering and oppression. Algorithms in practice can be many things. Algorithms are sociomaterial tangles (Seaver 2017) with different effects depending on context. As such, actors may not only mediate but also mitigate discriminatory classifications and consequences of the algorithm in some situations.

While our study has pointed to the role of human actors in recalibrating algorithmic output, this does not imply that the algorithm was open to any sort of interpretation. Specific design choices shaped the ensuing classification processes, involving a cascading of classification. Emoticons that functioned as visual indicators involved a translation of the 'readiness to pay' classifications into emotional categories. People's readiness to pay morphed into a characterization of people as either willing or uncooperative persons. Such design implications are well-known from other algorithmic projects, such as Amelang and Bauer's (2019) study of risk scoring in the health-care sector. In their case, developers attempted to avoid symbols such as traffic lights because they were considered 'too judgmental' (Amelang and Bauer 2019: 484). Our study also shows that despite similar attempts to avoid judgementalism, visualization remains open for resignification and becomes entangled with valuation.

Concluding remarks

The article was motivated by the spread of predictive algorithms into ever more contexts but offers an alternative to alarmist and tech determinist accounts of the effects of such algorithms. Drawing on an ethnographic study, it demonstrated how algorithms are enacted differently in different contexts and how actors may recalibrate algorithmic output, classification, and valuation. While values and normativities are always inscribed in predictive algorithms, their effects are not determined, but emerge in the classification work surrounding it in specific situations. With the focus on classification and valuation practices, our study extends the critique of 'monolithic accounts' of technologies (Lee and Helgesson 2020) that tend to underestimate the agency of technology users and which portray the values of algorithms as being blindly duplicated in practice without any sort of mediation, distortion, or resistance. We contribute to the literature by showing in empirical detail how classifications of and around an algorithm were enacted in multiple ways and how nominal and ordinal classification intersected in these processes. We theorized

this as cascading classifications, by which we mean classifications that multiply around an algorithm in ways that are neither deterministic nor arbitrary; they may be prompted by visual designs, as when emoticons prompt classifications of emotion, or they may emerge when employees are recategorized to fit the logic of an algorithm.

Our study is based on a single ethnographic study, but it may inspire future research to similarly pay attention to differentiated, coexistent, dynamic, and cascading classifications circulating among algorithms, employees, and managers. Classification and valuation have a 'career'. They can develop in surprising and sometimes internally inconsistent ways, and we need more empirical knowledge about how this unfolds in different empirical contexts. Much of the literature on predictive algorithms has relied on private sector cases. The present study contributes to our understanding of public sector organizations' adoption of this type of tool, and we suggest that given the spread of predictive algorithms in the public sector, future research should pay more explicit attention to the public sector as a particular context of digitalization (Plesner et al. 2018; Plesner and Justesen 2022) since it can be expected that a public sector ethos colours the valuations attached to predictive algorithms.

Acknowledgments

We would like to express our gratitude for the inspiration we got from our colleague José Ossandón in the early stages of writing this article, and for the extremely helpful and developmental comments from our editor, Jacob Reilley, who nudged and helped us to develop our conceptualization of cascading classifications. We also want to thank Andreas Sundström for inviting us to present and discuss this work while in progress at a seminar in the Accounting section at Stockholm University and to Veronica Casarin for a similar opportunity at the Accounting and Management Control Seminar Series at ESSEC in Paris.

References

Amelang, Katrin, and Susanne Bauer. 2019. "Following the Algorithm: How Epidemiological Risk-Scores Do accountability". *Social Studies of Science* 49(4): 476–502.

Aneesh, A. 2009. "Global Labor: Algocratic Modes of Organization." *Sociological Theory* 27(4): 347–370.

Beer, David. 2019. *The Data Gaze. Capitalism, Power and Perception*. London: Sage.

Benbouzid, Bilel. 2019. "Values and Consequences in Predictive Machine Evaluation. A Sociology of Predictive Policing." *Science & Technology Studies* 32(4): 119–136.

Bowker, Geoffrey C., and Susan Leigh Star. 1999. *Sorting Things Out: Classifications and its Consequences*. Cambridge, MA: MIT Press.

Braun, Virginia, and Victoria Clarke. 2006. "Using Thematic Analysis in Psychology". *Qualitative Research in Psychology* 3(2): 77–101.

Brayne, Sarah. 2017. "Big Data Surveillance: The Case of Policing." *American Sociological Review* 82(5): 977–1008.

Bucher, Tania. 2018. *If … Then. Algorithmic Power and Politics*. Oxford: Oxford University Press.

Burrell, Jenna. 2016. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society* 3(1):1–12.

Cevolini, Alberto, and Elena Esposito. 2020. "From Pool to Profile: Social Consequences of Algorithmic Prediction in Insurance." *Big Data & Society* 7(2): 1–11.

Cevolini, Alberto, and Elena Esposito. 2022. "From Actuarial to Behavioural Valuation. The Impact of Telematics on Motor Insurance". *Valuation Studies* 9(1): 109–139.

Deville, Joe. 2012. "Regenerating Market Attachments. Consumer Credit, Debt Collection and the Capture of Affect." *Journal of Cultural Economy* 5(4): 423–439.

Dudhwala, Farzana, and Lotta Björklund Larsen. 2019. "Recalibration in Counting and Accounting Practices: Dealing with Algorithmic Output in Public and Private." *Big Data & Society* 6(2): 1–12.

Eubanks, Virginia. 2018. *Automating Inequality. How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St Martin's Press.

Fourcade, Marion. 2016. "Ordinalization: Lewis A. Coser Memorial Award for Theoretical Agenda Setting 2014." *Sociological Theory* 34(3): 175–195.

Fourcade, Marion. 2021. "Ordinal Citizenship". *British Journal of Sociology* 72(2): 154–173.

Fourcade, Marion, and Kieran Healy. 2013. "Classification Situations: Life-Chances in the Neoliberal Economy." *Accounting, Organizations and Society* 38(8): 559–572.

Fourcade, Marion, and Kieran Healy. 2017. "Categories All the Way Down." *Historical Social Research* 42(1): 286–296.

Friedman, Batya, and Helen Nissenbaum. 1996. "Bias in Computer Systems." *ACM Transactions on Information Systems* 14(3): 330–347.

Jarke, Juliane, and Felicitas Macgilchrist. 2021. "Dashboard Stories: How Narratives Told by Predictive Analytics Reconfigure Roles, Risk and Sociality in Education." *Big Data & Society* 8(1): 1–15.

Jürgenmeyer, Julian, and Karoline Krenn. 2016. "Classification Situations." *Valuation Studies* 4(2): 177–189.

Kellogg, Katherine, Melissa A. Valentine, and Angéle Christin. 2020. "Algorithms at Work: The New Contested Terrain of Control." *Academy of Management Annals* 14(1): 366–410.

Kjellberg, Hans, Alexandre Mallard, Diane-Laure Arjaliès, Patrik Aspers, Stefan Beljean, Alexandra Bidet, Alberto Corsín, 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." *Valuation Studies* 1(1): 11–30.

Kornberger, Martin, Lise Justesen, Anders Koed Madsen, and Jan Mouritsen. 2015. "Introduction." In *Making Things Valuable*, edited by Martin Kornberger, Lise Justesen, Anders Koed Madsen, and Jan Mouritsen (1–17). Oxford: Oxford University Press.

Lamont, Michèle. 2012. "Toward a Comparative Sociology of Valuation and Evaluation". *Annual Review of Sociology* 38(1): 201–221.

Latour, Bruno. 1986. "Visualization and Cognition: Thinking with Eyes and Hands." *Knowledge and Society* 6: 1–40.

Latour, Bruno. 2005. *Reassembling the Social*. Oxford: Oxford University Press.

Lee, Francis. 2021. "Enacting the Pandemic." *Science & Technology Studies* 34(1): 65–90.

Lee, Francis, and Lotta Björklund Larsen. 2019. "How Should we Theorize Algorithms? Five Ideal Types in Analyzing Algorithmic Normativities." *Big Data and Society* 6(2): 1–6.

Lee, Francis, and Claes-Fredrik Helgesson. 2020. "Styles of Valuation: Algorithms and Agency in High-throughput Bioscience." *Science, Technology, & Human Values* 45(2): 659–685.

Lee, Francis, Jess Bier, Jeffry Christensen, Lukas Engelmann, Claes-Fredrik Helgesson, and Robin Williams. 2019. "Algorithms as folding: Reframing the analytical focus." *Big Data & Society* 6(2): 1-12.

Neyland, Daniel. 2018. *The Everyday Life of an Algorithm*. London: Palgrave Macmillan.

Noble, Safiya Umoja. 2018. Algorithms of Oppression. *How Search Engines Reinforce Racism*. New York: New York University Press.

O'Neil, Cathy. 2017. *Weapons of Math Destruction – How Big Data Increases Inequality and Threatens Democracy.* London: Penguin Books.

Pasquale, Frank. 2015. *Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Plesner, Ursula, and Lise Justesen. 2022. "The Double Darkness of Digitalization: Shaping Digital-Ready Legislation to Reshape the Conditions for Public Sector Digitalization". *Science, Technology & Human Values* 47(1): 146–173.

Plesner, Ursula, and Lise Justesen. 2023. "Digitalize and Deny: Pluralistic Collective Ignorance in an Algorithmic Profiling Project." *Ephemera* 23(1): 19–48.

Plesner, Ursula, Lise Justesen, and Cecilie Glerup. 2018. "The Transformation of Work in Digitized Public Organizations." *Journal of Organizational Change Management* 31(5): 1176–1190.

Ruppert, Evelyn, John Law, and Mike Savage. 2013. "Reassembling Social Science Methods: The Challenge of Digital Devices." *Theory, Culture & Society* 30(4): 22–46.

Seaver, Nick. 2017. "Algorithms as Culture: Some Tactics for the Ethnography of Algorithmic Systems." *Big Data & Society* 4(2): 1–12.

Sorensen, Tore, and Susan Robertson. 2020. "Ordinalization and the OECD's Governance of Teachers." *Comparative Education Review* 64 (1): 21–45.

Winiecki, Donald. 2009. "The Call Center and its Many Players." *Organization* 16(5): 705–731.

Lise Justesen is Associate Professor at the Department of Organization, Copenhagen Business School. Her research interests include the interplay between management technologies and organization, public sector digitalization and actor-network theory. She has published in journals such as *Accounting*, *Organizations and Society* and *Organization* and she is the co-editor of the volume *Making Things Valuable* published at Oxford University Press. She is co-PI of the research project Valuing Invisible Work: Efficiency ambitions and digitalization projects in practice.

Ursula Plesner is Associate Professor at the Department of Organization, Copenhagen Business School. She conducts research on digitalization and organization, mostly in a public sector context. Her work has been published in journals such as *Organization*, *Strategic Organization*, and *International Journal of Management Reviews*. She is the co-author of *Digital Organizing – Revisiting themes in organization studies*. She is co-PI of the research project Valuing Invisible Work: Efficiency ambitions and digitalization projects in practice.

Valuation Studies 11(1) 2024: 38–59

Theme issue contribution

Digital Valuation: Lessons in relevance from the prototyping of a recommendation app

Celia Lury, Sophie Day, Andre Simon, Martín Tironi, Matías Valderrama and Scott Wark

Abstract
This article describes the use of a prototype recommendation app to explore how users are included and/or excluded in categories of various kinds of 'People Like You'. In the study, interviews with users of the prototype app indicate that the experience of receiving personalized recommendations is routinely evaluated in terms of relevance, that is, as either of interest to them or as beside the point, as accurate or inaccurate, with accuracy often understood as recognition of their context(s). We build on the interviews to develop an analysis which suggests that the capacity of recommendation systems to make relevant recommendations is a function of the parallel projections – from the app on one side and users on the other – that are made as part of an interaction order. In developing this analysis, we reflect on the implications of the interaction order for the inclusion and exclusion of users in categories or kinds of people. We highlight the importance of the temporal formatting of interaction as a continuous present for the relation between belonging and belongings, and thus for the creation of a datasset (Beauvisage and Mellett 2020).

Keywords: digital valuation; prototype; relevance; recommendation systems; classification

Celia Lury is Professor in the Centre for Interdisciplinary Methodologies at the University of Warwick.
Sophie Day is Professor Emerita at Goldsmiths, University of London.
Andre Simon is co-creator of the app Big Sister.
Martín Tironi is Director and Associate Professor at the School of Design of Pontificia Universidad Católica de Chile.
Matías Valderrama is PhD Fellow at the University of Warwick.
Scott Wark is Lecturer in Media Studies at the University of Kent.

https://doi.org/10.3384/VS.2001-5992.2024.11.1.38-59
Hosted by Linköping University Electronic press
http://valuationstudies.liu.se

Introduction

The instruments and methods deployed in digital economies provide a set of devices that create new opportunities for valuation. One especially important set of such opportunities is associated with recommendation systems, and their capacity for classifying people and things. For example, Marion Fourcade and Kieran Healy say that 'digital economy's classificatory architecture allows market institutions to apprehend their clients, customers, or employees through new instruments of knowledge, efficiency and value extraction' (2017: 10). Such systems are everywhere: in earlier research we (Lury and Day 2019) showed that the dynamic classification of 'People Like You' who 'Like Things Like This' is central to the now ubiquitous practices of online personalization. As Kris Cohen observes:

> We find recommender systems in search engines, in dating sites, in shopping, in social media feeds like Facebook's, in streaming music services, and, increasingly, at every point of networked interaction. In fact, unless one tries to turn off these personalization engines, which isn't always possible, it's now often harder to find a nonpersonalized environment online.[1] (2019: 173)

This article builds on our previous research to describe the ways in which the processes of classification associated with the personalizing practices of recommendation systems provide opportunities for digital valuation by describing a study involving the prototyping of a recommendation app (https://algorithmicidentities.net/). In the study, interviews with users of the prototype app indicate that the experience of receiving personalized recommendations is routinely evaluated in terms of relevance, that is, as either of interest to them or as beside the point, as accurate or inaccurate, with accuracy often understood as recognition of their context(s). To explore the significance of this finding for valuation practices we build on the interviews to develop an analysis which suggests that the capacity of recommendation systems to make recommendations is a function of the parallel projections – from the app on one side and users on the other – that are made as part of an interaction order (Goffman 1983). In developing this analysis of our participants' experience, we draw on Karin Knorr Cetina's (2009) description of synthetic situations, that is,

---

[1] In contrast to the prevalence of recommendation systems in many other countries, Clause 18 of China's first E-commerce Law (issued on August 31 2018 and effective from January 1 2019) asserts: "When e-commerce operators provide search results of goods/services to consumers based on their consumption interests and habits, options not targeting their personal characteristics should also be provided so as to protect consumers' legitimate rights and interests" (http://www.lawinfochina.com/display.aspx?id=e0c468f6d44d5b50bdfb&lib=law; Han Wen, personal communication, March 2020)

situations in which there is human interaction with on-screen computational projections.

We found her account of such situations helpful not only because it acknowledges human–nonhuman interaction but also because it draws attention to encounters in time, with Knorr Cetina arguing for the importance of time integration formats as a means of coordinating interaction in situations in which participants are not physically co-present. Such formats, she says, need to be ongoing: a synthetic situation's 'assemblage and projection is a continuous project' (2009: 70). In our study, we found that the interaction order of the making and taking of recommendations not only involved complex, ongoing coordination but also opened up opportunities for different kinds of valuation practice. We show that the turn-taking in such situations involves the projection of subject–object and part–whole relations between users and the app, with these relations offering different possibilities for users to experience a sense of belonging and making it difficult for them to translate a sense of belonging into belongings (Cooper 2007), that is, into ownership, (self-)possession or property. In contrast, however, we suggest that the organization of these relations in the time integration format of a continuous present affords the owners of the app the possibility of assetization.

Big sister: The relevance of prototyping

The investigation of recommendation apps on which this article is based was conducted by an interdisciplinary and international team including researchers from sociology, anthropology, design, media studies and computer science in Chile and the UK. Together we developed a recommendation app using a prototyping methodology. The reasons we chose to use prototyping as a methodology included that, as a first or original typing,[2] it is a practice that invites – indeed perhaps requires – reflection on the process of typing or classification, and does so in a way that draws attention to how this process is organized in time.

As it has developed in the discipline of design for example, prototyping describes a process of research and development leading to the production of a product or service that is a specific instance of a type or class of object: this is an iterative process of modification and revision that generally finishes when the object is brought to market as

[2] The etymology of the term is 'c. 1600, from French *prototype* (16c.) and directly from Medieval Latin *prototypus* "original, primitive", from Greek *prōtotypon* "a first or primitive form", noun use of neuter singular of prōtotypos "original, primitive," from *prōtos* "first" (see proto-) + typos "impression, mould, pattern" … In English from 1590s as prototypon' (https://www.etymonline.com/word/prototype).

a commodity. In recent years, this process has come to be tied to the practices not only of designers and producers but also of consumers or users. The involvement of users typically happens through the representation or projection of their needs or concerns in a variety of intermediary professional practices, as in the development of user personas in market research, branding and User Experience Design (UXD). However, this process is increasingly also associated with the automated collection and analysis of data relating to user preferences, with many recommendation systems employing a combination of these practices.

In some business and software development practices, prototyping has also come to be linked to a principle of 'perpetual beta' (O'Reilly 2005), in which an object is seen as never finished or complete, but as consistently open to version-ing, as producers respond to the employment that users make of an object or product, including adaptations and customizations (Nieborg and Poell 2018).[3] Adoption of the principle of perpetual beta is not necessarily about detaching products from markets however but, rather, a way for producers to recognize the dynamism of markets and the extent to which products and services are co-produced with consumers or users. In these practices, prototyping is not understood exclusively in relation to the fixed end or goal of creating a (new) type of object, but in terms of continuously respecifying demand by typing or classifying subjects as particular kinds of users (Woolgar 1990; Clough 2018).

As science and technology studies of human–computer interaction have demonstrated, the practice of specifying or shaping the subject(s) or user(s) has become an increasingly central concern in prototyping practices, particularly through forms of embedded and enacted scripting (Akrich 1992). Indeed, this kind of technical scripting plays a key role in the production of value in relation to the data-intensive forms of 'controlled consumption' that apps facilitate (Andersen and Pold 2014). In this use of prototyping, it is no longer only the object that is the product or commodity but also the class or type of user or subject that can be associated with the object. In the case of recommendation apps, classes or types of user – instances of which might be called 'People Like You' – can themselves become a product to be brought to the (multi-sided) market, to be sold on to third parties, including advertisers.

There is often a kind of twisted, dynamic looping effect here, so our previous research suggests: more specifically, categories of 'People Like You' who 'Like Things Like This' are continually projected onto 'Things Like This' that 'Like People Like You' and vice versa as users respond (or not) to recommendations emerging from automated calculations (Lury and Day 2019). That prototyping may involve the

[3] In this sense, a prototype is objectual (Knorr Cetina 1997), that is, an object that is never closed, complete or final.

ongoing respecification of mutually informed projections running in parallel made it even more suitable for our methodological purposes, since it amplifies the possibilities that our prototype could be employed as a political device 'that can make visible (or invisible) certain entities and issues, determining what the experimental entities can do and say' (Suchman et al. 2002; Tironi 2020).

Big Sister, as we called the app, uses data from a user's social media accounts to generate a profile of personality traits using a form of machine learning based on natural language processing, as well as providing music recommendations linked to this profile. More specifically, we designed Big Sister to gather posts from a user's Twitter and Facebook accounts or, in a deliberate diversion from most recommendation apps and specifically to supplement or replace the conventional reliance on social media, from additional texts written or selected by the user. These posts and/or texts are then used to generate a profile of personality characteristics based on the IBM Watson™ Personality Insights service, with the characteristics indicated graphically on the app by positions along a number of 'bars' representing continuums of personality traits such as agreeableness, conscientiousness and neuroticism, as well as recommendations for songs in genres deployed by Spotify. To encourage reflection, Big Sister allowed users to compare their results by date and information source on an interactive graphic and see how personality predictions and recommendations change between social media platforms and over time. In later versions we added a further option for users to modify input texts or exclude tweets or posts to see how their results changed. We also incorporated the ability to compare their personality results with those of public figures including the UK's Queen Elizabeth II and the Chilean poet Gabriela Mistral. In a still later version, we added the option of the user composing Frankenstein-type texts, combining a sequence of fragments from different well-known authors, breaking with Personality Insight's assumption of the user as a single, unified subject.

The name Big Sister was chosen with a nod towards George Orwell's Big Brother as well as IBM's nickname – Big Blue, while recognizing that, as Armand Mattelart and André Vitalis (2015) observe, it is the multitude of 'little sisters' (or little analytics, Amoore and Piotukh 2015) that collectively work together to produce recommendations. We hoped the name would indicate membership of a family of digital devices while signalling aspects of both surveillance and companionship associated with apps (Woods 2018).[4] We also

[4] One UK interviewee remarked, 'of course there's a lot of connotation with … the name … but I kind of … it's kind of funny. So it's kind of like a … [laughs] kind of like a joke' (UK Participant 7).

created a visual identity to give the app a persona, including a stylized representation of an eye (see Figure 1).

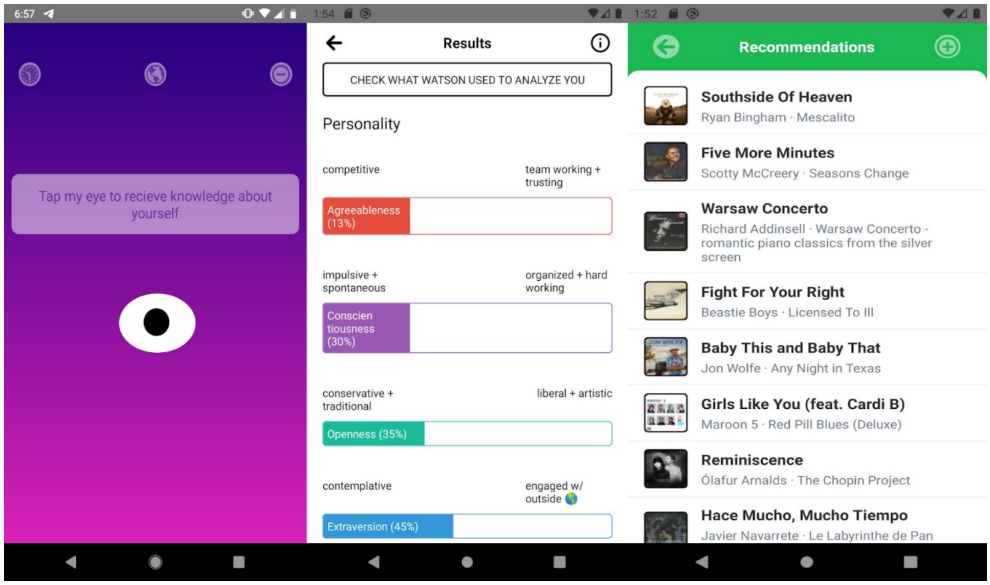

Figure 1: Screen shots from the first beta version of the Big Sister app
Source: Photo by author Matías Valderrama

Another element of the prototyping methodology was a Big Sister 'kit' (see Figure 2), which enabled users or participants to create three dimensional visualizations of the profiles and recommendations provided by the app. The kit included a board, with an abstract representation of a person at the centre, radiating outwards. In the trace interviews (Dubois and Ford 2015) we conducted following a period of 2–4 weeks use of the app, we asked participants to position post-its nearer or closer to the central figure during the interview to indicate perceived accuracy of recommendations and their personality profile.[5] Our questions were open-ended, and invited participants to reflect on their use and experience with the app. The aim was to enable a form of co-analysis of the results of the use of the app by us as the researchers and the participants (Latzko-Toth et al. 2017: 203). The preliminary analysis of a first round of interviews fed back into the design of later versions of the app, with additional features designed not only to ease its use but also to expand the possibilities of users testing the app itself.

[5] Dubois and Ford argue that the use of visual materials in interviews is 'useful for enhancing recall, validating trace data-generated results, addressing data joining problems, and responding to ethical concerns that have surfaced in the current era of surveillance and big data' (2015: 2067).

Figure 2: The *Big Sister* kit
Source: Photo by author Matías Valderrama

Prototyping relevance

While as to be expected, the experience of participants with the app was very varied (see Tironi and Valderrama 2021 for a discussion of Big Sister as a problematizing and decolonial prototype); the experiment seemed to lead some of them to reflect more deeply on the role of algorithms in their everyday life than they usually did:

> So, did you get a sense of how the app generated these inferences do you think from playing with it? (Interviewer).

> Yeah, I think so, like gathering the kind of frequencies of like word use and that kind of thing, I guess. So, yeah, so if you discount some of your posts and things, it changes it because you're kind of … alongside work ones I do post quite a lot of pictures of my guinea pigs, so, if they're taken out, it changes it. (UK, Participant 6).

> I just realized that I use a lot of things with algorithms and probably in the triple I am not aware […] Maybe if you had asked me on the street – with those quick surveys – I would have told you three things [about algorithms and data], but I have twelve and probably there are more. So that was like "Wow, let's not forget that". (Chile, Participant 1)[6].

[6] Matías Valderrama translated the interviews with Chilean participants included here from Spanish to English.

In some cases, the (in)adequacy of recommendations was understood to be a consequence of the (lack of) capacity of Big Sister and other apps to consider how recommendations may be of value or not depending on awareness of context, including whether an app is able to acknowledge the significance of time and place and the absence or presence of others. This was sometimes experienced by participants as a failure of recognition or understanding:

> For example, [Spotify] recommends me Reggaeton or Trap [music genres] from Chile, but I didn't listen to them or maybe I listened one time, because typically there is a birthday and someone connects your Spotify account and others put music on my account, then someone put Trap on at a party and it is there forever on my record and Spotify thinks I like Trap now. So it's out of context, there's a bias in that too [...] Also, there are apps that can be used by third actors: I go to my mom's house and she loves Inti Illimani, so she takes my phone and puts on Inti Illimani. Then Spotify doesn't understand that she's my mom and doesn't know that I'm not at home for the weekend or maybe it does, but it's not so precise about that yet. (Chile, Participant 1).

> It always came out on the last [personality dimension] bar ... that I was very anxious or very stressed. I said, 'Yeah, it could be because of everything that's going on. I'm probably giving a lot of likes to news about cops hitting people'. So I think, I made that connection, that I'm really giving a lot of likes to that kind of news, as well as the fires in Australia, things a little bit chaotic, there might be a connection with the anxiety of my profile. (Chile, Participant 1).

In other cases still, the lack of relevance was linked to the app's perceived inability to recognize the participant's membership of – and sense of belonging to – an already existing collective:

> [Big Sister] was saying that it was not probable that I like Latin music and I like Latin music. I mean I'm Latin. [Laughs.] I guess apps consider the same variables in all geographical contexts and there's still an important context factor. (Chile, Participant 3).

While these remarks can be seen as criticisms of our and other recommendation apps for being inaccurate, they can also be seen as the articulation of a desire for such apps to be more or differently relevant. That is, they articulate the value for users of the identification of relevance in a way that is more nuanced than a response to pre-existing interactions or expressions of interest, independent of context, and as wanting more than can come from a series of apparently

disconnected recommendations.[7] Indeed, in their criticisms of the app, many of our participants appear to value an understanding of relevance similar to that outlined by Noortje Marres (2012), who argues for an appreciation of relevance not only as a constantly ongoing activity – that is, for relevanc-ing, but also as an activity that requires a sensitive and highly dynamic recognition of context. Indeed, our participants appear to reject the idea that relevance involves a one-off projection of concerns, interests or aspects of identity onto a context or spatio-temporal background (domestic or geographic), independent of their ongoing (and sometimes changing) concerns, interests or collective or social identity. In suggesting that relevance should be more context-aware, or put differently, that the app recognize the ongoing complexity of the situatedness of interaction, the participants thus problematize the terms of their inclusion by the app in a category of 'People Like You' who 'Like Things Like This'.

Inclusion, exclusion and belonging

To explore this issue further consider the distinction drawn by one of our participants between the experience of receiving a recommendation from an app and receiving a recommendation from a friend. The former is described as a solitary experience while the latter is described as 'more about sharing':

> When someone else recommends it to you … I don't know how to explain it … but I feel that … I'm sharing a part of my tastes and I feel that it generates other things afterwards: it generates conversations, it generates "Hey, I didn't like it", "But why didn't you like it?" I feel that it invites you to connect with another person from something that has become common [...] the experience of these digital recommendations is more solitary, like they are for you and almost from yourself, because it's your own data that is generating these things, and the other is more about sharing. (Chile, Participant 4).

Receiving a recommendation from an app is understood negatively by this participant in comparison to receiving a recommendation from a friend because it is seen to be without the possibilities of a continuing connection to others, the lack of something held in common. However, the solitary nature of such recommendations is not

---

[7] Recommendation apps typically make use of behavioural classification and/or contextual classification algorithms. While there has recently been what has been called a 'contextual turn' (Prey 2017), in which apps use data relating to place, time, activity and emotional state our participants' experience suggests that contextual classification is as yet too crude – too behavioural – to be able to incorporate their understanding of context adequately.

aways understood in negative ways as the participant above suggests when they say about the app that 'it reflects me like, there's nothing in it that's presenting me in a way that I wouldn't want to be presented. So, yeah, so, yeah, I guess, I found that quite comforting or supportive, I suppose'.

The feeling of 'for you and almost from yourself' is captured in the title of a book by Kris Cohen: *Never Alone Except For Now*. He explains the phrase as a way to describe the political atomization he believes to be characteristic of participation in digital media: 'felt by some as abandonment or impoverishment and by others, mostly in principle it seems, as freedom to go it alone, to vote and shop as one likes' (2017: 23). A number of the participants in our study appeared to deliberately adopt this solitary way of engaging with recommendation apps, describing their interaction operationally, saying they tried to train apps to better recognize them, or at least recognize some selected aspects of themselves, sometimes expressing their interaction in strategic terms:

> … What is interesting with actually in thinking about the Twitter feed is how I create closure about myself on this type of platform, so in this context I know that the algorithm doesn't know that I like certain things because I never share those things on purpose on this platform. (UK, Participant 8).

The attempt to create 'closure about myself' by taking specific actions was understood by some participants to be recognized by the app as they perceived recommendations to be becoming more precise as they interacted with it:

> There are other exceptions like Rdio [a defunct music platform], which was more precise, but I'm thinking that it's because I probably interacted more, as I'm the one who's constantly educating it. There's a kind of computer-human interaction that tells you, "Is this ok?", "No", "Ah, let's go on the next one, is this ok?", "Yes", "Check" and it's fed by something I'm doing explicitly. (Chile, Participant 1).

In these examples, participants appear to adopt an understanding of interaction in terms of subject–object relations, in which they, as independent subjects, have the ability to direct the app, which is seen as a discrete object, relatively independent of its infrastructure or a wider context, including the interactions of other users. However, other participants brought forward alternative – part-whole – understandings, in which they presented their experience of use of the app as happening in a wider set of arrangements or circumstances. Consider in this regard not only the participant who said 'I am Latin' but also the participant who, while not disputing the high levels of anxiety attributed to them by the app, ascribed it not to their identity

or their personality as an individual subject, but rather to 'everything that is going on', including 'news about cops hitting people' and 'fires in Australia'.[8] This is a strong statement of the participant's understanding of themself (and the app) as being part of a whole, an ensemble of relations that extend beyond the immediate situation of interaction with the app. Nevertheless that this participant did not make any connection between Big Sister and the ways in which the protests happening at the time of the interview were being visualized by protestors through a public iconology of eyes (Fig 3; see also https://en.wikipedia.org/wiki/Eye_injury_in_the_2019%E2%80%932020-Chilean_protests) suggests that the potential multiplicity of part–whole relationships is not easily recognized by users, and that wholes are themselves always partial.

Figure 3: Signs of public protest using eye imagery, Santiago, Chile, November 2019
Source: Photo by author Celia Lury

To explore these relations of inclusion, exclusion and belonging still further, consider the framework of folding introduced by Francis Lee and co-authors in their discussion of algorithms (2019: 2). They suggest that 'it is through multiple operations of folding – of relating things – that [algorithms] work: It is in the many practices of relating, constructing, tinkering and applying that algorithms gain their power to reshape things'. Crucially however, they stress that it is not

[8] In fact, the app's report of excessive neuroticism as a characteristic of the user was likely due to an error in the configuration of the app that was subsequently corrected.

algorithms alone that loop, capture or fold, but that there is a kind of turn-taking, producing parallel lines of (inter)action: 'Sometimes humans fold things into the algorithm, and sometimes algorithms fold things into something else. Hence, agency is not fixed with the algorithms or with the humans'. Certainly our prototyping experiment, in its various iterations, provides support for this analysis. Across the range of responses we have described there is a kind of anamorphic (mis)recognition at play in which interaction or turn-taking involves a series of parallax projections as what is recommended by the algorithm changes in relation to what users do, creating a series of loops or foldings in which space and time are creased and concertinaed, leading to a variety of projections and displacements by users and the app. To understand what is involved here consider the following description of video art, which also creates a synthetic situation that has a relational dynamic. As the art critic Barry Schwabsky writes, 'Video technology and the mirror have this in common: that in reduplicating some fragment of the world, they introduce at least a very small spatial or temporal division into reality'. However, as he points out, while the reflection may be 'at a greater or lesser distance': 'if I try to take what I see in the mirror as a guide for my movements, I will always be in the paradoxical situation of trying to follow something that is following me' (2018: 35).

This paradox is deliberately brought into existence in one of the artist Bruce Nauman's works: Live-Taped Video Corridor which features

> two stacked television monitors at its far end, both linked to a camera mounted at the corridor's entrance: the top monitor plays live feed from the camera, while the bottom monitor plays pretaped footage of the empty passageway from the identical angle. Walking down the corridor, one views oneself from behind in the top monitor, diminishing in size as one gets closer to it. The camera's wide-angle lens heightens one's disorientation by making the rate of one's movement appear somewhat sped up. Meanwhile, the participant is entirely, and uncannily, absent from the lower monitor. The overall result is an unsettling self-conscious experience of doubling and displacement. (https://www.guggenheim.org/artwork/3153).

In our study, nearly all participants reported experiencing the situation in which interaction with recommendation apps takes place to be a 'cramped and disorienting space … a space [requiring] habitation, adaptation and negotiation' (Cohen 2017: 6). They all tried to engage with, manipulate or opt out of the app while the app continued to provide recommendations, continually folding or looping them in or out, co-opting them into its ongoingness whether they followed particular recommendations or not.

More specifically, the parallelism of this relational dynamic appears to create 'a space of prophylaxis between [an algorithmic] logic [of

classification] and more familiar logics of representation, identification, subjectivity and relation' (Cohen 2017: 126) as subject–object and part–whole understandings coexist, complicating the relations between inclusion, exclusion and belonging.[9] As we have noted, some participants disputed the relevance of the recommendations they received: disowning some of the personality characteristics they were ascribed and identifying with others, drawing lines between their selves as individual subjects and others – 'Spotify doesn't understand that she's my mom'. Other participants asserted their membership of a group form and an associated sense of belonging that they believed should automatically lead to inclusion – 'I am Latin'. Indeed, while some participants did not attach much significance to the recommendations at all in terms of identity or a sense of belonging to a category, in this latter case the participant asserts the overriding importance of socio-political grounds for identity that exist outside their individual interaction with the app. They insist upon a sense of belonging to a social group that is not recognized by the app. Others still were prompted by their use of the app to engage in a reflexive consideration of their identity, and even of the relevance of self-possession to the concept of individual identity itself:

> Because there's the whole debate about, okay, algorithms, the way they take information about you, do they take too much, but there's a whole other problem which is more like the problem of what is identity and, and I think what is good with Big Sister app is that it pushes you to think about, okay, how do I stage myself and this just shows that okay, even though there's a sort of fantasy or authenticity and being true to yourself, actually, we spend most of our social life in staging our self in different ways. And also, there's even like a sort of internalization of being true to some rules of social roles. All those things are super naturalized. So, I think something good with recommendation is like, yeah, it makes you think about how you could appear. (UK, Participant 8).

In sum, our participants' articulation of relevance in terms of both subject–object and part–whole relations exposed that some of them perceived an excess of belonging that was not captured by inclusion in the categories or classifications generated by the app. Correspondingly, however, as we have already seen, the app's recommendations create an excess of inclusion that does not correspond to a sense of belonging by the user when they question its relevance to themselves. The excesses informed by turn-taking run in parallel but are not the same.

---

[9] In this respect, recommendation apps appear to share the Möbius form of organization identified by David Stark and Ivana Pais (2020) in relation to platforms. They describe this form as having neither an inside nor an outside but animated by an organizing principle of co-option.

In the next section, we consider the implications of these parallel but alternate excesses (of belonging over inclusion and inclusion over belonging) for the relations between belonging and belongings (Cooper 2007), that is, for possibilities of ownership, possession and property. Our suggestion is that opportunities for valuation – for both the user and the owner of the app – are to be found in the organization of the interdependence of these excesses. To support this claim, in the next section we explore the temporality of this interdependence, the making and taking of recommendations.

The entitlements of turn-taking: Belonging and belongings

In her discussion of the synthetic situation of currency traders Knorr Cetina (2009) emphasizes the importance of time transactions to the interaction order. With this term she draws attention to interaction with on-screen projections in which a future outcome becomes linked to a present commitment. She suggests that such transactions may be coordinated via temporal integration formats. In what follows we suggest that the temporal integration format typical of recommendation apps is a continuous present (Day et al. 2023), and that this format opens up the possibility for such apps to become assets for their owners while foreclosing possibilities for users to assert or recognize relations between belonging and belongings.

It is certainly evident that the temporalities of turn-taking in digital media are viewed by app owners as highly significant for processes of digital valuation, as indicated by the ongoing attempts on a variety of platforms and apps to create what Tania Bucher (2020) calls 'right-time'. The examples she gives include Facebook's stated ambition that the News Feed function 'show everyone the right content at the right time so they don't miss the stories that are important to them' and the replacement of Twitter's one real-time feed with a 'While you were away' section at the beginning of the timeline, an algorithmically generated 'recap of some of the top Tweets you might have missed from accounts you follow' (2020: 1699). In these temporal orderings, relations are, as Alberto Corsín Jiménez puts it, 'always turning themselves "into" other relations, moving in and out of different social forms, in a "flow of analogies"' (2013: 389). Knorr Cetina describes this flow in the case of currency traders as

> like a carpet whose small sections are both being woven and rolled out at the same time in front of us. The carpet grounds experience; we can step on it and change our positioning on it. But the carpet composes itself only as it is rolled out; the spatial illusions it affords hide the intrinsic temporality of the fact that its threads (the lines of text appearing on-screen) are woven into the carpet only as we step on it and unravel again behind our backs (the

> lines are updated and disappear). As the carpet is woven it assumes different patterns; the weave provides specific response slots to which traders react, taking the patterns in different directions. In sum, the screen reality is a process, but it is not simply like a river flowing from one location to another as an identical mass of water. Rather, it is processual in the sense of an infinite succession of nonidentical matter projecting itself forward as a changing situation. (2009: 72).

In the case of recommendation algorithms, the processual flow of recommendations ('People Like You Like Things Like This' and 'Things Like This Like People Like You') may be understood as 'obviational'. This is the term that Roy Wagner (1978) deploys to make 'obvious' the supplementary and substitutive flow of social relations, while acknowledging that some relations – between people and things – are always being 'obviated' or (temporarily) disposed of in favour of others.Wagner suggests that if obviation is successful, 'awareness of time and its passing become one and the same thing' (2019: 12). How this is achieved, he says, depends on the medium of recollection in which obviation occurs. In the case of recommendation apps, we suggest, it is the organization of turn-taking in the digital medium – the ways in which user reaction is folded into prediction - that is key to the making of categories of 'People Like You'. That is, while the parallelism of interaction proceeds in turn-taking – as the participant quoted above describes their use of the app: "Ah, let's go on to the next one, is this ok?" – since the time units are able to be systematically varied by the app in the digital medium to produce recommendations in sequenced but also overlapping ways, the outcome is always that 'Many times are in "People Like You"', and 'Many "People Like You" are in each time now present'. We give this temporal integration order the description 'continuous present', an ongoingness that is continually punctuated for users as recommendations are called up and replaced, as they are simultaneously included in the categories recommended for them and excluded from categories recommended for others.

One participant thought that algorithmic recommendations are more accurate if the user is able to indicate at least some preliminary choices: 'If you are using it for the first time, I think Spotify would let you choose what type of music you like to listen to, like you could choose pop music or you could choose which specific singers. So, I think that's the most direct way' (UK, Participant 5). But our analysis suggests that such choices are never 'first' in any absolute sense since, as the parallelism of turn-taking proceeds, time units are always being varied such that an individual user is always included and excluded in many categories at any one time, even what they perceive to be 'the first time'. This means that, rather than the rule of first possession that characterizes the property relations of the queue (Strathern 2011), the

algorithmic rule of a continuous present is what organizes relevancing, opening up and closing excesses of inclusion over belonging and belonging over inclusion in a variety of continuously ongoing ways. And in doing so this temporal format opens up the possibility (or not) of establishing a potential relation of ownership or property.[10]

On the one hand, since users are never simply inside or outside the categories created by the app at any single point in time, not only are they likely to find it difficult to assert a sense of belonging on a continuous basis, they will find it even harder to identify ways in which they can translate a sense of belonging into belongings (Cooper 2007), or exert any kind of ownership of the category in which they are (temporarily) included.[11] At the same time, they may also feel a sense of belonging which is not recognized by inclusion in the app's categories. On the other hand, and also at the same time (!), the implementation of the continuous present as a temporal integration format, the 'permanently snarled and bewildering temporality' of 'never but always' (Cohen 2017: 5), makes the creation of categories of 'People like you' realizable by the app owner as an asset. Announcing the beta launch of Branded Moments for example, Spotify

---

[10] Consider some of the complexities of ownership in relation to the prototype Big Sister. To start with, it piggy-backs on the terms and conditions of the social media platforms already collecting/extracting/sharing data from their/our participants, allowing us to collect/extract/share data relevant to our concerns. Whether and to what effect such piggy-backing will be enabled by calls for interoperability has become a key regulatory issue (https://www.adalovelaceinstitute.org/blog/walled-gardens-open-meadows/). However our prototypical experiment, conducted in an academic environment, required its own contractual forms between us as researchers and participants (including consent forms), between the app and a commercial platform, and between our institutions, including an extremely complex data-sharing agreement.

Further legal, financial and institutional complexities required (or at least made it seem impossible to argue otherwise) that the UK institution involved be described as the lead institution in a memorandum of agreement, even though, among researchers it was acknowledged that it was the team at the university in Chile who had initiated and were leading the project. Issues of ownership continue in the naming of authors of this and other academic papers relating to the project, an issue that is linked to the ways in which the storage (or banking) of journal articles reorganizes recommendation through the operation of metrics in digital media (Biagioli and Lippman 2020). It is possible, though unlikely, that our prototype app will be commercialized. Possible other sources of revenue of Big Sister include (slight!) increased probability of securing further research grant funding, with other benefits being largely reputational. While we aimed for the co-production of knowledge, our participants are only partially recognized in this account; at the same time, many of them appeared to value their experience of this experiment in ways that are not fully recuperated or recollected here.

[11] Warner asserts that, 'All the verbs for public agency are verbs for private reading, transposed upward to the aggregate of readers. Readers may scrutinize, ask, reject, opine, decide, judge, and so on. Publics can do exactly these things. And nothing else' (2002: 123).

promised to leverage 'our unique data and insights' in order to 'identify – in real-time – what a listener is doing, and give brands an opportunity to own that moment' (Spotify for Brands, 2016, emphasis added, cited in Prey 2017: 9).

Of course, the maintenance of a continuous present is not easy to achieve, and users may choose to drop out altogether:

> There is a sort of cheating thing in the measurements that can be, or maybe exists but I don't know, that finally they read your past, then they are reading the things that you have done. But there is something in particular with music, especially for my profile that I like to get into music a lot, that I discover new things. So, precisely, discovering new things has to do with not repeating the past, like sometimes I've left behind styles [of music] that I don't want to go back to. If this algorithm gets into my past and says, "Oh, look, listen to Backstreet Boys again", I don't want to listen to that, I passed that stage. Then there's also a thing about discovering something new that's antagonistic to the previous pattern. (Chile, Participant 1)

Nevertheless, we suggest that it is the creation of a continuous present that offers the app's owner the opportunity to operate an open-ended and expanded ground for digital valuation, providing as it does the possibility for the app to be a 'resource controlled by [an] entity as a result of past events and from which future economic benefits are expected to flow to the entity' (Birch and Muniesa 2020: 2, 3).

Conclusion

To sum up, use of the method of prototyping in a study of recommendation apps enabled us to see the significance of the temporality of the activity of relevanc-ing for forms of digital valuation. This significance is tied to the finding that in the activity of relevanc-ing conducted by recommendation apps there is an excess of inclusion over belonging for users, and that this excess is more than can be owned either for themselves or as their selves by participants as individual subjects. At the same time, it is difficult for individual participants to recognize or make durable their sense that they are part of, or belong to, a whole that exists outside the app in their interaction with the app. Instead, so our experiment in prototyping suggests, the excess of inclusion over belonging is continuously re-collected in the medium of algorithmic calculation as a continuous present. In this process there is potential for the recommendation app to create a datasset (Beauvisage and Mellett 2020).

By pointing to the importance of both subject–object and part–whole relations, our experiment suggests that while the valuation

opportunities afforded by relevanc-ing make possible 'stealing', that is, data extraction, they may also include a taking part in, belonging to, being part of, maybe even 'sharing' in something else (Strathern 2011). Our analysis suggests however that there are very significant differences in the possibilities of realizing such opportunities associated with different projected positions in the turn-taking parallelism by which recommendation systems proceed. We conclude that in making this visible, there is a role for the experimental use of prototyping to change the terms and conditions of valuation not only by highlighting the ways in which the activity of making relevant establishes multiple relations between inclusion, belonging and belongings, but also by 'designing for belonging rather than individuating' (Tafasee 2021), by encouraging different kinds of context awareness and by drawing attention to the temporality of turn-taking as a kind of class(ification) action.

Acknowledgments

The project 'Algorithmic Identities' was conducted under the auspices of the Fondecyt project N°1180062: 'Datafication of urban environments and individuals: an analysis of the designs, practices and discourses of the production and management of digital data in Chile', and the project '"People Like You": Contemporary Figures of Personalisation', funded by a Collaborative Award in the Medical Humanities and Social Sciences from the Wellcome Trust Foundation 2018–2022 (205456/Z/16/Z). It was also supported by the award of 'Interdisciplinary Research 2018 Funding' from the Vice-Rectory for Research (VRI) of the Pontificia Universidad Católica de Chile.

The authors are very grateful for the comments from anonymous reviewers and the editors of the journal.

This work was supported by the National Fund for Scientific and Technological Development (Fondecyt), Ministry of Education, Government of Chile, Fondecyt Grant N°1210006; the Nucleus Millenium Futures of Artificial Intelligence Research (FAIR) N°NCS2022_065.

References

Akrich, Madeleine. 1992. "The De-Scription of Technical Objects." In *Shaping Technology/Building Society: Studies in Socio-Technical Change*, edited by W. Biker and J. Law, 205–224. Cambridge, MA: The MIT Press.

Amoore, Louise, and Volha Piotukh. 2015. "Life Beyond Big Data: Governing with Little Analytics." *Economy and Society* 44(3): 341–366.

Andersen, Christian, and Søren Bro Pold. 2014. "Controlled Consumption Culture: When Digital Culture Becomes Software Business." In *The*

Imaginary App, edited by P. D. Miller and S. Matviviyenko, 75–95. Cambridge, MA: The MIT Press.

Beavisage, Thomas, and Kevin Mellet. 2020. "Datassets: Assetizing and Marketizing Personal Data." In *Assetization: Turning Things into Assets in Technoscientific Capitalism,* edited by Kean Birch and Fabian Muniesa, 75–95. Cambridge, MA: The MIT Press.

Biagioli, Mario, and Alexandra Lippman. 2020. *Gaming the Metrics: Misconduct and Manipulation in Academic Research*. Cambridge, MA: The MIT Press.

Birch, Kean, and Fabian Muniesa. 2020. "Introduction: Assetization and Technoscientific Capitalism." In *Assetization: Turning Things into Assets in Technoscientific Capitalism*, edited by Kevin Birch and Fabian Muniesa, 1–41. Cambridge, MA: The MIT Press.

Bucher, Tanya. 2020. "The Right-Time Web: Theorizing the Kairologic of Algorithmic Media." *New Media & Society* 22(9): 1699–1714.

Clough, Patricia. 2018. *The User Unconscious: On Affect, Media and Measure*. Minneapolis, MN: University of Minnesota Press.

Cohen, Kris. 2017. N*ever Alone, Except for Now: Art, Networks, Populations*. Durham, NC: Duke University Press.

Cooper, Davina. 2007. "Opening up Ownership: Community Belonging, Belongings, and the Productive Life of Property." *Law & Social Inquiry* 32(3): 625–664.

Corsín Jiménez, Alberto. 2013. "Introduction. The Prototype: More Than Many and Less Than One." *Journal of Cultural Economy* 7(4): 391–398.

Day, Sophie, Celia Lury, and Helen Ward. 2023. "Personalization: A New Political Arithmetic? *Distinktion: Journal of Social Theory* 24(2), 167–194.

Dubois, Elizabeth, and Heather Ford. 2015. "Trace Interviews: An Actor-Centred Approach." *International Journal of Communication* 9: 2067–2091.

Fourcade, Marion, and Kieran Healy. 2017. "Seeing Like a Market." *Socio-Economic Review* 15(1): 9–29.

Goffman, Erving. 1983. "The Interaction Order: American Sociological Association, Presidential Address." *American Sociological Review* 48(1): 1–17.

Knorr Cetina, Karin. 1997. "Sociality with Objects: Social Relations in Post-Social Knowledge Societies." *Theory, Culture and Society* 14(4): 1–30.

Knorr Cetina, Karin. 2009. "The Synthetic Situation: Interactionism for a Global World." *Symbolic Interaction* 32(1): 67–87.

Latzko-Toth, Guillaume, Claudine Bonneau, Melanie Millette. 2017. "Small Data, Thick Data: Thickening Strategies for Trace-Based Social Media Research." In *The SAGE Handbook of Social Media Research Methods*, edited by Anabel Quan-Haase and Luke Sloan, 199-214. London and New York: Sage.

Lee, Francis, Jess Bier, Jeffrey, Christensen, Lukas Engelmann, Claes-Fredrik Helgesson and Robin Williams. 2019. "Algorithms as Folding: Reframing

the Analytical Focus". B*ig Data & Society*, 6(2). https://doi.org/10.1177/2053951719863819

Lury, Celia, Sophie Day. 2019. "Algorithmic Personalization as a Mode of Individuation." *Theory, Culture and Society* 36(2): 17-37.

Mattelard, Armand, Andre Vitalis. 2015. *De Orwell al Cibicontrol*. Barcelona: Gedisa.

Marres, Noortje. 2012. *Material Participation: Technology, the Environment and Everyday Publics*. London: Palgrave Macmillan.

Nieborg, David B., and Thomas Poell. 2018. "The Platformatization of Cultural Production." New Media & Society 20(11): 4275–4292.

O'Reilly, Tim. 2005. "What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software." *O'Reilly* 30 September. https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html?page=1, accessed 21 August 2023.

Prey, Robert. 2017. Nothing personal: algorithmic individuation on music streaming platforms, *Media, Culture and Society*, 40(7), 1086–1100.

Schwabsky, Barry. 2018. "Seeing through a Glass." *The Nation*, 14 May: 33–35.

Stark, David, and Ivana Pais. 2020. "Algorithmic Management in the Platform Economy." *Sociologica* 14(3): 47–72.

Strathern, Marilyn. 2011. "Sharing, Stealing and Borrowing Simultaneously." In *Ownership and Appropriation*, edited by Veronica Strang and Mark Busse, 23–42. Oxford: Berg.

Suchman, Lucy, Randall Trigg, and Jeanette Blomberg. 2002. "Working Artefacts: Ethnomethods of the Prototype." *British Journal of Sociology* 53(2): 163–179.

Tafesse, Emnet. 2021. "Between Belonging and Individuating: Data, Identification, Diaspora." *Data & Society: Points* 8 December. https://medium.com/datasociety-points/between-belonging-and-individuating-data-identification-diaspora-9fcdf010e0ce, accessed 12 March 2024.

Tironi, Martín. 2020. "Prototyping Public Friction: Exploring the Political Effects of Design Testing in Urban Space." *British Journal of Sociology* 71(3): 503–519.

Tironi, Martín and Matías Valderrama. 2021. "Descolonizando los sistemas algorítmicos: diseño crítico para la problematización de algoritmos y datos digitales desde el Sur" [Decolonizing Algorithmic Systems: A Critical Design to Problematize Algorithms and Digital Data from the South"] *Palabra Clave* 24(3):e2432.

Wagner, Roy. 1978. *Lethal Speech: Daribi Myth as Symbolic Obviation*. Ithaca, NY: Cornell University Press.

Wagner, Roy. 2019. *The Logic of Invention*. Chicago, IL: Hau Books.

Warner, Michael. 2002. *Publics and Counterpublics*. New York: Zone Books.

Woods, Heather S. 2018. "Asking More of Siri and Alexa: Feminine Persona in Service of Surveillance Capitalism." *Critical Studies in Media Communication* 35(4): 334–349.

Woolgar, Steve. 1990. "Configuring the User: The Case of Usability Trials." *The Sociological Review* 38(1): 58–99.

Celia Lury (corresponding author) is a Professor in the Centre for Interdisciplinary Methodologies at the University of Warwick. Recent publications include co-editing with Sophie Day and Helen Ward a Special Issue of *Distinktion: Journal of Social Theory*, on 'People Like You: A New Political Arithmetic'.

Sophie Day is Professor Emerita of Anthropology at Goldsmiths, University of London. Her latest publication is *Rendering Houses in Ladakh* and she recently conducted research with co-investigators Celia Lury and Helen Ward on "'People Like You': Contemporary Figures of Personalisation" in a collaborative medical humanities project supported by the Wellcome Trust.

Andre Simon studied Computer Science at Pontificia Universidad Católica de Chile, and collaborated in the Algorithmic Identities project as a software engineer, co-creating the app Big Sister. He is interested in the way people interact with technologies across history; his current focus involves improving the user experience at Microsoft.

Martín Tironi is Director and Associate Professor at the School of Design, Pontificia Universidad Católica de Chile, and director of the Millennium Nucleus Futures of Artificial Intelligence Research (FAIR). His research areas are design anthropology, digital technologies and mobility, and urban infrastructures.

Matías Valderrama is a PhD researcher in the Centre for Interdisciplinary Methodologies at the University of Warwick. He is exploring how different actors establish, demonstrate, contest or undermine the connection between social media platforms and societal harms. He holds an MA and BA in Sociology from the Pontificia Universidad Católica de Chile. He has worked on many Fondecyt research projects and studies for NGOs on the social implications of multiple digital technologies in Chile, such as environmental sensors, drones, predictive models and self-tracking devices.

Scott Wark is a Lecturer in Media Studies at the University of Kent. He researches digital culture and the relationship between race and ethnicity and digital technology. Prior to joining Kent, he held a position as a Postdoctoral Research Associate on the Wellcome-Funded 'People Like You: Contemporary Figures of Personalisation' project. This position was based at the University of Warwick's Centre for Interdisciplinary Methodologies, where he completed a PhD in Interdisciplinary Studies.

Valuation Studies 11(1) 2024: 60–90

Theme issue contribution

Valuing data: Attaching online data to stakes, selves, and other data

Susann Wagenknecht, Laura Kocksch, Stefan Laser, and Ann-Kristin Kühnen

Abstract
As datafication proceeds rapidly, a large, unwieldy amount of data is available online. In this article, we ask: How valuable is this data, how is it made valuable? To answer this question, we study how online data is endowed with worth in virtual collaboration workshops. Our workshops challenged participants to assert and question the worth of online data – a challenge that participants addressed by using a set of techniques of which we describe collage, hierarchy building, and calculation. Data, we show, gains value through attachment. Thinking with attachment, we foreground affect, materiality, and the situatedness of valuing online data. As ethnographers, we study how data, as haphazard as it comes, is attached to the circumstances and stakes at hand, to ourselves and to other data. Our study contributes a conceptual perspective that attends to the shifting boundaries of the personal and the public, tensions between locality and generality, the role of contiguity, and the limits of combinatorial connectivity.

Keywords: digitised valuation; ethnography; data; attachment; (post-)actor–network theory; virtual collaboration

Susann Wagenknecht is Junior Professor at the Institute of Sociology, TUD Dresden University of Technology.
Laura Kocksch is Post-Doctoral Researcher in the Technoanthropology Lab, Aalborg University.
Stefan Laser is Post-Doctoral Researcher at Ruhr-University Bochum.
Ann-Kristin Kühnen is Doctoral Candidate and Research Assistant at the Institute of Sociology, TUD Dresden University of Technology.

https://doi.org/10.3384/VS.2001-5992.2024.11.1.60-90
Hosted by Linköping University Electronic press
http://valuationstudies.liu.se

Introduction

Datafication proceeds rapidly, and untameable amounts of data are available online. But how valuable is this data anyhow? Or better, how can it be *made* valuable? In this article, we unpack the purview of these questions for valuation studies. We study two virtual workshops to examine how online data attains worth in digitised valuation. To analyse the valuation of online data, we adopt the notion of attachment. Thinking with attachment (de Laet et al. 2021), we contribute a conceptual perspective that attends to the shifting boundaries of the personal and the public, tensions between locality and generality, the role of contiguity, and the limits of combinatorial connectivity (cf. Conward 2018).

With attachment, we approach digitised valuation with a focus on affect, materiality, and situatedness. As an analytic term, attachment speaks to value because valuation is, crucially, concerned with relatability. Following Antoine Hennion, attachment is "connection, restriction, restraint and dependence" – relations that are continuously reworked (Hennion 2017a: 113, 118). Attachment keeps things together. Attachment is affective; it requires dedication. At the same time, it relies upon material qualities and sensorial capacities. Some things stick with us, and others glide through our hands. Most importantly, however, attachment needs to be understood through situatedness – it is bound to circumstance, always "specific to a locale" and its outfit (de Laet et al. 2021: 801). Attachment emerges from situated engagement, close and local, and sustains it at the same time. In this sense, attachment denotes relations of a specific kind and emphasises the inevitably situated character, the material qualities, and the affective investment of valuing data.

Situatedness, however, is commonly ignored when things virtual are discussed (Strathern 2002). Likewise, data is often portrayed as disembedded, mobile, abstract, immaterial, and unaffected. Re-emphasising the situatedness of data, scholars in science and technology studies (STS) remind us that "[a]ll data are local" (Loukissas 2019), "cooked" (Gitelman 2013), and "partial" (Pink et al. 2018). Stitched together in a plethora of practices, data carries traces that engender, hinder, obscure, or ease its situated valuation. With attachment, we shed light on the affective, material, sensorial, and reflexive capacities needed to endow data with worth.

In examining how online data gains value in digitised practice, we refer to valuation studies and post-actor–network theory (ANT), as well as literature from data studies. Empirically, we draw upon collaborative, participant observation in two workshops, both hosted with the help of home-to-home videoconferencing. The workshops build on experimental and participatory approaches developed in digital methods and use water data as a case in point. In our analysis, we detail how workshop participants dealt with the challenge of

making online data, as haphazard as it comes, relatable. Finally, the article discusses how online data finds attachment and gains worth, thereby characterising some techniques that workshop participants devised in virtual collaboration.

Data in digitised valuation

Scholars in the field of STS have characterised data as "local", "cooked", "rotten", "broken", haphazard, and patchy (Gitelman 2013; Doganova et al. 2018; Pink et al. 2018; Loukissas 2019). Data is always already processed and it relies upon the infrastructures necessary to collect, clean, and maintain it (Latour 1999; Bowker 2008; Edwards 2010; Ribes and Jackson 2013; Dumit 2018). Data, then, is inherently situated and relational.

The manifold relations that data collection and use establish make data what it is. Relations are at the heart of data. When, however, data value is narrowly defined as measurement, the reference relation between measuring (device, researcher, institution) and measured (object) gains pre-eminence. This eclipses the complexly layered relations work that make data *valuable*. In this article, we therefore foreground the relations that data entertains to things and people, to circumstances and situated practice, and, crucially, to other data. Literally, "data" comes in *pluralis* and becomes meaningful only in relation to one another (Mämecke et al. 2018). Data formats and databases are designed for circulation, compatibility (Helmond 2015), combination and "commensuration" – i.e. "the transformation of different qualities into a common metric" (Espeland and Stevens 1998: 137). Data helps to reframe and aggregate, also transcend, "personal trouble" in the public negotiation of controversial issues, generalised claims, shared concerns, and the common good (Madsen 2023; cf. Mills 1959). In this capacity, data is used both to make powerful claims about reality and to scrutinise its shortcomings. It is appreciated for being "evocative" (Mützel et al. 2018), stimulating cooperation, and fostering joint intervention (Star and Griesemer 1989; Jensen et al. 2021). Yet data is "mercurial", Rachel Douglas-Jones et al. (2021) argue. As it shifts formats quickly, travels fast, and proliferates, it glides through our hands all too easily. Its mercurial character challenges us to re-examine the situated relations work that sustains data and its value.

To value means to assemble and to cast aside. Here, we draw on scholarship in valuation studies that emphasises how "[c]reating value is a process of joining together: classifying, grouping, combining, making, re-forming. Yet", as Emma Greeson, Stefan Laser and Olli Pyyhtinen (2020: 157) argue, "it is also a process where persons, things, parts of bodies, or landscapes are disentangled, abandoned, dismissed, or corrupted." Thus understood, valuing is relations work. It

takes forging relations and cutting some, too (Strathern 1996). It can mean "sorting things out", a never-ending process that, if performed skilfully, requires delicate attunement to circumstance (Bowker and Star 2000: 47). With this perspective, we conceive of valuing as a bundle of practices, ranging from assessing and appraising, appreciating and depreciating, adopting as well as discarding, to amending, supplementing, adapting, and, crucially, fitting – equipping things and people with qualities, and "trueing" them up to what is at hand and at stake (Thévenot 2002). As Frank Heuts and Annemarie Mol (2013) point out, valuing is ill-understood as a judgement after the fact. Studying how tomatoes are valued in diverse practices, Heuts and Mol draw attention to how people *make* good tomatoes, i.e. how they engage with them and manipulate them to bring out, preserve, and increase their worth. The qualities that make tomatoes valuable, Heuts and Mol argue, "are not given; they may be tinkered with" (2013: 130). Valuing, in this sense, is not about fixed qualities but about perpetual qualification. In this vein, we focus on data to examine how valuing qualifies and disqualifies pieces of data as well as those who handle them, their capacities, and concerns.

There are different ways to make things (data, tomatoes) good and ascertain, or challenge, their worth. Things can be tried, measured, and tested (Potthast 2017); they can also be appreciated, carefully probed, and tasted (Hennion 2007). Multiple, incongruent, and incommensurable valuations may coexist, conflict, or fold into one another (Helgesson 2016). Scholars of valuation studies have observed how different "registers" of valuation interact (Heuts and Mol 2013), how distinct "styles" of valuation interweave (Lee and Helgesson 2020), and specific "constellations" of valuation play out (Waibel et al. 2021: 35). Moreover, scholars have parsed different "regimes" of engagement and valuation, distinguishing "intimate" from "public" engagement and elaborating their distinct notions of worth – from the personal, appropriate, habitual, and convenient in intimate engagement to legitimacy, justifiability, and common good in public engagement (Thévenot 2002). As we will show, the relationship between the personal and the public proved particularly salient in our study.

In this article, we focus on situated performances that unfold in specific "moments of valuation" (Antal et al. 2015). The moments we observe are "synthetic situations", i.e. characterised by the multi-sited and far-reaching interactions that sophisticated digital technologies afford (Knorr Cetina 2009). Note that the synthetic moments of digitised valuation involve people and things both analogue and digital, in immediate or mediatised presence. Note, too, that moments of valuation are not insular. They are linked by shared layers of knowledge infrastructure and build upon one another in re-articulating the socio-material "agencements" (Kjellberg et al. 2013) that they both shape and are shaped by. Data, we argue below, gains worth when it

fits into agencements of valuation and ties in with the stakes of the moment, i.e. with what the situation at hand "is about" and what it "puts to the test" (Goffman 1974; Marres and Stark 2020). As scholars have shown, digitised valuation can absorb, reprocess, and sometimes subvert the measures of worth inscribed in databases, search engines and platforms, as well as, e.g. the ranks and feeds of social media (Kropf and Laser 2019; Lee and Helgesson 2020; Paßmann and Schubert 2021; see also Balsiger and Jammet 2022 in this theme issue). When data becomes the subject of digitised valuation, both the "infrastructured-ness" of data and its valuation can inspire reflection and doubt. As we will show, workshop participants grapple with the technologically-augmented reflexivity of digital infrastructures that allow them to "see" – people, things, and data "out there" – and manipulate what they are shown at the same time (Knorr Cetina 2009: 64).

To account for the valuation of data, we rely upon the notion of attachment, a term rooted in post-ANT (Gad and Jensen 2010). As it inquires into the in/capacity to establish relations, attachment helps account for the texture of associations beyond the mere "interessement" that earlier ANT centred on (Callon 1984). As an analytic heuristic, attachment draws attention to the adhesive qualities of things and people, i.e. their ability to hold together, stick with, involve – or, slip away, repel, and ward off various forms of association. "Thinking with attachment", as Marianne de Laet et al. (2021) put it, means attending to the situated, more-than-human dynamics that brings things and people together, or separates them. Avoiding a priori fixation on human subjects and their preferences, attachments "do not belong to people nor define them" (Hennion 2017a: 118). Attachments have to be "continuously done and re-done" (118), efforts that are not only situation-dependent but reflectively work with their situatedness. Attachments are "specific to a locale and its material devices" (de Laet et al. 2021: 801). They are done by subjects "that have the agency to act as well as surrender, and rely on collaborations of sorts with objects that give (feed)back as they are tried, tested, tasted, put into use, crafted or falling in disrepair" (de Laet et al. 2021: 801). Put differently, attachments bring "subject-networks" into being (Gomart and Hennion 1999), co-formative relationships that endow subjects with competencies (of discrimination and connoisseurship) and objects with qualities (of worth). Attachments equip subjects and objects with capacities for valuation. In this vein, Antoine Hennion (2007, 2017a, 2007b) approaches attachment through "amateurship" (from Italian amatore, to love). Amateurship has amateurs re-emerge and reflect themselves through the select artefacts they are passionate about. As such, attachments are not easily transferrable or scalable. They are difficult to judge from afar. Fragile, fluid, and ephemeral, they challenge us to simultaneously observe detachment, disattachment, and reattachment.

In this article, we propose attachment as a conceptual perspective on the valuation of data. When we conceive of valuation in terms of attachment, we follow Hennion's argument that valuation is not to be understood "as a measure of inert things made from outside […] but as multiple "additive" relations, experimentations that help sustain those very things" that are being valued (Hennion 2017b: 79). Attachment cherishes objects as valuable while providing subjects with a taste for their worth. Working with the notion that data can be made valuable through the "affective and attentive relationships" built with and among them (Pinel et al. 2020: 175), we ask: How does data gain value through attachment, and how does attachment play out in digitised valuation? And, importantly: What are the limits of attachment in the digitised valuation of data?

Participant observation in videoconference workshops

To study how data is valued, we have relied upon the participant observation of two experimental workshops. Throughout the workshops, we have shifted roles between facilitators, local organisers, participants, and ethnographers as both events have been devised, planned, conducted, documented, and analysed by the authors of this article. Leaning onto collaborative fieldwork (Lassiter 2005; Estalella and Sánchez 2018), the workshops have drawn inspiration from short-term, participatory, and open-ended formats such as data sprints (Munk et al. 2019; Jensen et al. 2021). We conceive both workshops as *experiments* – encounters under unusual temporal, technological, and organisational constraints (Lezaun et al. 2013). Each workshop lasted an hour and a half; participant interaction relied heavily upon the functions of a common videoconferencing tool. Since we designed the workshops as occasions to probe, tinker, and create, participants were given instructions that remained intentionally vague and afforded much leeway. They were prompted to sieve through data repositories and online sources. Yet participants were free to interpret workshop activities as they liked and chose, e.g. which websites to visit and what data to pick, which search engines, data repositories, or online tools to use, and how to work with the data they found. The workshops, neither problem-centred nor application-oriented, were not geared towards utility or any definite "solution".

The workshops took place during the Covid-19 pandemic when partial lockdown and travel restrictions confined us to domestic retreat and pushed us firmly into the armchair (Howlett 2020: 12). The armchair has been criticised, and rightly so, for privileged complacency. It has been ridiculed for being out of touch and credited with (the illusion of) objective detachment. In our ethnography of

home-to-home videoconferencing, however, we seize upon the domestic to grapple with the locality and intimate investment of attachment and work through the situatedness of digitised valuation. To address the material, affective, sensory, and reflexive dimensions of relations forged through and with online data, we have borrowed from virtual ethnography, leaned onto digital methods, and resorted to participant observation online (Boellstorff 2008; Rogers 2013; Hine 2015). In videoconferences, ethnographic vision is sustained by layered screens, cameras, microphones, and speakers – auxiliary devices that are curiously "present-absent" (Ziewitz 2011), simultaneously separate and connect (Winthereik et al. 2011), seam spaces together and keep them apart (Vertesi 2014). As participant observation in online videoconferences affords distinctly "partial ways of seeing" (Rossmann 2021), this article draws on a compilation of ethnographic notes and materials collected during workshops. Participants of the first workshop were asked to write short, ex-post reflections. In the second workshop, ethnographers accompanied all breakout sessions, and two out of four sessions were recorded. In addition, we implemented a file-sharing system that relied upon the university cloud service and asked participants to upload the data they found and all sketches, notes, or photos they took. For our analysis, we have combined the composition of ethnographic vignettes and memos with coding techniques and category-building in an iterative, comparative approach. All participants, except for co-authors, were given pseudonyms.

Both workshops started with an individual exercise in online data scraping (Marres and Weltevrede 2013) that focused on publicly available online water data. Water is both a private need and a public issue (Barnes and Alatout 2012; Krause and Strang 2016). It is ubiquitous and scarce, widely measured and monitored yet rarely known, and hence particularly suitable, we hypothesised, to probe the worth of data across different registers of valuation. In searching for water data online, participants were encouraged to pursue their own understanding of what data, in fact, were. Participants could choose the materials they wanted to rely on, and problematise, *as data*. Since web access quickly generates an overwhelming abundance of information, they felt the need to be selective. They picked materials ranging from ready-made Excel files to statistical information from various online media or newspapers (see Table 1). Data was found on different kinds of websites hosted by various organisations. Cut from utterly different sources, the collected data formed part of diverse texts (ranging, e.g., from policy reports to activist communication) and couched in various arguments about the waste, conservation, and treatment of water. Because web searches interpret phrases such as "water data" or "data about water consumption" pre-eminently in quantitative terms, they may have introduced a bias towards numerical data.

Type of Data	Example material
Activist Communication	https://www.watercalculator.org/footprint/data-centers-water-use/
Advertisement	https://www.waterlogic.de/blog/29-beangstigende-fakten-uber-globale-wasserverschmutzung/
Calculation Tools (private databases)	Verivox Water Price Tool
Easily accessible science communication	Teaching material for physics classes: https://physikunterricht-online.de/jahrgang-11/wasserwellen/
Fiction Literature	G.G. Márquez: Relato de un Náufrago
Laboratory Testing Data	Laboratory results of commercial water quality test
Medical Advice	Letter from insurance informing participant to stay hydrated
Newspaper Articles	Newspaper Article on anticipated water scarcity in Saxony: https://www.saechsische.de/plus/reicht-das-wasser-kuenftig-in-duerrejahren-5102942.html
Photographs and Stock Photos from news media	https://www.dw.com/de/die-virtuelle-wasserverschwendung/a-37235591, 2017
Reports by Municipalities and Local Governments	https://www.dresden.de/media/pdf/umwelt/UB_Grundwasser.pdf
Product Information	Label on coffee packaging
Scientific Articles	Impact of coronavirus on water infrastructure in Brazil: https://www.sciencedirect.com/science/article/abs/pii/S0921344920304158
Statistics	Investment in water and sanitation with private participation (current US$): https://data.worldbank.org/indicator/IE.PPI.WATR.CD?view=chart
Videos	https://www.youtube.com/watch?v=408PZ_zrs5Y
Visualisations	https://ourworldindata.org/grapher/global-freshwater-use-over-the-long-run
Weather Databases	https://www.wetterkontor.de/de/wetter/deutschland/monatswerte-station.asp?id=10384
Misc.	https://de.wikipedia.org/wiki/Virtuelles_Wasser

Table 1: Overview of data collected in workshops provided by authors (duplicates eliminated)
Source: Authors' own work

With scraped data, workshop participants joined virtual breakout rooms. They were asked to work in groups to reconcile pieces of data with one another. (The second workshop also prompted participants to identify conflicts in their data.) The declared goal was to articulate relations between the different data found. Participants were invited to

jointly devise visualisations, such as models, schemes, or sketches. Here, the workshops varied because we assembled groups of varying sizes within different institutional ramifications. For example, the first workshop allowed all participants to present their data one by one in plenary before we set up two breakout sessions. The second workshop, in contrast, allotted less time for plenary presentations and confined much of the group discussion to breakout sessions (see Table 2).

The first workshop, June 2020	**The second workshop, January 2021**
(90 minutes) Agenda: Short introduction Activity 1. "Find one splash of water data" (data scraping, individually, 10 minutes). Activity 2. Presentation of results. Activity 3. "Use creative means to craft a model for (and provide reflection upon) relating, and e/value/ating, utterly distinct splashes of data" (in breakout sessions, 15 minutes). Activity 4. Presentation of results, open discussion.	(90 minutes) Agenda: Short introduction Activity 1. "Find data that documents where and how water is wasted" (data scraping, individually, 10 minutes). Activity 2. "Discuss: Where is there conflict in your data, what data is ir-/reconcilable?" (in breakout sessions, 15 minutes). Activity 3. "How are you able to relate your data in discussing the worth(s) of water?" In addition, participants were asked to develop a visualisation of their discussion. (in breakout sessions, 15 minutes) Activity 4. Presentation of results.

Table 2: Workshop activities
Source: Authors' own work

Finding data and putting it together

People and things, participants and data encountered one another differently in the two workshops: the first workshop, in June 2020, formed part of a bi-weekly research colloquium and was attended by a group of five social science researchers (including the first, second, and third author) – doctoral students, postdocs, and professors, all of whom share an interest in STS. Five were associated with the research group that hosted the workshop. The first author was acquainted with two workshop participants and organised the event. The second workshop took place in January 2021 and was part of a weekly Master's course in sociology at the university, where two co-authors are employed. Twelve students and one professor attended this workshop while all co-authors of this article acted as facilitators–observers.

The workshops differed distinctly regarding participants' expectations, professional experience, and academic status – differences that shaped their willingness to engage in what they

perceived to be going on. Cheery and curious, participants of the first workshop regarded the event as an opportunity to hone their academic skills. Though still somewhat unusual in June 2020, the workshop was held online. All participants had their cameras switched on and eagerly used various digital tools (e.g. Etherpads or online unit converters). In contrast, participants of the second workshop, also held online, were in attendance due to the requirements of their study programme. They all had earned a BA in the social sciences, but only a few had experience working with data gathered online. Note that, at the beginning of 2021 and well into Germany's second lockdown, students were not yet as well-versed in virtual learning as they would become. Some participants kept their video cameras switched off throughout the workshop. With many of the participants aloof and reluctant, and some outrightly irritated, workshop facilitators found it challenging to engage participants in group work. Participants seemed ill at ease and unsure what to expect from a workshop embedded in a course that usually consisted of short lectures or presentations by faculty members.

What's (good) data?

When workshop participants browsed web pages and sifted through data found online, they had to decide which data to pick and present. Which data would resonate with fellow participants and facilitators? Presenting data means deeming data, as inconclusive or incomplete as it comes, valuable – a challenge that some participants enjoyed, others shunned. During the second workshop in particular many participants were unsure about the data they found: "I'm not so good at finding real data", a participant claimed. "But this is not good data", another one reasoned. Annoyed, for example, one participant told his breakout group that he "did not find anything" online. When the second author reacted baffled, the participant explained that data should either confirm his assumptions or yield new insights. Yet none of the data he found so far, he elaborated, lent itself to either one or the other – he found it all equally inconclusive. Other participants disqualified data as "too fine-grained" or "not trustworthy". Often, participants found data quality hard to gauge. Doubtfully, they commented and compared the order of results that search engines would display for different search terms and users, questioning the reliability of online data searches. If search engines were highly personalised, websites were not disinterested, and data quality was challenging to assess; would it be appropriate to pitch the data they found to teammates and workshop organisers?

Frustrated, participants in the second workshop were testing the scope of data organisers were willing to accept: What qualified as (good) data? Trying to reanimate an increasingly sluggish discussion in

one of the breakout groups, one participant, a faculty member, made a point of getting up from his chair and consulting the floor-to-ceiling bookshelves behind his desk. He retrieved a foreign-language book title and held it to his camera for others to decipher: *Relato de un Náufrago*, Gabriel García Márquez's *Story of a Shipwrecked Sailor*. Indeed, Márquez's *Story*, the factual account of a sailor who knows the ins and outs of water, testifies to the multiple worths of water, ranging from the fearsome force of the ocean to the indispensability of drinking water. As its blurred cover appeared on participants' screens, the book raised intricate questions. Could this be data? Why did we, sociologists and anthropologists, confine ourselves to quantitative data? What could be the status of analogue text in this online meeting?

Questions about data qualification found articulation also in a personal story one participant brought up at the beginning of the second workshop. The participant detailed how her home, situated in a rural area, was acutely affected by declining groundwater levels. Her family's well had run dry, making it harder to keep animals. While she offered newspaper coverage by way of evidence, she conceded that she had no "scientific" data at hand – a shortcoming that fellow participants used to sideline her story quickly. Without "reliable" data on local groundwater, they argued, they should focus on corporate water consumption and the responsibility of large, international corporations instead. Later, when tasked to put the data they found in relation to one another, the group drew up a mind map that did not mention the family well at all, even though the third author intervened to suggest its inclusion. Participants pointed out that their mind map featured "social values" and argued that the well was implicitly included in this category. The personal account about a dried-up well, a sad family story about loss and irredeemable change, did not tie in with what fellow participants deemed data valuable to the task at hand.

Many participants felt drawn to figures and charts that offered absolute quantities. Large quantities seemed especially appealing but unfathomable at the same time: 10.6 billion cubic litres were used for the production of textiles and cotton imported by Germany in 2010. "Nice figure [schöne Zahl] but …", a participant that we will call Uwe remarked, shrugging, laughing, and discarding this piece of data. Particularly in the second workshop, most participants were quickly willing to let go of data they found. Hannah, however, was an exception. Right in the introductory round, she emphasised that she understood the workshop's question about water consumption and water waste as a question of global scope. Enthusiastically, she recommended ourworldindata.com, a website run by a British charity in cooperation with the University of Oxford. The website curates international data sets and presents them in visually appealing charts and maps, many of which are interactive. It offers various information graphics on how, e.g., water consumption levels in industrialised

countries compare to levels in threshold countries. Fittingly Hannah's virtual background sported a world map strikingly similar to those used by the website's info-graphics. Hannah presented herself as someone who does not stay at home but ventures into a global, virtual world. Curiously, in the group work phase, her teammates would be the ones to settle on a handwritten mind map that did not mention any data at all – a stark contrast to the digital, neat, and colourful info-graphics Hannah had brought up.

Venturing into the plethora of data available online is not a banal task, and participants had to decide what data they would attach to the situation, i.e. which data to include in their situated performance and fit into the self-image they wanted to endorse. Many student participants approached this question on strict methodological grounds, valuing data as "un/real", (not) "trustworthy", or (not) "fine-grained" enough. Other participants more readily attached data to personal concerns and global politics, carefully canvassing the issues that could be considered within the realm of the workshop. Some participants carefully tested the facilitators' permissiveness by invoking what challenged conventional, quantitative understandings of (online) data – a tactic probing the workshop's stakes and subtly shifting them.

Data from home, data at home

Seated at home, workshop participants found different ways to seize upon the personal and the domestic amid their private lives. Yet while some participants enthusiastically brought up data related to individual research interests, their places of living, or their personal history, others shunned doing so. And while some found it easy to interest fellow workshop participants in data that bore an intimate relation, others found it utterly challenging to convey private matters that resonated with their peers. The above-detailed example of the dried-up family well, a story quickly pushed aside, is a case in point. Some participants strongly resisted making personal experience and private life the subject of their research. The following episode from the second workshop shows how participants acknowledged the role of domestic water consumption while keeping workshop discussions and private lives separate affairs.

When a group of participants reviewed the data they had collected, one of them, Uwe, used their breakout session to think aloud:

> I don't know, but my first association was when asked to set data in relation that we could bring everything together in photos. Everything is related to water, the morning routine, for example. I get up, I shower, the products I use and wear are produced with water, I have breakfast with fruit that is also cultivated with water, and this is also somewhat data. (Uwe, workshop participant).

Having his morning routine pass before his inner eye, he realised: "All the data [we have found] ultimately relate to us; industrial production and so on – it remains with us." Uwe and his teammates quickly agreed to sample photos in a scrapbook fashion to illustrate how domestic practice involves water consumption. One of them wondered whether they should take pictures of their apartments, but by then, they had already begun searching for stock photos online – eyes flying over the screens, fingers busy on keyboards and mice.

As Uwe and his teammates were arranging photos, the first author wanted to know why they had left out all the quantitative data they had researched earlier. It would be "hard to estimate for private households", Uwe replied. "We could try to assign an estimated amount of water … but I would not know how much water I need to wash the dishes – this would be mere speculation," he said, shaking his head. He then elaborated that, from his point of view, environmental policies should aim at curbing agricultural and industrial water use. Sure, he explained, one could reduce private water consumption, but requiring people to time their showering would be difficult. Uwe felt that private water use cannot and should not be measured and evaluated. But while Uwe defied the notion that private water use counts, others in his breakout session disagreed. Avoiding outright confrontation, one participant, Taina, kept adding photos to the canvas – attaching more and more concerns to their group work. As a result, a compromise emerged in the form of a scrapbook collage that assembled impersonal stock images to illustrate the volume of private water consumption without mentioning any quantitative information (see Figure 1).

Figure 1: Scrapbook collage
Source: Participants' own work

In contrast, other participants welcomed the workshop as an occasion to engage with online data playfully. With an interest in data studies, participants of the first workshop picked large quantity data that related to personal experience and life-world circumstances. One participant researched water data relating to arid regions in her country of origin, and two participants brought up water data from cities they inhabit. Flora, e.g., explained how a particularly dry month could become a "problem for my garden". Even though the beginning of the year had been rainy and the rain barrel in her garden was well-filled, Flora had noticed that the previous month, May 2020, had been particularly dry. She recalled a note in her local newspaper that confirmed her observation. She worried that, on average, precipitation had declined, and her garden might dry out in summer. During the workshop, Flora retrieved data from an online weather service (wetterkontor.de), showing that the previous month of May had yielded only 59% of the long-term median of 1981 to 2010.

Similarly, the speciality coffee sitting on his kitchen counter struck another participant of the first workshop (third author). His coffee, "Brother Baba", a mellow roast with distinct flavours of caramel, waffle, and honey, came with a little ID card describing its origin, roast

profile, intensity, and sweetness. The packaging also featured an entry stating, "PROCESSED: WASHED". A quick search on the web listed various methods of processing coffee, washing being a common one for espressos. Wikipedia offers average amounts of water needed for coffee washing procedures in different regions. This particular coffee was from India ("Baba Budan Giri"); its processing used 14–17m³ water per tonne of unpicked fruit.

To make large quantity data relatable, the coffee atop the kitchen counter helped frame water consumption as a global issue, linking resource strains in the global South to consumer tastes in industrialised countries. Digital data, here, was appreciated through its attachment to the offline pleasures of domestic life. The group discussion kept returning to the coffee everyone would like to try: Can you smell the caramel flavour, can you feel the washed beans running through your fingers? Yet while participants were intrigued by the quality of coffee, and the quantity of water necessary to produce it, the quality of data presented no issue at all. This illustrates how in this instance, participants did not hold found data against academic standards but instead appreciated them through their relationship with the sensorial, enjoyable qualities of everyday life and its private concerns.

Putting data together

Throughout the workshops' second half, groups of participants were asked to jointly establish relations between the data they had found individually; a task that workshop organisers gave little explanation and no advice on. As detailed above, one group used an online pad to create a scrapbook-like collage of photos found online. The images illustrated water consumption in everyday life, starting with morning activities at the bottom and finishing with evening activities at the top of the page (see Figure 1). Another group settled on a handwritten mind map (see Figure 2), believing that it would be "practical" if one of them, Thomas, drew the mind map by hand and later uploaded a photograph of it. Thomas placed the term "water", surrounded by the shape of a water drop, in the middle of a sheet of squared paper and began to cluster categories around it: "scientific", "social", and "economic". Gazing down at his writing, Thomas provided a running commentary for fellow group members. From time to time, he held his notepad to the camera. Unfortunately, his internet connection was poor, so the mind map was difficult to read and his speech was frequently interrupted. Still, Thomas and his teammates were keen to cluster keywords evenly around categories, without neglecting or prioritising one of them. As categories grew increasingly connected, someone commented: "Actually, we could have drawn something like a spider's web right away." While their keywords were supposed to refer

to the data they had collected, the mind map actually did not mention any data at all.

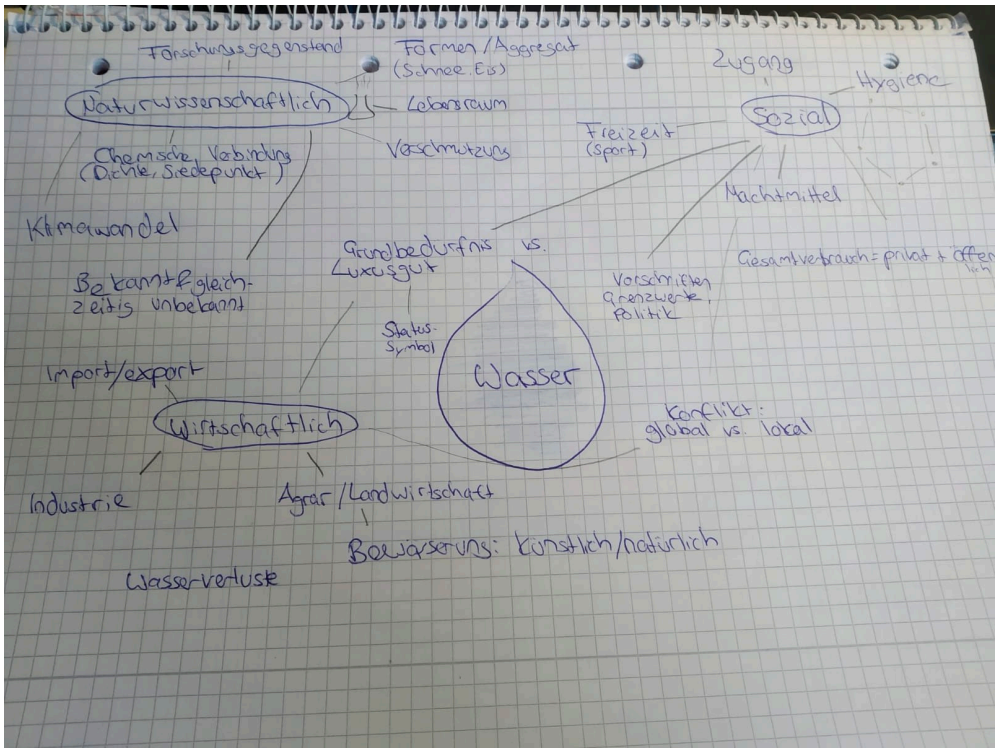

Figure 2: Mindmap
Source: Participants' own work

Both mind maps and scrapbook collages refrain from comparative modes of ordering. However, other groups of participants were intrigued by questions of priority, quantity, and (water) quality. They resorted to more abstract, hierarchical models for integrating the data they found and their discussions about the value of water. One such model is a rotating square (see Figure 3); another consists of a two-dimensional matrix (see Figure 4).

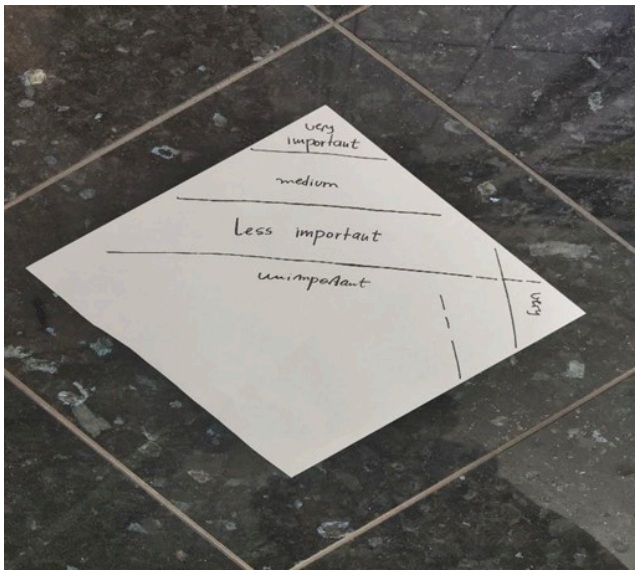

Figure 3: Rotating square
Source: Participant's own work

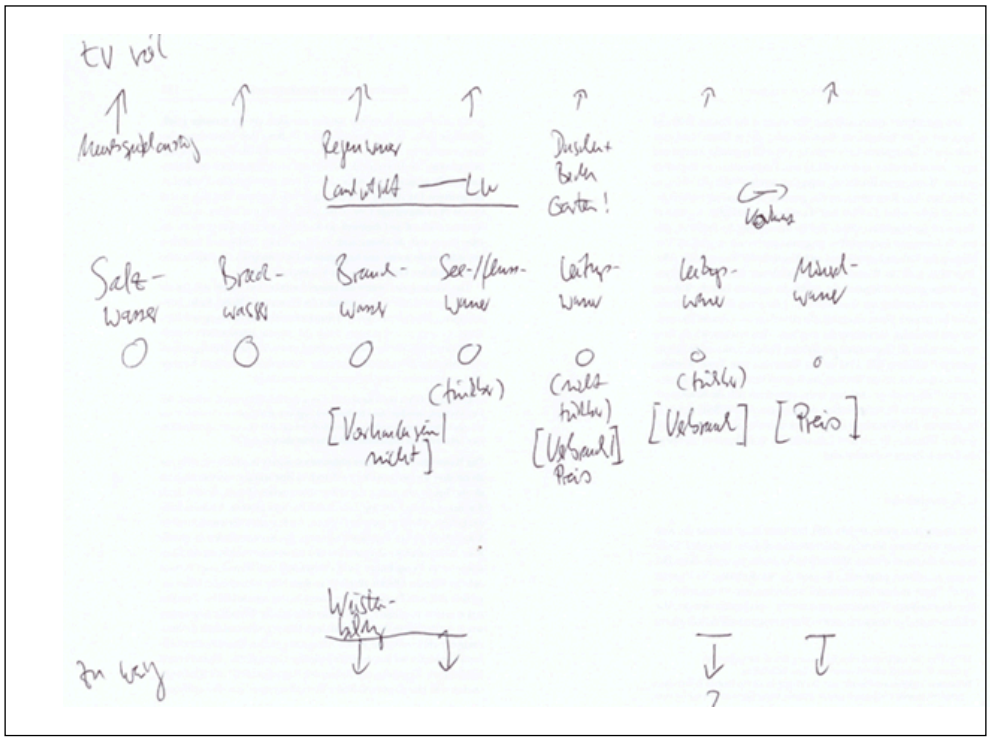

Figure 4: Two-dimensional matrix
Source: Participant's own work

The square model was devised by a group of participants struggling with an abundance of concerns for water (and data about it): "And yet," as one of them put it, "some water is more important", and some

data is more important than other. While drinking water and reliant measures about its quality are urgently needed, it may be considered less urgent to satisfy the water requirements of coffee processing plants or the cooling facilities of data centres. Yet what if data centres keep environmental data? What if coffee beans earn livelihoods? What if bean washing uses rainwater in regions where water shortage is not a problem? Is governmental public data more relevant than the privately commissioned testing of select taps? The group cut out a squarish shape of paper to explore these questions, each of its right-angle corners signifying a pyramidal hierarchy of value. With one of its corners pointing up, rotating the square by 90° would place a different hierarchy of value "on top". As they were working on their paper model, the third author retrieved a photo from the web – a car with square tires: "It can't work, and yet it works, with a lot of friction."

Another group, too, resorted to hierarchisation and devised a two-dimensional matrix, drawn by hand on a sheet of paper. Parsing issues of water use and water waste as a problem of both quality and quantity, the group developed a scheme to compare both conditions and amounts of water. On its horizontal axis, the scheme featured various qualities of water (beginning with saltwater on the left, ending with tap water and then mineral water on the right side of the paper). On its vertical axis, the scheme plotted relative amounts (with "too much" at the top, "too little" at the bottom). Participants drew water qualities as columns, with vertical arrows indicating how distinct qualities of water might move towards "too much" or "too little" (see Figure 4).

Only one group tackled quantitative data directly. Proceeding from precipitation data and data on the environmental impact of data centres, the group converted the amount of water used for cooling data centres into the amount of precipitation that fell onto gardens in Berlin – comparing data in a way that felt absurd yet intriguing. To do so, the group had to convert gallons to litres, relating the water used by all US-based data centres during one month to the rain that reached 500m2 gardens in Berlin throughout the particularly dry month of May (see Figure 5). According to their calculations, 3,287 gardens received as much rainwater as US-based data centres needed for cooling. Unsure how to interpret this result, the group wrote: "seems a lot – even possible … ?" Similar to the quadrangular shape model and the two-dimensional matrix, this exercise in metric conversion grappled with problems of comparison and weighting. Which needs for water, and which water data, should receive our attention?

What is more important? What is equivalent? How can the rationale of equivalence meet personal attachment in considering public concerns? As this workshop exercise shows, participants managed to attach data to one another but friction between the local and the global, the personal and the impersonal become quickly apparent.

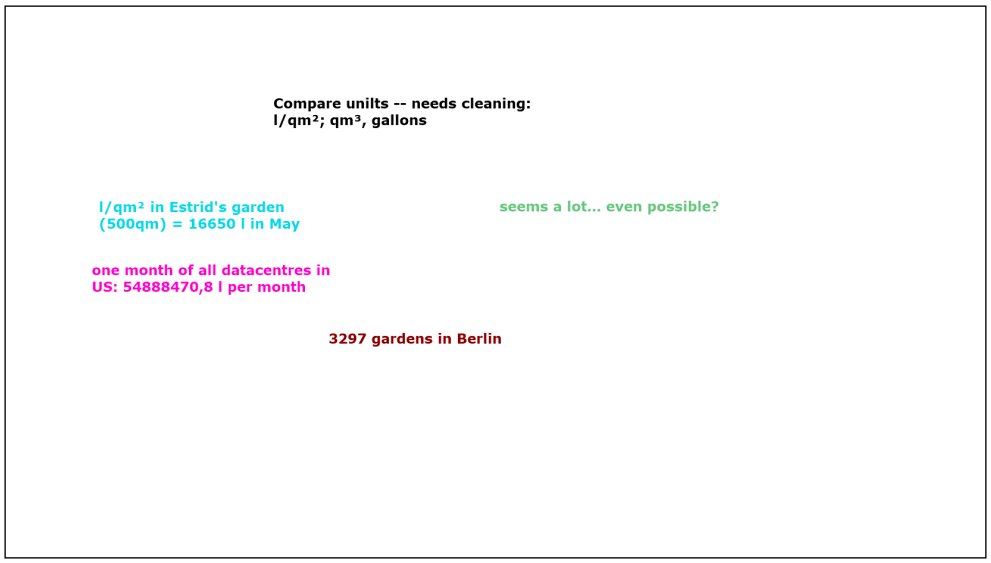

Figure 5: Calculations
Source: Participants' own work

Attaching online data to stakes, selves, and other data

The workshops provided an opportunity to observe and engage with the valuation of online data. Thinking with attachment, we suggest viewing valuation in terms of the material and affective, sensory, and reflexive relations through which value finds re-articulation. In particular, we draw attention to how data value emerges from attachment to the situated circumstances, the material objects, and the stakes at hand in moments of valuation. Furthermore, we elaborate on how data finds attachment to selves in situated subject networks. Last, we discuss the various techniques of dis/assembling data – attaching data to one another – that we observed in the valuation of online data.

Attaching data to the stakes at hand

To conceive of valuation in terms of attachment means, we argue, not only that it takes place *under* specific circumstances but that it works *with* them. Valuation crucially relies upon what is at hand – objects, infrastructures – and at stake in a given situation. Workshop participants, seated in front of screens at home, found themselves in domestic surroundings and with access to innumerable objects, some physically, some virtually present. Participants worked across various media, shifting back and forth between analogue and digital. Because screen sharing, for example, was unavailable during breakout sessions, participants resorted to email, shared dashboards, and handwritten

notes held to the screen. In similar vein, incomplete, scattered, and pre-formatted data challenged participants to forge attachments across data sets, metrics, and graphic renderings – a challenge that some participants avoided.

In fact, much of the data found online was hardly taken up during group discussion. Their adhesive qualities, we might say, proved poor. But while some participants complained that they found online data "hard to relate to" and "intangible", other participants presented carefully selected data as relatable and tangible. They researched, e.g., water data concerning everyday objects within immediate reach (clothing, coffee atop kitchen counter), domestic practices (showering, cooking, watering), or surrounding regions (garden, home town, home country). Making their surroundings a starting point for researching data online, participants were able to anchor the data they found within the realm of domestic life and attach it to "home".

Moments of valuation, however, do not only feature particular surroundings; they also feature a specific frame – they are about something (Goffman 1974). They put something at stake. Endowing data with value means attaching it to the stakes at hand and situating it firmly within the frames that shape its moments of valuation. As we observed, data gained value when it gained a significance that resonated with what the workshops were deemed to be about. While the two workshops may appear similar in design, participants found their attachments tried in distinctly different ways: some participants regarded the workshops as an occasion to probe and hone their academic taste, whereas others perceived the workshops as yet another test in higher education.

Depending upon their frame, moments of valuation may call upon different registers of valuation (Heuts and Mol 2013). Educational examinations promote impersonality and disinterestedness; tastings (of wine, coffee, or data) cultivate personal interest and sensual proximity. Tastings probe perceptions. How does it feel? Or better, how to feel oneself into it? Tasteful things "offer themselves only to those who offer themselves" to them (Hennion 2007: 106). Taste fosters intimate attachment and is steeped in collective discourse at the same time. The acquisition of taste depends on occasions to probe perceptions with others, and some participants made the workshop such an occasion. The coffee atop one participant's kitchen counter proved particularly conducive in this respect. Relating water data about coffee processing to domestic consumption, the taste for a particular brand of coffee lends itself to a preference for online data. Data, here, was cherished for its connection with the offline, sensory pleasures of everyday life.

In contrast, many participants were too uncomfortable to "taste" data. They felt the workshop was an educational test. When perceived as an examination, the workshop became an event that put participants' academic credit on trial and challenged them to apply the conventions of quantitative social science. Many participants found it

difficult to find and present data in ways that they believed satisfied these conventions; and many felt it inadequate to associate themselves – their lives, their homes – with the data they found. As Luc Boltanski points out, what is at stake in educational examinations is "[…] the competence [of students and involved personnel] to produce arrangements that are acceptable – 'convincing' – to others", i.e. arrangements that lay claim to objectivity and universality (Boltanski 2012: 33). Educational examinations typically minimise "extraneous worth" by barring testees from "wearing overly expensive jewellery or very shabby clothes", objects that point towards differences in financial capacity and class membership (Boltanski and Thévenot 2006: 137). Thus conceived, the workshop remained blatantly at odds with the implications of home-to-home videoconferencing, a mode of interaction that constantly risks revealing intimacies of domestic life. It became an exercise in controlled detachment, minimising exploratory familiarisation and playful attachment.

Attaching data to selves

The personal has a fraught relationship with data. While data can help translate between personal and public (as, e.g., in Marres 2009), the personal can be perceived as an infringement upon the purchase of data. In this vein, Anders Koed Madsen emphasises that "personal experience" should be assigned a "restricted role" in data workshops (Madsen 2023). In this article, however, we argue that data can gain value when it is attached to selves. Such self-attachment requires one to invest oneself in data to have oneself re-emerge in subject-data networks. Contrary to the notion that quantitative data science promotes detachment and impersonality, we observe how self-data relations can thrive upon passionate dedication – an attitude that some workshop participants adopted and others refuted.

Academic and educative exchange through videoconferencing implicates the domestic and infringes upon the private. Seizing upon the household and the private, some workshop participants readily mobilised personal attachments in selecting and presenting data (Hennion 2007, 2017b). They relished being *data amateurs*, guided by a desire for reflexivity and pleasure. Data amateurs might wonder: How do I feel myself into the data? What data is worth my dedicated attention, what data would be exciting for me to bring in, and how does a piece of data recast my subjectivity, my entanglement in the object-relations surrounding me? Data amateurs delight in reflecting themselves in bits of data. As conscientious coffee lovers, they research the amount of water consumed by washing their coffee beans. As concerned parents, they produce the digital scan of a commissioned tap water analysis. As international scholars, they retrieve water data from the regions they have researched and lived in. And as ardent

gardeners, they make calculations about rainfall per square metre. Like care (cf. Mol et al. 2010), such attachment is ambivalent. Amateurs are not immune to narcissism, and their dedication risks being overly egocentric. Nevertheless, we find that distinct self-investment challenges collaborating participants to appreciate others' discriminatory capacities and skilfully interweave personal attachment. Amateurism animates explorative reflections of data in the negotiation of public issues, relating private pleasure to the common good.

Some participants regarded self-invested data amateurism as inappropriate and opposed any attempt to create a relation between the data they found and their private lives. For example, Uwe copiously resisted any effort to relate data about domestic water consumption to specific activities in his daily life. "We could", he contended at one point, "try to assign an estimated amount of water … but I would not know how much water I need to wash the dishes – this would be mere speculation." With speculation deemed inappropriate and specific water data out of reach, Uwe steered his group away from assigning water volumes to domestic activities. He argued that it would be "difficult to require people to time their showering". Instead, he drew attention to agricultural water consumption. Aware of the restrictive environmental policies that might underlie the measurement of domestic water use, Uwe defies the notion that domestic consumption data counts and would be worth having. Implicitly, Uwe questioned the legitimacy of domestic datafication.

It is important to note, however, that detachment is productive in its own way. Throughout the workshop, several participants remained defensive, even defiant, and their detachment helped articulate questions and uncertainties. It initiated a critical, often somewhat sceptical examination of data. More than their enthusiasm, participants' defiance raised the question: What data is needed, and what data is warranted? What data does it take, and what does it take to stitch these data together? When defiance cast doubt upon data, it also brought to the fore what was lacking – e.g. parent populations, reference values, and more detailed information about data collection and data analysis. Instead, passionate data amateurs readily attached themselves to present and available data. When passionate attachment was invoked, the quality of data was hardly ever questioned.

As they attach, detach, and reattach themselves, "[a] constant testing, assessing, calibrating appreciation is at the heart of participants' engagement with their environment" (de Laet et al. 2021: 809). So, while we maintain that data can gain value through attachment, particularly to the attachment to selves, we acknowledge that detached defiance has its merits. The interplay of attachment and detachment makes for some of the complexity of valuation. The value of online data is shaped and reshaped by whether or not we hold on to

data tightly, scrutinise it at arm's length, or keep it bay, invite it home or have it slip away.

Attaching data to one another

Data rarely comes alone, and it gains value when it attaches to one another. As Wendy Espeland and Mitchell Stevens have argued, it typically associates through commensuration. Understood as a metric translation, commensuration "is fundamentally relative" and "creates relations between attributes or dimensions where value is revealed in the comparison" (Espeland and Stevens 1998: 317). Here, value "emerges from comparisons that are framed in terms of how much of one thing is needed to compensate for something else" (317). In fact, both workshops were concerned with comparison, compensation and resemblance, categorisation, and prioritisation. In particular, these concerns found articulation when participants were tasked to relate the water data they found. In both workshops, this task proved rather tricky for participants to solve in a manner that they perceived as satisfying. Online data did not associate easily with one another and challenged workshop participants to work out different relations of equivalence (Boltanski and Thévenot 2006: 33). While some groups of participants related data employing commensuration, others relied upon forms of relation-making that deliberately stopped short of comparison and compensation. In our observation of the workshops, we identify three techniques for attaching data to one another: *collages*, *hierarchies*, and *calculation*. While we characterise each technique separately in the following, they overlap in practice.

Some workshop participants devised *collages* – collections of notes or images spread out on a surface and loosely associated by resemblance or rough categorisation. Such collages rely on implicit, rule-of-thumb rationale of equivalence. What's un/like what? What belongs where? In our workshops, collages could take the form of scrapbooking or conventional mind mapping. The scrapbook (Figure 1) consists of photographs, many of which stylistically resemble one another, and uses them as stand-ins for domestic, water-consuming activities (bathing, cooking, etc.). The mind map (Figure 2), in turn, invokes analytic differentiation. Its structure relies upon three umbrella categories (i.e. the "social", "economic", and "scientific" value of water), the equivalence of which was diligently cared for when subcategories and keywords were placed evenly around them.

In creating collages, workshop participants used different materials (digital photographs, paper and pencil) to present and discuss diverse uses of water and its multiple worths, quite literally, "on a singular plane" – avoiding, notably, in cases we observed, any material mix. What is more, the collages do not attach to any of the specific data participants were able to find. They also avoid prioritisation,

weighting, and comparison. How do showering and bathing compare? What kind of tensions would be between water's "social" and its "economic" value? What are the conceptual politics of such a distinction anyhow? Steering clear of questions such as these, collages defer the challenge to quantify value and engage with concrete data. Instead, they convey the subtle notion that water remains *in*valuable.

Some participants devised diagrams of *hierarchies*. These graphics use abstract shapes to describe value orders as relational, dynamic, and heterogeneous. One of these diagrams, a two-dimensional matrix (Figure 4), features a single definite ordering of qualities of water from saltwater to mineral water. Another diagram tackles the simultaneity of incommensurable hierarchies with a rotatable quadrangular shape (Figure 3). These diagrams, too, refrain from mentioning any specific pieces of data. Instead, they strive to articulate the rationale of association that would allow for prioritizations of worth.

Finally, participants resorted to *calculation* in order to attach the data they found (Figure 5). As calculation requires reformatting data (Dumit 2018), participants carefully converted units of measure to relate the amount of water used for cooling data centres to the amount of rain that fell on private gardens – juxtaposing utterly different water qualities by equating water quantities. This equation consists of an odd if intriguing commensuration: if US-based data centres consume just as much rain as 3,297 gardens in Berlin receive, does that imply that cooling water is equally as valuable as precipitation? What impact – direct or indirect – has industrial water use upon rainfall? Clearly, one is no compensation for the other, and the relinquishment of the former would not benefit the latter in any straightforward way. While participants' calculations queried the limits of commensuration, the equation of cooling water with precipitation served participants as a means of valuation. This equation not only emphasised both the value of cooling water as well as Berlin rain; more importantly, it valorised selected data and appreciated their capacity to connect in far-reaching, thought-provoking ways.

Calculation, collage, and hierarchy building are techniques for attaching data that invoke and manipulate relations of worth. These techniques handle data in different ways – joining data through commensuration or grouping it using resemblance and categorisation. Calculation, collage, and hierarchy building help attach data to one another and query these attachments simultaneously. They can also separate data or sideline it. As techniques for handling diverse, abundant, and fragmented data, they hardly settle value. Rather, they grapple with commensuration and its limits and invoke issues of in/commensurability (Boltanski and Thévenot 2006) while they put material, adhesive qualities of data to the test.

Conclusion

Building upon a vast array of digital technologies, digitised valuation appropriates new possibilities, necessities, and constraints to forge relations; or to cut them. Digital infrastructure comes with inscribed configurations of worth (see, e.g., Balsiger and Jammet (2022) as well as Krüger and Petersohn (2022) in this theme issue of *Valuation Studies*), and users are called to adopt or discard, process, and rearrange them. Crucially, virtual collaboration takes place in technologically mediated, synthetic moments whose configuration shapes the performance of valuation. Moreover, the "mercurial" character of things digital – online data, that is – challenges digitised valuation to continuously re-situate it. Online data abounds, and massive datafication is likely to entail fundamental changes in the ways of domestic life, the manners of private reflection, and our modes of civic deliberation (Marres 2009; Gabrys et al. 2016). Before this backdrop, we studied how (online and other) data figures in the shifting boundaries between valuable and invaluable.

In this article, we have analysed how online data finds attachment, thereby gaining worth in virtual collaboration. Our analysis relies upon virtual ethnography and participant observation in two experimental workshops. Held as home-to-home videoconferences, the workshops allowed us to observe how data relates to the domestic and how data value is negotiated at the nexus of private and public. In our analysis, we elaborate how analogue and digital materials, immediate pleasure and collective debate, domestic concerns and global resources are brought together/apart in digitised valuation.

We contribute a perspective that conceives of valuation in terms of attachment. Data becomes valuable when it attaches to the circumstances and stakes at hand, to selves, and to other data. Attachment makes data stick. Conceptually, attachment recasts digitised valuation as radically relational, material, and affective. It draws attention to the ways in which digitised valuation not only "is situated" but works with the situation—its stakes, locality, and provisions. In doing so, attachment foregrounds the role of material and sensorial quality, affect, and reflexivity in endowing data with value. Attachment can be enthusiastic or wanting. Often enough, it is ambivalent, multiple, and overlapping. It may evolve or shift. Attachment alone may not do the trick. Finally, we have pointed out how questions about data quality and the legitimacy of datafication are triggered by practices of *de*tachment.

Acknowledgments

We are deeply indebted to the participants of both workshops. A special thank you goes to members of the Ruhr University Science and Technology Studies Lab (RUSTlab) Bochum, namely Estrid Sørensen ("Flora"), Susana Carmona and Raphael Hemme, for their comments on an early version of this article. Finally, we thank the reviewers of *Valuation Studies* for their profound reading and commenting, which inspired us to develop our argument further.

References

Antal, Ariane Bertoine, Michael Hutter, and David Stark (eds.). 2015. *Moments of Valuation: Exploring Sites of Dissonance*. Oxford: Oxford University Press.

Balsiger, Philip, and Thomas Jammet. 2022. "What is Digital Valuation Made of? The Integration of Valuation Poles on a Reservation Platform and Its Effects on the Hotel Industry in Switzerland." *Valuation Studies* 9(1): 47–77.

Barnes, Jessica, and Samer Alatout. 2012. "Water Worlds: Introduction to the Special Issue of Social Studies of Science." *Social Studies of Science* 42(4): 483–488.

Boellstorff, Tom. 2008. *Coming of Age in Second Life: An Anthropologist Explores the Virtually Human*. Princeton, NJ: Princeton University Press.

Boltanski, Luc. 2012. *Love and Justice as Competence*. Cambridge, MA: Polity.

Boltanski, Luc, and Laurent Thévenot. 2006. *On Justification. Economies of Worth*. Princeton, NJ: Princeton University Press.

Bowker, Geoffrey C. 2008. *Memory Practices in the Sciences*. Cambridge, MA: MIT Press.

Bowker, Geoffrey C., and Susan Leigh Star. 2000. S*orting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Callon, Michel. 1984. "Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay." *The Sociological Review* 32(1): 196-233.

Conward, Martin. 2018. "Against Network Thinking. A Critique of Pathological Sovereignty." *European Journal of International Relations* 24(2): 440–463.

de Laet, Marianne; Annelieke Driessen, and Else Vogel. 2021. "Thinking with attachments: Appreciating a generative analytic." *Social studies of science* 51(6): 799–819.

Doganova, Liliana, Martin Giraudeau, Hans Kjellberg, Claes-Fredrik Helgesson, Francis Lee, Alexandre Mallard, Andrea Mennicken, Fabian Muniesa, Ebba Sjögren, and Teun Zuiderent-Jerak. 2018. "Five Years! Have We not Had Enough of Valuation Studies by Now?" *Valuation Studies* 5(2): 83–91.

Douglas-Jones, Rachel, Antonia Walford, and Nick Seaver. 2021. "Introduction: Towards an Anthropology of Data." *Journal of the Royal Anthropological Institute* 27(S1): 9–25.

Dumit, Joseph. 2018. "The Other Ninety Per Cent: Thinking with Data Science, Creating Data Studies–an Interview with Joseph Dumit." In E*thnography for a Data-Saturated World*, edited by Hannah Knox and Dawn Nafus, 252–275. Manchester: Manchester University Press.

Edwards, Paul N. 2010. *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, MA: MIT Press.

Espeland, Wendy Nelson, and Mitchell L. Stevens. 1998. "Commensuration as a Social Process." *Annual Review of Sociology* 24(1): 313–343.

Estalella, Adolfo, and Tomás Sánchez Criado (eds.). 2018. *Experimental Collaborations: Ethnography through Fieldwork Devices*. New York: Berghahn Books.

Gabrys, Jennifer, Helen Pritchard, and Benjamin Barratt. 2016. "Just Good Enough Data: Figuring Data Citizenships through Air Pollution Sensing and Data Stories." *Big Data & Society* 3(2): 1–14.

Gad, Christopher, and Casper Bruun Jensen. 2010. "On the Consequences of Post-ANT." *Science, Technology, & Human Values* 35(1): 55–80.

Gitelman, Lisa (ed.). 2013. "*Raw Data" is an Oxymoron*. Cambridge, MA: MIT Press.

Goffman, Erving. 1974. *Frame Analysis. An Essay on the Organization of Experience*. New York: Harper & Row.

Gomart, Emilie, and Antoine Hennion. 1999. "A Sociology of Attachment: Music Amateurs, Drug Users." *The Sociological Review* 47(1): 220–247.

Greeson, Emma, Stefan Laser, and Olli Pyyhtinen. 2020. "Dis/Assembling Value: Lessons from Waste Valuation Practises." *Valuation Studies* 7(2): 151–166.

Helgesson, Claes-Fredrik. 2016. "Folded Valuations?" *Valuation Studies* 4(2): 93–102.

Helmond, Anne. 2015. "The Platformization of the Web: Making Web Data Platform Ready." *Social Media + Society* 1(2): 1–11.

Hennion, Antoine. 2007. "Those Things that Hold us Together: Taste and Sociology." *Cultural Sociology* 1(1): 97–114.

Hennion, Antoine. 2017a. "Attachments, You Say? … How a Concept Collectively Emerges in One Research Group." *Journal of Cultural Economy* 10(1): 112–121.

Hennion, Antoine. 2017b. "From Valuation to Instauration." V*aluation Studies* 5(1): 69–81.

Heuts, Frank, and Annemarie Mol. 2013. "What is a Good Tomato? A Case of Valuing in Practice." *Valuation Studies* 1(2): 125–146.

Hine, Christine. 2015. *Ethnography for the Internet: Embedded, Embodied and Everyday*. London: Bloomsbury.

Howlett, Marnie. 2020. "Looking at the 'field' through a Zoom Lens: Methodological Reflections on Conducting Online Research During a Global Pandemic." *Qualitative Research* 22(3): 387–402.

Jensen, Torben Elgaard, Andreas Birkbak, Anders Koed Madsen, and Anders Kristian Munk. 2021. "Participatory Data Design: Acting in a Digital World." In *Making and Doing STS: Activating STS through Knowledge Expression and Travel*, edited by Gary Lee Downey, and Teun Zuiderent-Jerak, 117–136. Cambridge, MA: MIT Press.

Kjellberg, Hans, Alexandre Mallard, Diane-Laure Arjaliès, Patrik Aspers, Stefan Beljean, Alexandra Bidet, Alberto Corsín, 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." *Valuation Studies* 1(1): 11–30.

Knorr Cetina, Karin. 2009. "The Synthetic Situation. Interactionism for a Global World." *Symbolic Interaction* 32(1): 61–87.

Krause, Franz, and Veronica Strang. 2016. "Thinking Relationships through Water." *Society & Natural Resources* 29(6): 633–638.

Kropf, Jonathan, and Stefan Laser (eds.). 2019. *Digitale Bewertungspraktiken: Für eine Bewertungssoziologie des Digitalen. Soziologie des Wertens und Bewertens*. Wiesbaden: Springer VS.

Krüger, Anne K., and Sabrina Petersohn. 2022. "From Research Evaluation to Research Analytics. The Digitization of Academic Performance Measurement." *Valuation Studies* 9(1): 11–46.

Lassiter, Luke Eric. 2005. "Collaborative Ethnography and Public Anthropology". *Current Anthropology* 46(1): 83–106.

Latour, Bruno. 1999. "Circulating Reference. Sampling Soil in the Amazon Forest." In *Pandora's Hope. Essays on the Reality of Science Studies*, 24–79. Cambridge, MA: Harvard University Press.

Lee, Francis, and Claes-Fredrik Helgesson. 2020. "Styles of Valuation: Algorithms and Agency in High-throughput Bioscience." *Science, Technology, & Human Values* 45(4): 659–685.

Lezaun, Javier, Fabian Muniesa, and Signe Vikkelsø. 2013. "Provocative Containment and the Drift of Social-Scientific Realism." *Journal of Cultural Economy* 6(3): 278–293.

Loukissas, Yanni A. 2019. *All Data Are Local: Thinking Critically in a Data-Driven Society*. Cambridge, MA: MIT Press.

Madsen, Anders Koed. 2023. "Digital Methods as 'Experimental a priori' – How To Navigate Vague Empirical Situations as an Operationalist Pragmatist." *Convergence* 0(0) (online first).

Mämecke, Torben, Jan-Hendrik Passoth, and Josef Wehner. 2018. "Bedeutende Daten– Einführende Überlegungen." In *Bedeutende Daten. Modelle, Verfahren und Praxis der Vermessung und Verdatung im Netz*, edited by Torben Mämecke, Jan-Hendrik Passoth, and Josef Wehner, 1–14. Wiesbaden: Springer VS.

Marres, Noortje. 2009. "Testing Powers of Engagement: Green Living Experiments, the Ontological Turn and the Undoability of Involvement." *European Journal of Social Theory* 12(1): 117–133.

Marres, Noortje, and David Stark. 2020. "Put to the Test: For a New Sociology of Testing." *The British Journal of Sociology* 71(3): 423–443.

Marres, Noortje, and Esther Weltevrede. 2013. "Scraping the Social? Issues in Live Social Research." *Journal of Cultural Economy* 6(3): 313–335.

Mills, John Stuart. 1959. *The Sociological Imagination*. Oxford: Oxford University Press.

Mol, Annemarie, Ingunn Moser, and Jeannette Pols (eds.). 2010. *Care in Practice. On Tinkering in Clinics, Homes and Farms*. Bielefeld: transcript.

Munk, Anders Kristian, Axel Meunier, and Tommaso Venturini. 2019. "Data Sprints: A Collaborative Format in Digital Controversy Mapping." In *DigitalSTS: A Field Guide for Science and Technology Studies*, edited by Janet Vertesi and David Ribes, 472–492. Princeton, NJ: Princeton University Press.

Mützel, Sophie, Philippe Saner and Markus Unternährer. 2018. "Schöne Daten! Konstruktion und Verarbeitung von digitalen Daten." In *Datengesellschaft*, edited by Daniel Houben and Bianca Prietl, 111-132. Bielefeld: transcript.

Paßmann, Johannes, and Cornelius Schubert. 2021. "Liking as Taste Making: Social Media Practices as Generators of Aesthetic Valuation and Distinction." *New Media & Society* 23(10), 2947–2963.

Pinel, Clémence, Barbara Prainsack, and Christopher McKevitt. 2020. "Caring for Data: Value Creation in a Data-Intensive Research Laboratory." *Social Studies of Science* 50(2): 175–197.

Pink, Sara, Minna Ruckenstein, Robert Willim, and Melisa Duque. 2018. "Broken Data: Conceptualising Data in an Emerging World." *Big Data & Society* 5(1): 1–13.

Potthast, Jörg. 2017. "The Sociology of Conventions and Testing." In *Social Theory Now*, edited by Claudio E. Benzecry, Monika Krause, and Isaac Ariail Reed, 337–360, Chicago, IL and London: University of Chicago Press.

Ribes, David, and Steven J. Jackson. 2013. "Data Bite Man: The Work of Sustaining a Long-Term Study." In *"Raw Data" is an Oxymoron*, edited by Lisa Gitelman, 147–166. Cambridge, MA: MIT Press.

Rogers, Richard. 2013. *Digital Methods*. Cambridge, MA: MIT Press.

Rossmann, Sophia. 2021. "Going Virtual: The Ethnographic Gaze in Pandemic Times." *EASST Review Volume* 40(1): [no pp.]

Star, Susan Leigh, and James R. Griesemer. 1989. "Institutional Ecology, Translations and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39." *Social Studies of Science* 19(3): 387–420.

Strathern, Marilyn. 1996. "Cutting the Network." *Journal of the Royal Anthropological Institute* 2(3): 517–535.

Strathern, Marilyn. 2002. "Abstraction and Decontextualization. An Anthropological Comment." In *Virtual Society? Technology, Cyberbole, Reality*, edited by Steve Woolgar, 302–314. Oxford: Oxford University Press.

Thévenot, Laurent. 2002. "Which Road to Follow? The Moral Complexity of an 'Equipped' Humanity." In *Complexities: Social Studies of Knowledge*

Practices, edited by John Law and Annemarie Mol, 53–87. Durham, NC and London: Duke University Press.
Vertesi, Janet. 2014. "Seamful Spaces: Heterogeneous Infrastructures in Interaction." *Science, Technology, & Human Values* 39(2): 264–284.
Waibel, Désirée, Thorsten Peetz, and Frank Meier. 2021. "Valuation Constellations". *Valuation Studies* 8(1): 33–66.
Winthereik, Britt Ross, Peter A. Lutz, Lucy Suchman, and Helen Verran. 2011. "Attending to Screens and Screenness." *STS Encounters* 4(2): 1–6.
Ziewitz, Malte. 2011. "How to Attend to Screens? Technology, Ontology and Precarious Enactments." *STS Encounters* 4(2): 203–228.

Susann Wagenknecht holds the Junior Professorship in Micro-Sociology and Techno-Social Interaction at the Institute of Sociology at TUD Dresden University of Technology, Germany. She obtained her dissertation ('A Social Epistemology of Research Groups') from Aarhus University, Denmark. Drawing on ethnography, sociology and science and technology studies, she researches valuation at the private/public interface as well as the design, maintenance and use of infrastructure (of data, traffic and energy).

Laura Kocksch works as a post-doctoral researcher in the techno-anthropology lab (TANTlab) at Aalborg University in Copenhagen. She recently defended her doctoral dissertation ('Fragile Computing – Theorizing Cybersecurity as Practice') summa cum laude at Ruhr University Bochum. Her work centres in ethnographies of data matters – material infrastructure, cybersecurity and planetary resources. She uses a combination of qualitative methods and data design to craft prolonged situations of observation and intervention.

Stefan Laser is a dost-doctoral researcher with the Collaborative Research Center 'Virtual Live Worlds' at Ruhr-University Bochum, Germany, investigating the entanglement of knowledge practices and planetary resources through data centres. Formerly, he was studying train infrastructures in transformation at Siegen University, Germany; and from 2013–2018 he conducted his doctoral studies on the global value chains of electronic waste at the sociology department at Kassel University, Germany. Stefan has published in the fields of valuation studies, waste studies, economic sociology and digital methods.

Ann-Kristin Kühnen is a doctoral candidate and research assistant at the chair of Micro-Sociology and Techno-Social Interaction at the Institute of Sociology at TUD Dresden University of Technology, Germany. Her research focuses on (feminist) science and technology studies (STS), new materialism and political ecology.

Valuation Studies 11(1) 2024: 91–137

Judging by the Rules? The Emergence of Evaluation Practices

Stacy E. Lom

Abstract
How does evaluation work differently, and how do evaluation practices emerge, in different contexts? Drawing on a mixed-methods study of evaluation in figure skating and classical music, I discuss the divergent evaluative cultures in these settings, especially in terms of how formal and standardized they are, to consider how and why evaluation practices change over time and why different settings use different evaluation practices. I emphasize the importance of organizational structure, including context, competition structure, degree of centralization, and governance structure. My findings suggest that highly centralized settings governed by more powerful organizations and where competitions build on each other tend to use more formal and standardized evaluation practices compared to other settings with fewer constraints. Understanding how evaluation practices develop and what they look like in different contexts is important because in addition to influencing the objects of evaluation and perceived fairness and legitimacy, these practices often affect outcomes, which have significant consequences for participants.

Keywords: evaluation; formalization; rule changes; figure skating; classical music

Stacy Lom is Assistant Professor of Sociology at the University of Central Arkansas.

https://doi.org/10.3384/VS.2001-5992.2024.11.1.91-137
Hosted by Linköping University Electronic press
http://valuationstudies.liu.se

Introduction

Evaluation often looks very different in different settings. That statement might seem fairly obvious, but what is much less obvious is what this looks like on the ground. Consider two examples: (1) As American Nathan Chen skates his long program at the 2022 Olympics as the final skater in the men's competition and the overwhelming favorite for the gold medal, the technical panel identifies each technical element he performs, and the judges evaluate each of those elements with a score from -5 to +5. When his program ends, the judges also give him a score for five separate program components. After the judges enter their scores into a computer, Chen scores 218.63 points for his program and wins the gold medal with 332.6 total points, 22.55 points more than Yuma Kagiyama from Japan (ISU 2022a). (2) In contrast, the third and final contestant in a high-level organ competition finishes her performance, and now the judges must decide who will finish first, second, and third. The judges deliberate, arguing about whether to reward technique or artistry. They ultimately reach a consensus that the more artistic performer should win.

In both examples, the goal is to figure out who is going to win a competition, but the methods for achieving that goal are very different. These examples show how evaluation works differently in different settings, which raises the issue of how to explain these differences. In this article, I discuss how organizational structure affects evaluative cultures, especially in terms of the formality and standardization of rules. In particular, I examine how organizational context, competition structure, degree of centralization, and governance structure influence approaches to evaluation practices. This is important because in addition to influencing the objects of evaluation and perceived fairness and legitimacy, evaluation practices often affect outcomes, which can have significant consequences for participants. Based on a mixed-methods study of evaluation in figure skating and classical music and building on discussions of evaluative cultures (Lamont 2009), rules (e.g., Meyer and Rowan 1977; Dobbin and Kelly 2007; Edelman et al. 2011), and objectivity (Porter 1995; Daston and Galison 2007), I focus on how two contexts have developed very different evaluation practices. Based on my main findings, I suggest that settings with high degrees of centralization and shared international governance and where competitions build on each other tend to adopt more formal and standardized evaluation practices compared to other settings with fewer constraints.

Lamont (2009) describes evaluative cultures as cultural scripts surrounding evaluation, including how people think about and practice evaluation, as well as the rule systems and methods they use. In this article, I conceive of evaluative cultures as the meanings, values, and practices surrounding evaluation in particular settings, with an emphasis on rules and how they develop. I focus especially on

variations in evaluative cultures in terms of the formality and standardization of their rules. Based on a study of peer review in academia, Lamont's discussion of evaluative cultures highlights how people deal with disciplinary differences when they come together in an evaluative context where outcomes are based on evaluator consensus. However, she does not address where those original differences come from. This is where my research comes in. While Lamont's work focuses on understanding how evaluative cultures work in general and the process of how decisions are produced in particular, my research examines how evaluative cultures develop and change, with a focus on rules and practices. In addition to a broader emphasis on how perceived fairness and legitimacy affect how rules develop, my work expands Lamont's discussion of differences in evaluative cultures by focusing on why and how different rule systems emerge.

Evaluation, objectivity, and formalization

Valuation is a fundamentally social process, one that is dependent on historical and cultural context and interacts with many other areas of social life (Fourcade 2011; Lamont 2012; Helgesson and Muniesa 2013; Kornberger et al. 2015). What this looks like and how it affects valuation processes is very different in different contexts, which I highlight in this article, focusing on rules and practices surrounding evaluation in competition settings. In general, evaluative rules range from highly formal to highly informal. In many settings in recent years, evaluation has become increasingly formal, often in response to concerns about fairness and legitimacy, but there have been exceptions to this trend. Figure skating is a case of highly formal and standardized evaluation, including a judging system that has been overhauled in recent years following a legitimacy crisis. Classical music operates under a range of evaluative rules and practices, from very informal discussion-based evaluation to more formal and numerical evaluation. Some music competitions use intricate scoring systems, for example. On the whole, however, evaluation in classical music is much less formal and much less standardized than in skating. An evaluative rule system's degree of formality affects both evaluators and the people they evaluate in relation to everything from interactions among evaluators, to evaluators' discretion, to performers' potential creativity. I focus here on distinctions between formal and informal rules and varying degrees of standardization, especially where those differences come from and what they look like on the ground.

At least on the surface, the skating world has emphasized evaluation with formal, standardized rules much more actively than the music world generally has. Why has this been the case, and how has this happened? The notion of objectivity, which involves neutrality based

on following rules and "knowledge that bears no trace of the knower" (Daston and Galison 2007: 17), often plays a key role in how people think about fairness and legitimacy. This affects just about every area of public life, especially settings that rely on overtly subjective evaluation or evaluation that involves a lot of discretion – control, flexibility, or lack of constraint in making decisions (Mastrofski 2004; Shen and Cho 2005; Li and Tang 2010), which applies to both skating and music. The idea is that if judges are objective in their evaluations, participants should determine the outcomes based on what they do rather than what the people judging them do, similar to the popular belief that scientists are supposed to produce knowledge based on objective procedures which are devoid of emotion or political forces (Brown and Malone 2004).

Porter (1995) delineates two distinct types of objectivity: disciplinary objectivity, which revolves around reaching consensus, and mechanical objectivity, which is based on strictly adhering to rules and tends to emerge where reaching consensus would be difficult or would not seem legitimate to external observers or stakeholders. Porter stresses how disciplinary and mechanical objectivity are often at odds with each other and emphasizes that purely mechanical objectivity is impossible because all types of rules have ambiguities, regardless of how clear the people who designed them think they are. While disciplinary objectivity and the discretion that usually goes along with it have mostly remained legitimate in classical music, pressures toward mechanical objectivity and limiting discretion – and the legitimacy that often goes along with them – have had a profound impact on figure skating.

Porter's (1995) discussion of objectivity centers largely around explaining quantification's appeal in the modern world. Other scholars have built on Porter's work, suggesting that institutional legitimacy and accountability based on numbers have become increasingly linked over the past few decades (Power 2003) and that this link has contributed to a proliferation of measurement systems in a wide variety of settings (Espeland and Sauder 2007; Colyvas 2012; Mau 2019). These kinds of legitimacy and accountability have been significant in skating and music, but in different ways, largely due to their distinct organizational structures. Porter (1995) suggests that because people tend to associate quantification with "impersonality, discipline, and rules" (32) and think of it as one of the most credible strategies for achieving pure objectivity, it has been most attractive in fields plagued by outside pressure, suspicion, controversy, and unseemly politics. Especially because of its centralization and international governance, these factors have seemed to affect evaluation practices in skating much more than in music.

In keeping with how perceptions of objectivity tend to be linked to legitimacy and accountability, many organizations have used

formalization – implementing written rules and more specific procedures or instructions (Adler and Borys 1996) – as a strategy for improving legitimacy and accountability. Formalization is a key component of Weber's ideal typical bureaucracy (Weber 1968; Adler 2012), where organizations run like machines without any individual discretion (Feldman 1992), and is often geared toward controlling people's actions (Stinchcombe 2001). In general, rules, written documents, and standardization, especially related to jobs and work roles, have played increasingly pivotal roles in organizations in recent years (for instance, see Adler and Borys 1996; Adler 2006 on the software industry; Maccoby 2006 on health care organizations). In particular, a number of researchers have shown how modern organizations tend to emphasize that they have specific rules or codified procedures in place, regardless of what they look like or how they affect practices and outcomes (Jacobsson 2000; Star and Lampland 2009). Based on these trends, one might be surprised to find any organizations that have not engaged in formalization processes. Many people in a wide variety of settings believe that formal, written rules are fairer and more legitimate than informal rules that have not been clearly defined, especially when those formal rules revolve around numbers. Many people also assume that when rules and procedures are written down, they increase transparency and limit discretion, whether or not this happens in practice.

Similar to how Porter (1995) emphasizes that procedures have become increasingly more important than outcomes in many areas of social life, institutional theorists have argued that in many contexts, formal rules exist for purely ceremonial purposes and to enhance legitimacy rather than to improve organizational fairness or efficiency (Meyer and Rowan 1977). Researchers have found that when organizations have the kinds of policies in place that they are "supposed to" have (like anti-discrimination laws or human rights laws, for example), people – including legal professionals – tend to assume that those policies are effective in practice, even though they are often merely symbolic and often perpetuate the issues they were intended to address (Hafner-Burton and Tsutsui 2005; Dobbin and Kelly 2007; Edelman et al. 2011).

In keeping with these trends, according to other research, people's ideas about whether rules are fair or not (rather than how effective they are) affect how likely they are to accept organizational policies in various domains and how they evaluate the people and organizations behind them in terms of how legitimate they are (for example, see Elsbach and Elofson 2000; Tyler 2000; Sunshine and Tyler 2003). One prominent example of this is law, where formal policies play an especially significant role. Tyler (1988), for instance, suggests that perceptions of procedural justice in legal contexts influence how people understand and interact with the legal system more than the

outcomes of those procedures. Tyler (2003) also notes that perceived neutrality and lack of bias among decision-makers play an especially influential role in how people evaluate the fairness of policies.

This link in many people's minds between how fair rules are and their legitimacy should continue to play a significant role in scholarship on evaluation, rules, and organizations more broadly. Unless people believe that the evaluation systems that matter within the institutions that they care about are legitimate, it will be difficult to sustain the institutions that use them (Lamont 2009). This has major implications for competition settings like figure skating and classical music, which will be virtually impossible to maintain unless participants, and outsiders in some cases, think they are legitimate activities with fair rules. One complicating factor here is that a rule system's fairness is often perceived differently in different contexts (Tyler 1988), which helps explain why different fields use different evaluation practices and why so many different ways of evaluating merit could be considered legitimate. While people might think of strict quantitative rules as fairer in one setting, they might view deliberation as more reasonable in another, depending on the evaluators, what they are evaluating, and the larger goal or context.

According to this research, it makes sense that the figure skating world responded to an Olympic judging scandal by implementing a judging system with more explicit, numerically specific rules, especially given the legitimacy crisis and coercive isomorphism surrounding this case. Isomorphism is a constraining process that is generally linked to legitimacy and leads organizations operating under similar environmental conditions to become more similar to each other (Deephouse 1996). Coercive isomorphism in particular arises from indirect pressures from cultural expectations in an organization's environment and direct pressures from powerful organizations (DiMaggio and Powell 1983). Both types of pressures – informal pressures from values emphasizing fair play in sport and a formal mandate from the International Olympic Committee (IOC) – contributed to formalization in figure skating. Judging-related controversy has plagued both figure skating and classical music, however, and theoretically at least, the music world and its outside constituents should also be concerned about its legitimacy.

As Porter (1995) argues, not only do most people believe that standardized measurement helps protect against bias and neutralizes politics; it is also a common method of dealing with distance and crossing "the boundaries of nation, language, experience, and discipline" (1995: 220). This suggests that centralized, internationally organized settings like figure skating tend to be more susceptible to pressures to use more mechanical and quantitative evaluation practices. Despite how there are international music competitions, a judging system's capacity to cross boundaries is much more important

in skating because while international music competitions can use different rules, skating's centralization and competition structure (where competitions often build on each other) require a single evaluation system that all people everywhere will be able to use and understand. The distinct organizational structures surrounding figure skating and classical music, particularly in relation to competition structure, centralization, and governance, have played key roles in how formal and standardized their rules have been. Based on Porter's work, as well as the data I have collected surrounding both settings, these factors, along with embeddedness within the sport world versus the arts world, have emerged as playing key roles in why they have adopted different rules. After a discussion of methods and data, I detail how evaluation works and how it has changed over time in these settings, addressing how these factors have contributed to variations in evaluation practices in skating and music.

Methods and data

The issues I am examining in this article revolve around differences in evaluation practices, which require comparing different evaluation practices. Fourcade (2011: 1725) emphasizes that comparative analysis "affords us precious analytical leverage … and reveals patterns that are not visible otherwise." Comparisons are especially crucial in analyzing distinctions between or among types of systems, as I am analyzing here. The variations I focus on are degrees of formality and standardization. To investigate how evaluative rules and practices develop through a comparison of two contexts, I rely on a mixed-methods approach, drawing on 96 semi-structured interviews with figure skating and classical music insiders, participant-observation, archival materials, and content analysis. These different types of data are useful for addressing evaluative practice development from different angles and allow for triangulation, which helps verify evidence (Jick 1979). Although looking at additional cases could have increased the breadth of my analysis, it would be virtually impossible to gain a full understanding of evaluation practices and how they work within particular settings without conducting in-depth case studies. Other scholars who have studied evaluation, as well as the effects of measures and other types of systems on organizations, have used similar approaches and types of data (e.g., Timmermans and Berg 2003; Stevens 2007; Lamont 2009; Sauder and Espeland 2009).

Figure skating and classical music provide an ideal comparison for addressing differences in evaluation practices. Skating is an extreme case in two ways: its evaluation system is highly formal and standardized, with very specific numerical rules, and it has overhauled its judging system in recent years, largely in response to a legitimacy crisis, which has completely changed many other aspects of the sport.

Many of the music competitions I focus on represent the opposite extreme, with much more informal evaluative rules, but music competitions use evaluation practices that are all over the map. This variation is very different compared to the standardized evaluation in figure skating, despite how the goal in both settings is the same: figuring out how to rank participants and ultimately who should win. Given that their goal is the same, why is evaluation in these settings so different, and what do those differences look like? These are the main issues I address here, focusing especially on the latter question.

For the figure skating component of my research, I interviewed 33 skating insiders in the United States, most of whom I recruited from an official list of about 200 active technical panelists. My sample includes four Olympic-level, four world-level, two international-level, and nine national-level judging officials. My interview request response rate was about 15%, but considering the range of experiences among the skating insiders I did interview and the other data I collected, I do not believe that this relatively low response rate affected my findings. I conducted these interviews over the telephone, mostly in July and August 2006 but also in July and August 2010. Both rounds of interviews took place just a few months after the Olympics, held in February 2006 and 2010, respectively, so the most recent Olympics was fresh in respondents' minds for both rounds. Interviews averaged about 45 minutes long but lasted between 22 minutes and almost two hours, and with interviewees' permission, I recorded and later transcribed them. We covered several broad topics, including the 2002 Olympic judging scandal and ensuing responses, the two judging systems and especially their differences, how skaters and their choreographers construct their competitive programs, and figure skating's relationship with outsiders like skating fans.

I also collected data as a participant–observer in several settings, including two judging seminars for the overhauled judging system – known as the "international judging system" (IJS) – one in August 2006 that took place at a regional-level competition where some participants did trial judging, and a national technical panel training seminar in April 2009. This data is especially useful for showing how judging officials learn how to judge under the IJS and how they interpret the IJS, and it shows how formal the IJS is and its emphasis on numbers and following written rules. In addition, I collected a variety of archival data spanning from 1997–2023, including television coverage and web-based material from a number of sources. The "communications" published by the International Skating Union (ISU) were especially useful because they provide an official record of the ISU's response to the 2002 Olympic scandal and how the judging changes emerged, and several official documents include guidelines for identifying technical elements and their levels of difficulty, as well as determining technical and artistic scores.

I collected similar types of data for the classical music side of this study. I interviewed 63 music insiders, including 21 music students at the college level or above (many of whom had at least some teaching experience), 40 musicians who had taught at a university-level music school or conservatory (or several) in the United States, and two other professional musicians. Everyone I interviewed was an active musician at the time, and almost everyone had been involved in activities that revolve around evaluating musical performance on a regular basis. These interviews were conducted in person between August 2009 and May 2010 and averaged about one hour in length. Other than one outlier that took only 17 minutes, interviews with professional musicians ranged from 33 minutes to two hours and 20 minutes, averaging about one hour and 15 minutes. Interviews with students were generally a bit shorter, ranging from 23 to 69 minutes and averaging about 45 minutes. We discussed several general topics, including evaluation criteria that music insiders tend to care about, what makes a "great" performance, and evaluation at music competitions.

In addition, I collected data as a participant–observer in a number of settings. The most significant component of this data includes participant-observation at six music competitions, four where I observed judges' deliberations and other meetings: a trumpet competition over two years (March 2009 and 2010), two rounds of a national-level organ competition held in May and July 2010, a conservatory concerto competition held in April 2010, and a conservatory piano competition over two years (April 2009 and 2010). This data illuminates how judges negotiate in a deliberative context and ultimately reach a consensus, as well as how deliberation can affect competition outcomes. It also shows the informality and discretion involved in evaluation in music. Finally, I collected archival data from a variety of sources. This data includes information about rules and judging procedures from many competitions' and related organizations' websites. To supplement this data, I corresponded with staff from several of these competitions about how they are judged and how their evaluation practices have evolved over time. Additional sources included several music schools' and conservatories' websites, music blogs, and other miscellaneous materials. This archival data reveals the wide range of evaluation practices surrounding music competitions, the general informality surrounding music compared to skating, and whether and how particular music competitions' evaluation practices have changed over time.

Data analysis involved mainly open coding and content analysis of interview transcripts, field notes, and archival materials. Mostly following an inductive, grounded theory approach as outlined by Charmaz (2001), I used my research questions and my prior knowledge of figure skating and classical music as points of departure

for developing interview questions and analyzing data. This is also where my initial codes – including "judging," "judging changes," "competitions," "deliberations," "discretion," "evaluating evaluators," and "repertoire" – came from. As I collected and analyzed more and more data, I added many codes that emerged from the data, such as "numbers," "formal/informal," "conformity," "freedom/creativity," and "legitimacy." This two-phase coding process is similar to how several experts in qualitative research methods and data analysis describe how this process often works for qualitative researchers (for example, see Lofland and Lofland 1994; Emerson et al. 1995). I now turn to a discussion of factors influencing the formality of evaluation practices, revolving around organizational structure and focusing on how rules and practices have developed in figure skating and classical music.

Organizational context: Sports vs. the arts

Although figure skating and classical music have different evaluation practices, they are similar on other dimensions. Perhaps most obviously, both contexts have technical and artistic components, which competition evaluators must take into account as they are judging. Given the central role of aesthetic principles in art worlds (Becker 1982) and how often members of these settings need to evaluate quality, people in the arts tend to grapple with evaluation more explicitly than in many other social contexts. This makes the arts in general, and figure skating and classical music competitions in particular, especially useful for studying social processes surrounding evaluation.

Many people describe artistic performance as "elusive" or "ineffable," which contributes to tensions in evaluating performances with artistic features. A lot of experts believe very strongly that using formal criteria to break down and evaluate artistic products reduces them, but competition judges are forced to do this all the time. However, while some settings (like figure skating) have been required to formally codify artistic components in their judging processes, others (like classical music) have not. Although skating is situated within the art world, its position as a sport has had a profound impact on its evaluation practices and its need to rely on formal rules, whereas the music world as a whole has not needed to deal with similar pressures to adopt formal evaluation practices based on its more obvious position within the arts world.

In comparing figure skating and classical music, I focus on competitions in both domains rather than other evaluative settings because the people involved in evaluating competitions must make excellence explicit. The need to identify one clear "winner," and "losers" by extension, requires competitions to develop clear ways to separate the winner from other competitors. This is different from

other evaluative contexts in these arenas, such as skills tests in skating and auditions in music, where in many cases multiple participants can "win." When judges can pick multiple winners rather than just one, they can choose to use a broader definition of excellence and reward participants with a wider variety of strengths rather than one participant who fits a narrower set of criteria. Competition evaluators who must identify a single winner are usually forced to use a less flexible definition of excellence. This makes the evaluation practices surrounding these types of events especially significant.

There are a number of parallels between music competitions and competitive sports, including the complex organizational components that go into them, participants' hard work and preparation leading up to them, and the identification of a clear winner. Music competitions also involve technical aspects, such as sound quality, whether participants play or sing the right notes, and even how long a musician can sing or play without taking a breath, which are comparable to the athletic aspects of figure skating and other competitive sports. Despite these similarities, few people would consider music a "sport," and there are several unique characteristics that distinguish sports from most other areas of social life.

First, competition – where the goal is to win – plays a central role in sport unlike in any other social setting. While competition also plays a role in many other areas, such as education (for instance, being named valedictorian or getting into the "best" college), the workplace (job offers and promotions), politics (winning an election), and even the family (sibling rivalries or being the "best" mother), its role is much more explicit in sport. In fact, for many sports fans and participants, one of its main attractions is that "success" is measured more exactly than in other contexts: by doing well in competition (Chambliss 1989). Sport also involves more clarity than other areas: by the end of any athletic event, the audience and participants should know who won, who lost, by how much, and how they won and lost (Eitzen 2012). Going along with this is "the perceived inherent purity and goodness of sport" (Coakley 2015: 11), the notion that sport encourages "fair play" (Eitzen 2012: 60) and is "played on a level playing field" (16). Despite much evidence to the contrary (Coakley 2015), the idea here is that the outcome of sporting events should be determined solely based on athletic factors like talent, skill, hard work, preparation, and strategy (and perhaps luck), rather than social factors like where participants are from, who their parents are, or who they know. This "obvious meritocratic orientation" (Washington and Karen 2001: 189) necessitates an emphasis on "playing by the rules" (Coakley 2001: 20), which is why many organizations within the sports world have incorporated bureaucratic practices, including strict rule systems that are in place partly to encourage objectivity (Eitzen 2012).

Beyond the importance of fairness and rules, sport is prevalent as a cultural, political, and economic symbol around the world, to such a degree that some major sporting events, like the Olympic Games, the World Cup, and the Super Bowl, to name a few, attract so much attention and media coverage that they are impossible to ignore, even for people who have no interest in sports otherwise (Coakley 2015). Especially with global sporting events like the Olympics, sport can contribute to nationalism and patriotic sentiment (Frey and Eitzen 1991) through "an emphasis on demonstrating superiority over other countries and other political systems" (Coakley 2001: 31). This is why sport has such strong political implications and plays such an important role in international relations (Guttmann 2003). The importance of the gold medal count in the Olympics is just one example of how success in sport is "interpreted internally and externally as 'proof' of the superiority of a nation's social, economic, and political systems" (Frey and Eitzen 1991: 512). Music also tends to play a cultural, political, and economic role within specific cultures and more globally, but the publicity surrounding even major international music competitions is typically nothing compared to global sporting events, and very few non-experts care about music competitions compared to sporting events.

While figure skating and classical music are very different in some ways, these two cases – one with highly formal, numeric, standardized evaluation, and the other with significant variations in its evaluation practices – provide similarities and differences that are useful for addressing why settings develop different evaluation practices. Figure skating's high profile among Olympic sports and relatively recent judging changes make it a unique case where it is possible to examine a transition between two very different judging systems. Given the similarities between the Olympics and music competitions (McCormick 2009), classical music offers an intriguing comparison. These two cases are different enough on the formal–informal continuum and in relation to how they are organized, which are major foci of this article, but they are similar in other ways. The publicity and visibility surrounding competitions more broadly, especially at the highest levels, means that they must at least appear legitimate to both participants and outsiders. This legitimacy is commonly, but not always, maintained through formal rules. Organizations involved in running competitions need to make sure that participants and outsiders see them as fair and that the evaluators involved can defend their decisions. All of this makes skating and music competitions a useful comparison for examining evaluation. I now turn to a discussion of other factors – competition structure, degree of centralization, and governance structure, all of which directly come out of organizational differences between these two contexts.

Centralization, governance, and competition structure

Despite the prominent role of international competitions in both figure skating and classical music, these contexts are organized very differently. Figure skating is highly centralized and has an international governing body, the ISU, which also governs speed skating, whereas the music world lacks both of these elements. Based in Lausanne, Switzerland, the ISU was founded in 1892 following the emergence of international competitions in figure skating and speed skating during the second half of the nineteenth century. The ISU's main purpose has been to develop and maintain international standards for these two sports and to organize international competitions (ISU 2023a). In figure skating in particular, the ISU was responsible for overhauling the judging system after the 2002 Olympic judging scandal and is in charge of reviewing and updating the IJS on a regular basis. For individual athletes to compete at events that are linked to the ISU, like the World Championships and the Olympic Games, their countries must be ISU members. As of June 2023, the ISU has 101 member federations representing 80 countries (Australia, Belgium, Canada, Finland, Germany, the United States, and a number of other countries have two organizations, one for figure skating and one for speed skating) (ISU 2023b). Skating's status as an Olympic sport means that the ISU is accountable to a more powerful organization, the IOC, which has the highest authority over anything having to do with the Olympic Games (IOC 2021).

One thing that makes the ISU, and the IOC by extension, so significant in figure skating is its competition structure, where competitions build on each other. Winning an Olympic gold medal is the (rather obvious) pinnacle of the sport, which skating commentators tend to bring up over and over again leading up to and during the Olympics. The following statement by commentator and Olympic gold medalist Scott Hamilton at the beginning of the 2018 Olympics personifies how much of an emphasis skaters place on the Olympics: "It's the BEST!!! … This is the biggest stage these athletes will ever stand on … When they leave here their lives are forever changed" (Sheehan and Michaels 2018).

Especially in countries with skaters who are competitive on the world level, to have any hope of making an Olympic team, skaters must do well in other competitions first. Each year in the United States, for example, the National Championships (which most skaters must qualify for through regional and sectional competitions earlier in the season) acts as the qualifying competition for many competitions that follow, including the World Championships and the Olympics but also the six-event Grand Prix Series that takes place internationally each autumn and other competitions that receive less publicity but can be very important for skaters to establish themselves. Even before those

events, the number of spots that each ISU member nation has at competitions like the World Championships and Olympics depends on their skaters' placements at previous competitions. For instance, in the year before an Olympics, for a member nation to be granted three spots to those Olympics and the following year's World Championships, the combined placement of their top two skaters at that year's World Championships must be no higher than 13. If two Japanese skaters were placed sixth and seventh in the women's competition, Japan would receive three spots for the following year, but if that second skater were placed eighth, they would only receive two spots. Before skaters reach that level, to put their names in the mix for consideration for bigger events, they must do well at other less prestigious competitions.

Especially with this kind of competition structure, it makes sense that the ISU requires a standardized set of rules at all ISU-sanctioned competitions and that the vast majority of other competitions around the world also uses those rules so that athletes competing there are better prepared for ensuing competitions. This competition structure is very different compared to music competitions, which rarely build on each other. Some competitions are limited to a particular instrument, age range (classical music competitions tend to be limited to young musicians at the beginning of their careers), or repertoire, or have other entrance criteria, but as long as musicians meet those requirements, they can enter those competitions. Even for many prestigious international music competitions, any eligible musician can submit an application, usually consisting of a recorded performance.

In the music world, the most similar organization to the ISU is the World Federation of International Music Competitions (WFIMC), a network of many of "the world's most important music competitions," with 125 member competitions as of June 2023 (WFIMC 2021). Founded in 1957 and based in Geneva, Switzerland, one of the WFIMC's chief objectives is helping and advising member competitions, including by helping them communicate with each other and with other organizations, promoting their prizewinners' careers, and presenting international music competitions in a positive light more generally. This organization also requires "that member competitions maintain the highest professional standards and strictest ethics" (WFIMC 2022).

Music competitions can also become members of the Alink-Argerich Foundation (AAF). Founded in 1999, the AAF provides "the most complete details on music competitions ever compiled" (AAF 2023), with particular emphasis on piano competitions, and works to help and provide information to musicians and competition organizers. While the WFIMC is only open to international competitions, any musician or competition that is willing to make an annual contribution is eligible to join the AAF. In addition to providing "dedicated

assistance" to all members, one of the AAF's main goals is to provide member competitions with publicity through its website and inclusion in its annual brochure. As of June 2023, the AAF has 208 member competitions, some of which are also WFIMC members (AAF 2023).

Even though the WFIMC and AAF play significant roles at the international level, neither organization acts as an international governing body for classical music. As long as music competitions do not build on each other, the music world does not have much need for a governing body that would oversee international competitions and enforce standardized rules. In addition, although many international music competitions have applied for WFIMC membership and have paid to join the AAF, hundreds of successful international music competitions are not members of either organization. Without an organization to impose a single set of rules on all competitions, music competitions can use whatever evaluation practices they deem most appropriate. As many of the music insiders I interviewed pointed out, specific competitions' evaluation methods usually depend on the organizations that run them and sometimes even differ from one year to the next. One piano teacher noted, "The guidelines really can be quite different. Sometimes it's just numbers, or you throw out the high and the low, or 'yes, yes, yes, or no,' or I mean it's just, every competition organizer has to make those decisions as to how they will evaluate it" (Pianist and music conservatory faculty member). The music world's lack of centralization and international governance and its disconnected competition structure encourage a wide range of evaluation practices, whereas figure skating's centralization, international governance, and integrated competition structure more or less require shared evaluation practices.

Evaluation practices in figure skating vs. classical music

In the following sections, I provide a detailed account of how evaluation works in figure skating compared to classical music. Building on work on rules (Meyer and Rowan 1977; Dobbin and Kelly 2007; Edelman et al. 2011; Edelman 2016) and evaluative cultures (Lamont 2009, 2012), as well as Porter's (1995) discussion of objectivity, in addition to highlighting the evaluation practices that emerge in different contexts, especially shared versus varied practices, this article enhances our understanding of how the quest for formality and objectivity on the one hand and the use of deliberation and the discretion that tends to go along with it on the other hand affect evaluation practices on the ground. This matters because in a deliberative context, judges often influence each other's decisions about competitors' merits and how to evaluate them, which can change outcomes, whereas non-deliberative evaluation prevents judges from

influencing each other's opinions, at least in theory. The differences discussed here between figure skating and classical music also enhance our understanding of how and why evaluative cultures change.

Figure skating: Shared, rigid evaluation

While skating has been around as a mode of transportation for more than 3,000 years, competitive figure skating did not begin until around the mid-1800s (Hines 2006). Informal international competitions cropped up during the latter half of the century, and the first official International Figure Skating competition was held in Vienna in 1882 (ISU 2012). Skating styles were very different in different parts of the world around this time, with an emphasis on movement across the ice in continental Europe, geometric formations in England, and developing unique designs in North America. These differences created problems for judging international competitions fairly, which led to the need for international judging criteria (Hines 2006).

Following the rise of national skating organizations and international competitions, most prominently the first European Championships in 1891, the ISU was established in 1892 to develop international standards, which involved many compromises but most closely resembled the "international style" of continental Europe. After the first World Championships in St. Petersburg in 1896, in 1908, figure skating was the first winter sport included in the Olympic Games (ISU 2012; Hines 2006, 2011). In current international competitions, skaters in each discipline (men, women, pairs, and ice dance) skate a short program and a long program (also known as a "free skate") (Hines 2006). The ISU first introduced formal rules for figure skating competitions around 1895 with the 6.0 judging system (Hines 2011) and kept using this system, with occasional adjustments over time, until 2005, a few years after a very consequential judging scandal.

The 2002 Olympic judging scandal

Many people inside and outside of skating would identify the judging scandal during the 2002 Olympic Games in Salt Lake City, which preceded the ISU's total overhaul of its judging system, as the most significant scandal in skating history. The top two teams in the 2002 pairs event were from Canada and Russia. Despite a technically flawless long program by the Canadians, the Russians won the gold medal, even with a noticeable mistake. After the Canadians' marks were unveiled, indicating a second-place finish, NBC commentator Scott Hamilton exclaimed, "How did that happen?! … They won that

program! … There's not a doubt in anyone in the place, except for maybe a few judges … That will be debated forever … Debated forever" (Gunts and Michaels 2002). Five of the nine judges involved placed the Russians ahead of the Canadians in the free skate, but the only one who received widespread publicity was Marie-Reine LeGougne of France, who was involved in a vote-swapping deal with Russia. After the medals had been awarded, LeGougne acknowledged that the French Ice Sports Federation had pressured her to favor the Russians so that the Russian ice dance judge would favor the top French ice dancers and essentially assure them a gold medal.

In response to the uproar over this result, the next day, the ISU began an investigation of the event's judging. The day after that, IOC President Jacques Rogge met with ISU President Ottavio Cinquanta "to emphasize that the situation needed to be resolved quickly" (Roberts 2002a). During a "late-night emergency meeting" four days after the event, the ISU decided to throw out the French judge's marks for the Russians, which led to a four-to-four tie and a duplicate gold medal for the Canadians. Only days later, while the scandal was still a big story, the ISU "unveiled plans for a new scoring system designed to limit the threat of collusion between judges." In an attempt to divert people's attention from the scandal, Cinquanta stated, "'This is a total revolution in the history of the International Skating Union … I promise this system will reduce to a minimum the prospect of bloc judging'" (Roberts 2002b). This scandal was so significant in figure skating history partly because "no judge had ever admitted to outright cheating" before (Jackson 2005: 198). There had been many previous scandals involving disputed outcomes at major events like the World Championships and the Olympics, but unlike this 2002 scandal, none of the judges involved in those competitions ever admitted that they had done anything wrong.

Given the ISU's relationship with a more powerful governing body, the IOC, figure skating insiders must answer to a higher power. The power dynamics between these two organizations were critical in shaping the judging changes that followed the 2002 Olympic scandal. This scandal emerged as a central storyline of those Olympics, and in response, the IOC insisted that the ISU take action to restore its public image. This IOC pressure played a fundamental role in accelerating the transition from the 6.0 system to the IJS. An Olympic-level judging official who was involved in designing the IJS explained how this process unfolded:

> It was the International Olympic Committee that went to the ISU and said, "Either you come up with a better way of measuring your sport, or you're out of the Olympic Games." So it was a mandate by the IOC to the International Skating Union to improve the way they measure their athletes. For example, in speed skating you're measured by time. In figure skating in

> the 6.0 system, you were measured by the impression of someone else, so the IOC said, "That's not acceptable sport. We need a measuring stick" … So the ISU came up with this system of measurement that was acceptable to the IOC and kept the sport in the Olympic Games.
> (Olympic-level judging official and ISU Technical Committee member)

This IOC mandate and the ISU's response to it highlight how influential figure skating's centralization and shared governance have been in shaping its evaluation practices. If the ISU had been self-governing and had not had to worry about satisfying a more powerful organization like the IOC, implementing these judging changes would have probably been a much slower, more deliberate process, and the ISU would have at least had the option of just keeping the old 6.0 system without any changes. The ISU had already been thinking about changing its judging system, and the 2002 Olympic scandal, especially the French judge's admission, gave its leaders a perfect opportunity to go ahead with those changes. In keeping with the link between sport and notions of "fair play," the International Sports Federations that belong to the Olympic Movement, such as the ISU, must ensure that their sports are fair on an international level, including in relation to how competition outcomes are determined. The IOC's commitment to encouraging ethics across all Olympic sports helps explain why it put so much pressure on the ISU to do something to uphold its integrity after the French judge's confession became public.

The judging changes

After this scandal, the ISU completely overhauled its judging system to make it more formal and more numerically specific. The new "IJS" was officially unveiled in late 2003 and gradually implemented leading up to the 2006 Olympics. I outline the main differences between the 6.0 system and the IJS in Table 1.

	6.0 System	IJS*
Judging Panels	One panel	Two panels: one technical panel and one judging panel
Judging Tasks	All judges identified and evaluated technical elements and artistic presentation	Technical panelists: identify skaters' technical elements and their levels of difficulty Judges: evaluate skaters' technical elements and program components
Anonymity	Judges' nationalities displayed	Judging officials' nationalities hidden
Technical Evaluation	0-6.0 scale for overall technical merit	Pre-determined base values for all technical elements; judges evaluate elements with grade of execution (GOE) scores ranging from -3 to +3**
Deductions	Automatic deductions for specific errors on required elements (short program only)	Deductions for specific errors
Artistic Evaluation	0-6.0 scale for overall artistic presentation	0.25-10.0 scale for five program components: skating skills, transitions, choreography, interpretation, and performance/ execution***
Competition Outcomes	Ordinal rankings combined after short and long programs	Cumulative point totals after short and long programs

*The information here reflects the IJS as it was during much of my data collection and analysis process. Some IJS rules have changed over time, as explained here.

**Since the beginning of the 2018-19 season, judges have been evaluating technical elements with GOE scores ranging from -5 to +5.

***Following the 2021-22 season, the ISU decided to reduce the number of program components from five to three: composition, presentation, and skating skills. The ISU Technical Committees initiated these changes after observing that there were too many criteria at work for the original five components and that those criteria were often unclear. According to the ISU, those criteria "have been simplified and reorganized in a more logical way" to encourage "more objective evaluation" (International Skating Union 2022b).

Table 1: Figure Skating's 6.0 System vs. IJS*
Source: Author's work

The most significant difference is that the rules are now much more detailed, with much more specific judging criteria, but the 6.0 system and the IJS are generally quite different.

Under the 6.0 system, judges' placements determined competition outcomes. Each judge gave skaters two marks for each program: one for "required elements" (in the short program) or "technical merit" (in the long program), based on how well skaters performed the elements in their programs, and one for "presentation" that reflected skaters' overall programs, including composition, style, originality, and musical interpretation. Both marks were combined to determine each skater's total score from each judge, which led to skaters' "ordinals" from each judge. Skaters' placements for a particular phase of a competition were based on comparing ordinals, with the skater who received the most ordinals matching a given placement finishing in that spot. Because the long program was supposed to be worth twice as much as the short program, to determine overall competition results, short program placements were multiplied by 0.5, and free skate placements were multiplied by 1.0.

The IJS is generally much more exact than the 6.0 system, with each facet of skaters' programs broken down and evaluated based on precise guidelines. In contrast to the 6.0 system's ordinal rankings, the IJS is based on cumulative points. Another important distinction is that rather than relying on just one judging panel, the IJS uses two panels with distinct types of judging officials: (1) a three-person technical panel, typically made up of coaches or former skaters, which is responsible for identifying each technical element in a skater's program and its level of difficulty, and (2) a judging panel, usually with nine judges, which is in charge of evaluating each of those elements and several separate program components. Under the 6.0 system, judges needed to both identify elements and evaluate skaters' programs, but with the added technical panel under the IJS, judges can focus on evaluating elements and components. For the remainder of this article, I refer to judges and technical panelists as such and use "judging officials" or "officials" to refer to both groups simultaneously.

In Table 2, I describe how officials evaluate certain facets of skaters' programs under the 6.0 system versus the IJS.

Type of element	6.0 System	IJS*
Jumps **Triple Axel**	Judged as part of the overall required elements mark (short program) or technical merit mark (long program); if the skater made a mistake on this element in the short program, judges took a specified deduction according to what kind of mistake it was (this did not apply to mistakes in the long program); judges may or may not have considered this element and other jumps when determining presentation marks	Technical panel identifies "triple Axel" and determines whether or not the skater completed the required revolutions (if there is any doubt, the panel reviews the jump using video replay after the end of the program); judges evaluate the jump with a GOE score from -3 to +3; the average GOE score is factored and added to or subtracted from the jump's base value to determine the total score for this element**
Spins **Layback Spins**	Judged as part of the overall required elements mark (long program) or technical merit mark (long program); judges may or may not have considered quality of spins when determining presentation marks**	Technical panel identifies the level of difficulty depending on the number of "features" the skater achieves: judges evaluate the spin with a GOE score from -3 to +3; the average GOE score is factored and added to or subtracted from the spin's base value to determine the total score for this element
Footwork **Circular Step Sequence**	Judged as part of the overall required elements mark (short program) or technical merit mark (long program); judges may or may not have considered quality of footwork when determining presentation marks	Technical panel identifies the level of difficulty depending on the number of "features" the skater achieves; judges evaluate the step sequence with a GOE score from -3 to +3; the average GOE score is factored and added to or subtracted from the step sequence's base value to determine the total score for this element
Presentation/ Artistry* **Composition**	Judged as part of the overall presentation mark	Judges evaluate composition on a scale of 0.25 to 10.0 based on specific criteria

*Unless otherwise noted, the information in this table is from ISU Communication No. 1611 (International Skating Union 2010) and reflects the IJS as it was during much of my data collection and analysis process.

**Since the beginning of the 2018-19 season, judges have been evaluating technical elements with GOE scores ranging from -5 to +5.

***Guidelines for determining program component scores are provided in several ISU documents, including "Program Components Overview" (International Skating Union 2004) and "ISU FAQ – Program Components" (International Skating Union 2022b).

Table 2: Element Evaluation in Figure Skating: 6.0 System vs. IJS*
Source: Author's work

Under the IJS, judging officials determine skaters' technical scores as follows. All elements have "base values" that correspond with specific written criteria that the technical panel uses to identify them. Base values for jumps are based on the type of jump and the number of revolutions, with the "more difficult" jumps with more revolutions receiving higher base values.[1] Base values for other elements, such as spins and step sequences, are based on levels of difficulty, which revolve around how many "features" (such as position changes, "difficult" positions, or a particular number of revolutions) they involve. As the technical panel identifies each element and enters it into the computer system, the judges evaluate it with a "grade of execution" (GOE) score from -3 to +3 based on detailed guidelines.[2] Judges can give a GOE of –3 on a jump, for example, based on these criteria: "any of the following individual errors: fall; severe change of edge on take-off of flip or lutz; lesser rotation than required" (ISU 2010). The following guidelines, in contrast, are provided for giving a GOE of +3 on a jump: "Superior in all jump phases (eg. unexpected or difficult entry phase, great height/distance, strong flow in and out and superior extension on landing)" (ISU 2010). Judges must follow similar written criteria for all types of elements. To determine the score for a particular element, the GOEs from all judges for that element are averaged and then added to or subtracted from its base value. The total technical score for a given program is based on adding up all of the element scores from that program.

[1] Of the six types of jumps, the Axel is considered the most difficult (partly because it has an extra half revolution compared to the other jumps), followed by the Lutz, flip, loop, Salchow, and toe loop. The base values of these jumps vary accordingly, as the following examples illustrate:

Double toe loop 1.30 vs. double Lutz 2.10 vs. double Axel 3.30;
Triple toe loop 4.20 vs. triple Lutz 5.90 vs. triple Axel 8.00;
Quad toe loop 9.50 vs. quad Lutz 11.50 vs. quad Axel 12.50 (ISU 2022c).

[2] Since the beginning of the 2018–19 season, judges have been using a GOE scale from -5 to +5.

Under the 6.0 system, judges incorporated all technical elements (including jumps, spins, and step sequences) into skaters' technical marks, whereas each technical element is identified and evaluated separately based on strict guidelines under the IJS. According to the IJS rulebook, judges can award higher GOE marks if skaters show control, good speed, effortlessness, flow, and high-quality positions; well-centered spins; and elements that go along with the music. On the other hand, if skaters under-rotate, two-foot, step out of, or fall on a jump; take too long to prepare for a jump; fail to complete the required number of spin revolutions; or display poor positions, edge quality, or speed, judges are supposed to lower their GOE marks. Even for the best skaters in the world, judges rarely give the highest GOE mark of +3 (ISU 2010).

Similar to technical element judging, the artistic side of the sport is broken down much more specifically under the IJS than under the 6.0 system. Judges used to give skaters a single mark for presentation, which was designed to incorporate all artistic aspects of skaters' programs, often with an emphasis on choreography and interpretation. Some judges, at their discretion, included the quality of skaters' technical elements like spins, footwork, and even jumps in determining their presentation marks for a particular skater. Under the IJS, the artistic side of skaters' programs is judged on five separate "program components": skating skills, transitions, choreography, interpretation, and performance/execution. Judges evaluate each program on each one of these categories on a ten-point scale (in increments of 0.25), and this is supposed to measure a skater's overall technical mastery and presentation. Similar to the GOE marks, the IJS rulebook provides precise criteria for each component, based on the following qualities:

> Skating skills: balance, precision, flow, speed, effortlessness, and edge quality;
>
> Transitions: variety, difficulty, intricacy, and quality;
>
> Choreography: use of space, ice coverage, purpose, originality, and matching the music;
>
> Interpretation: musical expression and nuance;
>
> Performance/Execution: commitment, style, personality, carriage, movement clarity, variety, and projection (and how well a skater does technically).[3]

[3] Following the 2022 Olympic season, the ISU reorganized these five program components into just three: (1) Composition: "how the program is designed in relation to the music;" (2) Presentation: "how the program is performed," including skaters' "expressive abilities" and "musical sensitivity;" (3) Skating skills: technique and movement (ISU 2022b).

In addition to these factors, component marks should be based on the percentage of the program in which the skater meets those guidelines for each component. After the IJS was first introduced, even Olympic champions received many component marks below 8.0, but these marks have generally become much higher over time, with the top skaters at the 2022 Olympics receiving many component scores in the nine-range (ISU 2004). Skaters' total scores for each program are determined by adding their technical score and program component score together, and total competition scores are based on combining skaters' short program score and free skate score. The competition winner is simply the skater with the highest total competition score.

Despite how both the 6.0 system and IJS are generally based on "technical" and "artistic" marks, the IJS is much more rigid than the 6.0 system. Other than requiring judges to take specific deductions for certain types of mistakes on technical elements in the short program, judges' 6.0 system marks were not based on precise written criteria. Judges used the old technical and artistic marks as a tool to rank skaters and based those marks largely on their overall impressions of skaters' programs, which meant that judges had a lot of discretion in determining their marks. Especially following the 2002 Olympic judging scandal, the IJS was intended partly to make cheating and deal-making more difficult and to increase fairness surrounding competitions through stricter rules revolving around evaluating much more specific aspects of skaters' programs.

Judging the judges

Many of the skating insiders in my interview sample applauded the IJS for its relative transparency, including in relation to the training and testing process for judging officials, which has become much more formal since the judging changes. According to my data, assembling international judging panels under the 6.0 system did not involve many formal criteria. One of my informants who had a lot of experience with this on the ISU side explained, "Romania could have just sent someone and said, you know, 'This is our representative and they are qualified to judge at the Olympic Games,' and that's how it was done" (Olympic-level judging official and ISU Technical Committee Member). With the IJS, while each ISU member nation can decide how to train their own judging officials, before they can serve as judges or technical panelists for ISU events like the Olympics or World Championships, all officials must attend judging seminars and pass a series of tests.

To use the United States as an example, anyone who wants to become a judging official needs to pass a written exam just to qualify to attend a judging seminar that will count toward that process. There are different types of IJS judging seminars, some more focused than

others – on the technical panel (like one I attended as an "observer" in April 2009), a particular discipline like singles or ice dance, or certain elements like spins or program components, for instance – and most officials need to attend a number of seminars before they can pass all of the necessary tests and need to continue attending them to keep up with IJS rule changes and to advance to higher levels of officiating. These seminars are taught by experienced judging officials, who also evaluate exams at the end, and were described by some of the officials I interviewed as "unexpectedly grueling" and "very overwhelming" (Technical panelist).

During the technical panel seminar that I attended in April 2009, I heard a lot of discussion revolving around rules and rule changes, how to interpret certain rules, and how to call specific technical elements. The seminar faculty repeatedly emphasized the importance of knowing the rules and applying them correctly, whether or not judging officials agreed with them, and being able to justify their calls. For example, one faculty member said at one point, "If it's not in writing, don't say it," and another followed with, "You have to know the rules … 'it is' or 'it is not;' no 'it might be' … Your job is to apply the rules, not interpret the rules" (Field notes, April 17, 2009).

There was also a lot of discussion of how the tests at the end of the seminar would be evaluated, which gave me and everyone else who was there a great deal of insight into what these tests tend to look like and how they tend to be judged in general. To become a technical official or a judge, applicants must take a written test and act as an official for anywhere from four to six programs. During these "test" programs, technical officials need to call the elements in each program, and judges need to identify and justify GOE and program component scores. The seminar faculty highlighted several important criteria for evaluating these tests: identifying elements correctly and confidently, diction, the review process, pace of calling, focus (not getting hung up by incorrect calls), and teamwork, and they emphasized that fast calls are especially important because they affect everything else that happens during a competitive program, especially judges trying to enter GOE marks after a skater does each element (Field notes, April 17–19, 2009).

These exams are generally very difficult, which was emphasized by not only the faculty at this seminar but also several of my interviewees. One faculty member in particular, the ISU Technical Committee member quoted earlier, described these exams like this:

> [I]t's a very, very high-pressure type of environment, and we do that on purpose to make them nervous, which is what it really feels like when you're a judge or an official at an Olympic Games … If they say +2, they have to specifically identify in the rules and regulations, not by their feelings or, "Gee, I just think it was that good."
> (Olympic-level judging official and ISU Technical Committee member)

Another judging official at this seminar echoed these sentiments during an informal conversation I observed, where she told someone else who was there to take a test at the end that judging actual competitions is much more relaxing than calling programs during a test because there is so much pressure (Field notes, April 17, 2009).

The ISU trains hundreds of judging officials every year, but just a very small percentage of officials who go through the training process passes the exam. Since the IJS was introduced in 2003, the ISU has been collecting a lot of of data, including recordings of all technical panels for all performances at all ISU events. Especially compared to the relative lack of data from the 6.0 system, this comprehensive IJS data has made it much easier to distinguish among judging officials, identifying officials who are consistently quick and accurate and exposing officials who are not. When the ISU Technical Committees are deciding which judging officials to appoint to ISU events, they look at this data and select officials who have done consistently well on exams and at previous competitions, with a particular emphasis on accuracy and speed.

In keeping with all of this, figure skating has also implemented very strict rules for dealing with judging ethics, which I learned about firsthand before and during the technical panel seminar I attended. A few days before it started, all participants received an email asking us to read the "Technical Panel Code of Ethics" ahead of time because everyone would need to sign a compliance agreement during the on-site registration process. The code provides official ethics guidelines for technical officials involved in U.S. Figure Skating (USFSA) events, specifically by enforcing "the highest standards of ethics, fairness, honesty and integrity." One aspect of the code that the instructors at this judging seminar emphasized was that judging officials must "recognize that even the appearance of misconduct, impropriety, insincere attitude or purpose can be damaging." If officials fail to follow any of these rules, the USFSA can suspend their judging appointments or even expel them "from all U.S. Figure Skating events and activities." These guidelines have been designed in part to help participants, including athletes, and outsiders feel more confident that USFSA events will be conducted and judged fairly (USFSA 2009b).

The USFSA's preoccupation with maintaining high ethical standards illustrates how many organizations have become increasingly concerned about ethics over the past several decades

(Weaver et al. 1999; Erwin 2011). Many organizations have implemented corporate ethics programs, for instance, which often include establishing ethics committees and training programs, developing ethics codes and communication systems, and other features geared toward formalizing expectations for employees and promoting ethical organizational cultures (Weaver et al. 1999; Erwin 2011). In figure skating, the IOC's reaction to the 2002 Olympic scandal put pressure on the ISU and its member nations to make sure the sport's ethical conduct improved quickly. Requiring participants to sign an "Agreement to Comply with Code of Ethics" form (USFSA 2009a) before this judging seminar is one of the clearest examples of this from my data.

One result of the judging changes is that many coaches now serve as technical panelists. This means that coaches' and judging officials' positions sometimes overlap, which has complicated the ethical standards within the sport. The faculty at this judging seminar addressed this issue in detail, advising that coaches who have become certified technical panel officials must acknowledge when they are coaching a skater in an event. Whenever this is the case, they are not allowed to judge that event. As one of the instructors pointed out, this is an important rule because "the perception is that you have an inside track … Perception is reality" (Field notes, April 18, 2009). All of this exemplifies skating's highly formal evaluation practices.

Classical music: Varied, flexible evaluation

In contrast to figure skating's formal, standardized evaluation practices, evaluation in classical music is varied and generally much more flexible. While music contests reportedly go back to ancient times, the modern music competition did not become especially prominent until the nineteenth century, throughout which competitions became increasingly popular, especially in Europe. During the twentieth century, many international competitions were created, mostly focused on a particular instrument, especially piano or violin, and many to celebrate a famous musical figure. The International Tchaikovsky Competition, Chopin International Piano Competition, and Paganini International Violin Competition are three prominent examples of competitions commemorating renowned composers (Latham and Spencer 2002).

Table 3 compares evaluation in figure skating and classical music on several significant dimensions.

Figure Skating	Classical Music
Standardized evaluation practices	Varied evaluation practices
Formal, numeric rules	Generally much more informal rules
Technical and artistic components	Technical and artistic components

Table 3: Evaluation in Figure Skating vs. Classical Music Competitions
Source: Author's work

The most significant differences for the purposes of this article are that evaluation in skating is standardized and generally much more formal than in music. While some music competitions use numeric evaluation systems, many do not, and when they do rely on numbers, judges often give a single score (on a scale from zero to ten or zero to 100, for instance) to the performance as a whole, rather than scoring specific components of the performance like judging officials do in skating now.

In Table 4, I compare elements of three well-known international music competitions: the Queen Elisabeth International Music Competition, Chopin International Piano Competition, and Paganini International Violin Competition.

	Queen Elisabeth International Music Competition*	Chopin International Piano Competition**	Paganini International Violin Competition***
Frequency	Annual	Every five years	Annual until 2002 Biennial since 2002
Location	Brussels, Belgium	Warsaw, Poland	Genoa, Italy
Instrument(s)	Violin, Piano, Voice, Composition (competition for one or two instruments/ categories each year)	Piano	Violin
Age Range	18-29	17-30	16-30
Repertoire Guidelines	Guidelines but some choice	Guidelines but some choice; works by Chopin only	Guidelines but some choice
Number of Rounds	Four	Six	Four

Competition Format	*Preselection Round:* Confidential Judges evaluate applicants' DVDs to eliminate participants who do not play well enough for public rounds *Preliminary Round:* Public Unlimited number of participants Participants play their first two or three pieces, then the jury selects one or more etudes from their repertoire lists *Semi-Final Round:* Public 24 participants Recital phase: Participants play a piece written for that year's competition; jury selects recital programs from repertoire lists Concerto phase: Participants play a Mozart concerto *Final Round:* Public 12 participants Prizes awarded	*Screening Round:* Confidential Admissions Committee (10 members) evaluates applicants' documents and DVD recordings Participants choose repertoire order for all rounds *Preliminary Round:* Public No more than 160 participants 20-member jury 12-member Competition jury *Competition Stage I:* Public No more than 80 participants *Competition Stage II:* Public No more than 40 participants *Competition Stage III:* Public No more than 20 participants *Final Round:* Public No more than ten participants Prizes awarded	*Screening Round:* Confidential Judges evaluate applicants' documents and CD recordings Participants choose repertoire order for all rounds *Preliminary Round:* Public Unlimited number of participants *Semi-Final Round:* Public No more than 12 participants *Final Round:* Public No more than six participants Prizes awarded

Number of Judges	Varies from year to year	12	At least seven
Judging Procedures	Secret ballot system for all rounds *Preselection Round:* Yes/no system *Preliminary Round:* Numeric and yes/no system (0-100 points) *Semi-Final Round:* Numeric and yes/no system (50-100 points) *Final Round:* Numeric system (60-100 points) and rankings	*Pre-Final Rounds:* Numeric and yes/no system (1-100 points) Discussion and open ballot voting *Final Round:* Numeric system (1-100 points) Discussion and open ballot voting	*Preliminary and Semi-Final Rounds:* Voting system (yes/no/possible) *Final Round:* Numeric system (70-100 points)
Prizes	1st Prize: 25.000€; concert performances 2nd Prize: 20.000€; concert performances 3rd Prize: 17.000€; concert performances 4th Prize: 12.500€; concert performances 5th Prize: 10.000€; concert performances 6th Prize: 7.000€; concert performances Unranked Finalists: 4.000€; recital	1st Prize: 30.000€ and gold medal 2nd Prize: 25.000€ and silver medal 3rd Prize: 20.000€ and bronze medal 4th Prize: 15.000€ 5th Prize: 10.000€ 6th Prize: 7.000€ Remaining Finalists: 4.000€ each Special prizes (including concert performances, money, and gifts) are also offered independently of the prizes listed above.	1st Prize: 25.000€; concert engagements 2nd Prize: 10.000€ 3rd Prize: 5.000€ Other Finalists: 1.500€ each Additional special prizes

*This information is from the Queen Elisabeth International Music Competition's official website (https://concoursreineelisabeth.be/cgi?lg=en).

**This information is from the Chopin International Piano Competition's official website (https://chopin2020.pl/en).

***This information is from the Paganini International Violin Competition's official website (https://www.premiopaganini.it/) and the website for the 2010 competition (http://www.carlofelice.it/VediFocus.asp?imgPath1=Media&imgPath2=focus&parentZ=5020&itemID=5020&idNews=1247&idSpettacolo=0).

Table 4: A Comparison of International Music Competitions
Source: Author's work

This information illustrates the variation among competitions, including in terms of featured instruments, frequency, numbers of rounds, and how many judges they use. For the purposes of this article, one of the most important distinctions among these three competitions is their different judging practices.

The Queen Elisabeth Competition uses secret ballots across all four rounds, but its procedures for each round are somewhat different. During the first round, before the actual competition starts, judges use a yes/no system, where they simply specify whether they think participants should or should not be admitted to the competition. As the competition progresses, these procedures become increasingly more complicated: numeric systems with various scales added to the yes/no system for the middle two rounds and a numeric system and rankings for the final round. The Chopin and Paganini Competitions also use different procedures depending on the round. The Chopin Competition combines a numeric and yes/no system with discussion and open ballot voting, which allows judges to see how other judges vote, but it drops the yes/no system for the final round. For its first two rounds, the Paganini Competition uses a "yes/no/possible" voting system, but then it uses a 70- to 100-point numeric system in the final round. In contrast, it is not uncommon for competitions to use the same procedures for all rounds, and sometimes a particular competition uses different practices each time it is held depending on who is involved. As these examples highlight, competition judging practices can be wildly different.

While some music competitions have developed more formal evaluation practices over time, many competitions rely heavily on deliberation. This resembles the multidisciplinary peer review panels outlined by Lamont (2009), who shows how participants use their deliberations partly to develop shared evaluation guidelines, which not only contribute to building trust but also leave room for uncertainty and discretion to address it and more complex cases. Music competitions with deliberative evaluation tend to work this way too,

with rules and criteria developing depending on the judges and along with their deliberations.

I observed judges' discussions at several competitions like this, including an annual trumpet competition over two years. During the semi-final rounds of this competition, during breaks between each group of six or seven musicians, the judges typically talked about which performances stood out to them from that group. Then, at the end of the round, they listed the competitors that they thought should advance to the finals – individually – and the competition's judging coordinator and the head judge for that division tallied their lists. This silent ballot process set the starting point for their deliberations by giving the judges a sense of where they agreed and disagreed, with the ensuing discussion focused on performances they disagreed on. At the end of their deliberations, they nominated between three and six musicians for the final round in each division based on where they decided that there was a "natural break" between competitors, but not in any particular order.

Discussions at some competitions relying on deliberation as part of their judging processes work somewhat differently. Judges might start by talking about each performance in detail, noting positive and negative aspects of each performance, and then use that discussion to figure out the results or who should move on to the next round. However, judges often have limited time for their deliberations, so this likely tends to happen more at competitions or after rounds with fewer performances (I observed two discussions where this happened, and both were after rounds with only three or four performances and took 25–30 minutes). More broadly, there is wide variation in how long judges take to discuss competition performances. I observed deliberations ranging from about two minutes to about 30 minutes, but some of the musicians I interviewed reported that they had been involved in longer discussions. Not surprisingly, deliberation length tends to correspond to how much consensus there is when the discussion starts, with longer discussions typically happening when there is a lot of disagreement.

If competitions use numeric or yes/no judging systems, judges typically write a score or "yes" or "no" next to each musician on a list of names (or a list of numbers if judges are not supposed to see performers' names until after they finalize the results), and then they hand in their lists to whomever is responsible for tallying the results. Sometimes judges deliberate after results are tallied, in which case those results might change, but this often does not happen.

Music competition controversies

Classical music competitions have also had to deal with judging-related scandals, perhaps most notably at the International

Tchaikovsky Competition. In the violin portion of the 1974 competition, for example, judges did not award a gold medal because they "decided no one had performed consistently well enough." This result was very controversial, particularly because a North Vietnamese judge reportedly gave American Eugene Fodor, who tied for the silver medal, only five points out of the maximum 25. After this became public, Fodor was described as "the victim of political bias" (Osnos 1974). 20 years later, in 1994, judges failed to award a first prize in all three of the competition's "main categories" (piano, violin, and cello). Observers described this as "the most stunning outcome in the competition's tumultuous history since 1966, when audience members shouted 'Shame!' at Soviet jurors for unjustly favoring the Soviet contestant." Judges tried to defend these outcomes by claiming that none of the competitors had lived up to the standards of past winners, but several disgruntled violin judges boycotted the awards ceremony, and "people gasped and some members of the audience booed" when the results were announced (Stanley 1994).

For decades, the Tchaikovsky Competition, held every four years, was one of the world's most prestigious piano competitions. Some people still think of it as "one of the major events in the international music community" (Peterson 2011), but several issues, including a series of scandals like those described above, have tarnished its reputation (Stults 2010). In an attempt to deal with these issues and restore its reputation, competition organizers made major changes leading up to the 2011 contest. One of the most significant changes surrounded selecting judges. In the past, this competition's judging panels had included "a fair number of teachers, including Moscow Conservatory faculty members notorious for lobbying and voting in favor of their own pupils," which had led to public scrutiny (Stults 2010). For the 2011 competition, organizing committee chair Valery Gergiev (a conductor) tried to prevent this criticism by selecting mostly performing musicians for the judging panels. Gergiev also asked Richard Rodzinski to serve as senior advisor to the 2011 competition. Rodzinski was the Van Cliburn International Piano Competition's executive director for 23 years, and he introduced new rules for the Tchaikovsky Competition that were very similar to the Cliburn Competition's rules (Service 2011).

One of the cornerstones of these new rules was a new evaluation system that was originally developed in 1990 for the Cliburn Competition and other competitions in the United States to address problems in how many music competitions had been judged. Described by Rodzinski as "transparent" and "very sophisticated in its construction" (Stults 2010), rather than relying on judges' raw scores, this new system emphasizes the spacing between judges' scores. As a result of these changes, this system "forces each juror to have an equal impact on the final result … minimizes the impact of 'gaming' by a

juror … permits abstentions while minimizing any resulting bias … [and] removes the effect of strong juror personalities which could impact the results from traditional consensus systems" (International Tchaikovsky Competition 2010). Even with these changes, which were designed to prevent the typical scandals surrounding this competition, controversy tainted the 2011 contest once again, partly because of some widely debated judges' decisions (Morrison 2011).

The Tchaikovsky Competition is an example of a music competition that has at least tried to increase fairness and transparency by making its evaluation practices more formal and mechanical. Some competitions, however, have taken very different approaches, including even changing evaluation procedures in the opposite direction. The Sydney International Piano Competition, for instance, used a numeric evaluation system for its inaugural contest in 1977 but implemented a yes/no system leading up to its 1981 competition (Email correspondence with the Sydney Competition, September 14, 2010; information provided by Warren Thomson, Artistic Director). Despite how this competition has also been surrounded by judging controversies that have led to widespread criticism, its judging process has not changed much in recent years. Porter (1995) would describe this as anomalous based on his discussion of quantification and objectivity, according to which publicity and outside pressure tend to contribute to more reliance on numeric systems.

Many prominent classical music competitions have dealt with judging controversies, but to my knowledge, no one involved has ever admitted to dishonest judging or anything else that might be considered unethical or contribute to legitimacy concerns. The people involved in evaluating music competitions also typically have a great deal of power, especially compared to judging officials in figure skating. One woodwind player and teacher explained that "the people who are usually doing the evaluations are … the ones that are revered" (Woodwind player and music conservatory faculty member). Combined with how music competitions do not need to appease an overarching governing organization that could force them to change their rules, based on my data, this power has played a key role in minimizing pressures on music competitions to mechanize their evaluation procedures. More broadly, the different organizational structures surrounding music and figure skating, with its centralization and international governance, promote different types of evaluation practices.

Judging the judges

The lack of centralization and international governance in music has also had a major impact on how musicians become adjudicators. While figure skating insiders need to go through extensive training and

testing under the IJS, based on my data, musicians simply rely on their experience – their own teaching and evaluating their students, observing other people like their teachers or colleagues evaluating, and listening to performances. One musician explained, "Some of it is very intuitive. You just get an intuitive feeling about a player when you hear them. When you've lived long enough and you've listened long enough, I think you develop that sense of 'this is good'" (Woodwind player and music conservatory faculty member). Another music insider emphasized that musicians don't usually receive any formal training to help them become evaluators: "I never took a course in that … I'm not sure anybody gives one … It's just applying what you know about music" (Vocalist and music conservatory faculty member). This informality and lack of standardization makes sense given the competition structure in music, where competitions usually do not build on each other like they often do in figure skating and where it is not considered important for competitions to use the same rules.

Also, unlike figure skating, the classical music world does not have a formal mechanism for evaluating judges, which makes sense given its lack of centralization and international governance and music evaluators' resulting high degree of power. However, music judges are evaluated more informally all the time. This happens mostly when competition organizers are assembling judging panels before competitions, where musicians' informal peer evaluations tend to play a major role. Competition organizers often call musician friends and ask for judge recommendations, or if organizers are considering inviting specific musicians to serve as judges, they might ask their friends about whether those people would make good judges. Rather than being asked to take a "judging test" before a competition, musicians who are invited to judge competitions are assumed to already have whatever knowledge they need, as one music student described: "It's not like there is a training session beforehand: 'this is what it's supposed to sound like.' You're expected to know what sounds good and what doesn't and all those in-between areas" (Tuba student and adjudicator).

Despite how music competitions do not generally rely on formal procedures for assembling their judging panels, many competitions use strategies to try to ensure as much fairness as possible. The most common examples that came up in my interview data were trying to prevent teachers from evaluating their own students and throwing out the high and low scores in cases involving numeric evaluation systems, which tends to be helpful because judges sometimes give outrageously high or low scores if they want a certain player to win. As most of the musicians I interviewed admitted, however, it would be virtually impossible to "police" the judging at music competitions without interfering with the judging process itself. Regardless of the results,

judges would typically be able to justify their evaluations somehow, as a guitar player emphasized:

> I'm sure there's lots of judging competitions where people really want their students or their style or their clique to win. I am sure that happens all the time. I don't see what you're supposed to do about that. I mean once you hire them, they're gonna say what they like or don't like. I mean, you could say, "Oh, well that was clearly driven by this," but they can always say, "No, I just liked the way they sung." So, that's a really hard one. I don't see how really you could do that.
> (Guitar player and music conservatory faculty member)

Without any formal policing mechanisms, musicians stressed how important it is to find ethical judges with integrity and "good faith" but acknowledged that they cannot always predict whether or not someone will be trustworthy.

These trends highlight the importance of trust within these competition settings. The four music competitions that allowed me to observe judges' deliberations are very clear examples of this trust. For instance, based on the data I collected at the trumpet competition I attended over two years, including informal conversations with board members and judges, this competition revolves around friendship, trust, and camaraderie. One especially illustrative example of this trust, especially among the people involved in the judging process, happened before a semi-final round in 2010. A judge whose student was about to play brought up the issue of whether he should judge his own student. He asked, "Does anyone else have a student playing?" None of the other judges responded, so he continued, "I do, so I'll just excuse myself." Another judge acknowledged that there were three students from his school in this division but that they were not his students. This conversation led to some confusion about the competition's rules surrounding this issue. After the judge whose student was going to play suggested, "I'll take myself out," and joked that he would not be able to be objective, the head judge looked for a board member who would be able to make a recommendation based on the competition's policies. By the time the judging coordinator entered the room a few minutes later, several non-judges had come in, but he stood behind the judges' table and explained quietly, "If your own students are playing, you don't need to take yourself out. We trust you." Moments later, he added, "I've judged my own students before … If you want to take yourself out of the discussion about that student, that's fine … but you don't need to say that it's your student, or that it's not your student" (Field notes, March 12, 2010).

This approach is dramatically different from the Tchaikovsky Competition, where the 2011 organizing committee chair specifically tried to assemble a panel of judges without any students in the

competition. Other high-level international competitions have implemented similar strategies. The Paganini Competition, for instance, has established a formal rule that prohibits judges from evaluating their own students (Paganini International Violin Competition "Criteria for Ranking by the Jury," personal communication, July 7, 2010). Even with this rule, musicians are still allowed to judge competitions that their students are competing in; they just cannot participate in evaluating their own students. More generally, based on my data, competitions with these kinds of rules do not have any way to punish evaluators who break the rules, other than not inviting them back to judge in the future. Without any official policing mechanisms with serious sanctions, music competitions simply need to rely on the trustworthiness of their evaluators. This is acceptable partly because of the organizational and power dynamics surrounding the music world, where competitions do not need to defend their rules or practices to any more powerful organizations.

In general, these trends where music insiders are typically allowed to judge competitions that their students have entered are in stark contrast to figure skating, where coaches cannot judge events where skaters they have recently worked with are competing, even if they have only worked with a competing skater for a few hours. All of these findings illustrate how evaluation practices in classical music tend to be relatively informal, especially compared to figure skating.

Discussion

Based on this discussion of evaluation in figure skating and classical music, organizational structure has a tremendous impact on evaluation practices and their degrees of formality and standardization. In particular, my findings suggest that highly centralized settings governed by more powerful organizations and where competitions build on each other tend to use more formal and standardized evaluation practices compared to other settings with fewer restrictions. In Table 5, I highlight these factors in terms of how they differ in relation to skating and music competitions.

	Figure Skating	Classical Music
Organizational Context	Sport with artistic components	Art
Centralization	Highly centralized	Not centralized
Governance	Internationally governed	Not internationally governed
	International Skating Union (ISU): • Central governing body with power to impose internationally standardized rules • 101 member federations representing different countries	No governing bodies
	International Olympic Committee (IOC): • Governing body for all Olympic sports	Judging officials and other elites: • Not accountable to more powerful organizations
	ISU officials and other elites • Accountable to IOC • Low levels of power	• High levels of power
Competition Structure*	Competitions tend to build on each other.	Competitions usually do not build on each other.
	Need for internationally standardized rules	Less need for standardized rules

*These distinctions reflect my findings on figure skating and classical music competitions and do not necessarily apply to other activities in these arenas.

Table 5: Factors Affecting Evaluation Practices in Figure Skating vs. Classical Music
Source: Author's work

My findings indicate that organizational structure is pivotal to why these settings have developed formal and standardized versus informal and varied evaluation practices. For example, whether or not everyone is working toward the same pinnacle event, such as the Olympics, makes a major difference in whether an evaluative setting is expected to adopt formal, standardized evaluation practices. In contexts where competitions do not build on each other, with international competitions that are not advancing participants toward a pinnacle and are not run by the same organization, there is much less pressure for formal, standardized evaluation. These divergences in competition

structure interact with variations in levels of centralization and governance structure to influence approaches to evaluation practices.

Also based on these findings, when an organization experiences a legitimacy crisis like a judging scandal, that crisis interacts with these other factors to influence the trajectory of the evaluation practices involved. When rules are perceived as fair, it is usually less likely that anyone would want to change them, which tends to contribute to maintaining the status quo. Even if there is a legitimacy crisis, such as a judging scandal where someone does not follow the rules, if the organizations involved in that crisis do not have to worry about a more powerful entity like an international governing body, those organizations would often be able to come up with a solution that contributes to maintaining the status quo as well. This has happened with many music competitions that have dealt with judging scandals. However, when there is a legitimacy crisis in a setting that needs to worry about international governance, especially by a more powerful organization, outsiders would be more likely to demand change. This helps explain why there has been so much more pressure, and ultimately change, in response to judging controversies in figure skating compared to classical music.

This analysis is limited to just two cases, so researchers should conduct further studies in other areas before definitive conclusions can be made about variables that affect evaluation practices and how rules develop, including the factors I have emphasised here as well as others. Other types of power relationships, publicity, and trust dynamics, for example, also come up a lot throughout my data as influencing evaluation practices in skating and music and would be worth including in further analyses. Adding more cases to this discussion would increase our understanding beyond what an in-depth qualitative study like this allows for.

In this article, I have emphasized the significance of organizational structure in shaping evaluation practices, including degrees of reliance on formal, standardized, quantitative rules. Porter (1995) suggests that quantification is such a powerful legitimation device because it appears to separate knowledge from the people who created that knowledge, which many people believe decreases the likelihood of corruption, dishonesty, and otherwise "hazy thinking" (85). While the kind of debate and persuasion that characterizes judges' deliberations at many music competitions is often seen as more legitimate in a context where people in power positions can make decisions about evaluation practices without pressure from external governing bodies and without worrying about affecting other competitions, in a highly centralized, internationally governed context surrounded by a relative lack of power, this type of evaluation would be much more difficult to defend. A common perception is that verbal reasoning allows for too much potential bias for these kinds of settings without providing any "clear

checks against errors of reasoning" (Porter 1995: 52). In many contexts with similarities to figure skating, decision-making based on this kind of deliberation would be considered illegitimate. In this sense, highly formal rule systems act as legitimation devices, especially when they revolve around numbers.

Conclusion

The findings presented here have implications for how we make decisions and distinctions and how we construct quality – based on adding up a series of individual choices about specific parts of some product or performance (as in figure skating) or through reaching consensus based on overall impressions through group discussion (as in some settings in classical music). Variations in valuing parts or wholes can be especially significant in performance settings because in many instances, the sum of a performance's various components might not reflect its overall impact, so this can affect the types of performances that are rewarded and ultimately valued. The contrast between making decisions individually or collectively can also affect outcomes, partly because this affects whether people can influence each other's evaluations. Some people in some settings might think that letting evaluators influence each other is problematic, while others believe that this can lead to a more balanced decision-making process and ultimately fairer outcomes.

More broadly, our evaluation practices often have profound effects on many important distinctions, such as how we think about "good" versus "bad" and "worthy" versus "unworthy," as well as the boundaries between these categories (Lamont 2012). Lamont (2012) emphasizes that one aspect of fighting inequality is broadening how we define social worth, and she suggests that before we can do that, we need a better understanding of evaluation, which she calls a "complex, slippery, and often elusive sociological object" (203). My research contributes to this goal by outlining how evaluation works in two settings and identifying several factors – organizational context, centralization, governance, and competition structure – that other scholars should incorporate into research on different evaluative settings. Looking at how and why evaluation practices change and the effects of those changes would be a useful starting point, one that would also deepen our understanding of organizational responses to rule changes.

Evaluation, especially deciding what kinds of practices to use in various contexts, has been hotly contested in many areas. Debates surrounding how to make decisions about loans, hiring, evaluating teachers, and college admissions, as well as many other issues, illustrate this point. There is so much controversy around these issues and others linked to evaluation because these decisions are so

important for so many people, and many of them contribute to perpetuating existing inequality. When banks decide whether to give someone a loan, organizations decide whether to offer someone a job, or colleges decide whether to admit a student, the outcomes tend to have serious consequences for the people on the other side of those decisions, and those outcomes are often influenced by how formal these organizations' evaluation systems are. Many such organizations use very formal rules to make those decisions because they think this enhances their legitimacy, but formal guidelines usually favor people who look good on paper – people with high credit scores, degrees from prestigious universities, or high SAT scores, for example – which in turn tend to privilege people who have money and other resources. In contrast, more informal criteria typically give organizations more flexibility to reward more subjective indicators (such as work ethic, integrity, creativity, or interpersonal skills), take unique accomplishments into account, and make exceptions.

When organizations adopt more formal rules, in addition to their link to legitimacy, they often do so to limit corruption and encourage fairness, objectivity, and equality. Most people would think of these goals as very honorable. However, although following formal rules makes sense in certain contexts, especially where breaking rules could be very dangerous (take driving, for instance), formalization often leads to unintended consequences, many of which are negative in some way. A number of researchers have emphasized the potential drawbacks of trying to contain discretion (Perrow 1999; Lom 2016) and how many types of rules and measurement devalue qualities that are important but difficult to measure or account for with rules and tend to change what is being measured (for example, see Carruthers and Espeland 1991; Espeland and Sauder 2007; Sauder 2008; Sauder and Espeland 2009; Colyvas 2012; Lom 2016). Among the effects of the judging changes in figure skating has been a devaluation of the artistic side of the sport, in contrast to how many music evaluators continue to stress artistry. This personifies the potential drawbacks of highly formal rule systems and the potential advantages of more informal systems and indicates that in many cases involving evaluation, sometimes less is more.

Acknowledgments

This research was supported by a Research Grant from the University Research Grants Committee at Northwestern University. Earlier versions of parts of this manuscript were presented at the American Sociological Association Annual Meetings and the Arkansas Sociological and Anthropological Association Annual Conference. The author is grateful for helpful comments and suggestions from Bruce

Carruthers, Wendy Espeland, Gary Fine, Carol Heimer, Ed Powers, John Toth, and Claes-Fredrik Helgesson and the anonymous *Valuation Studies* reviewers.

References

Adler, Paul S. 2006. "Beyond Hacker Idiocy: The Changing Nature of Software Community and Identity." In *The Firm as a Collaborative Community: Reconstructing Trust in the Knowledge Economy*, edited by Charles Heckscher and Paul S. Adler, 198–258. New York: Oxford University Press.

Adler, Paul S. 2012. "The Sociological Ambivalence of Bureaucracy: From Weber via Gouldner to Marx." *Organization Science* 23(1): 244–266.

Adler, Paul S., and Bryan Borys. 1996. "Two Types of Bureaucracy: Enabling and Coercive." *Administrative Science Quarterly* 41(1): 61–89.

AAF (Alink-Argerich Foundation). 2023. "Foundation." https://www.alink-argerich.org/foundation, accessed 29 April 2024.

Becker, Howard S. 1982. *Art Worlds*. Berkeley, CA: University of California Press.

Brown, Richard H., and Elizabeth L. Malone. 2004. "Reason, Polities, and the Politics of Truth: How Science is Both Autonomous and Dependent." *Sociological Theory* 22(1): 106–122.

Carruthers, Bruce G., and Wendy N. Espeland. 1991. "Accounting for Rationality: Double-Entry Bookkeeping and the Rhetoric of Economic Rationality." *American Journal of Sociology* 97(1): 31–69.

Chambliss, Daniel F. 1989. "The Mundanity of Excellence: An Ethnographic Report on Stratification and Olympic Swimmers." *Sociological Theory* 7(1): 70–86.

Charmaz, Kathy. 2001. "Grounded Theory." In *Contemporary Field Research: Perspectives and Formulations*, 2nd edn, edited by Robert M. Emerson, 335–352. Long Grove, IL: Waveland Press.

Coakley, Jay J. 2001. "Sport in Society: An Inspiration or an Opiate?" In *Sport in Contemporary Society*, 6th edn, edited by D. Stanley Eitzen, 20–36. New York: Worth.

Coakley, Jay. 2015. *Sports in Society: Issues and Controversies*, 11th edn. New York: McGraw-Hill.

Colyvas, Jeannette A. 2012. "Performance Metrics as Formal Structures and through the Lens of Social Mechanisms: When Do They Work and How Do They Influence?" *American Journal of Education* 118(2): 167–197.

Daston, Lorraine, and Peter Galison. 2007. *Objectivity*. New York: Zone Books.

Deephouse, David L. 1996. "Does Isomorphism Legitimate?" *Academy of Management Journal* 39(4): 1024–1039.

DiMaggio, Paul J., and Walter W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* 48(2): 147–160.

Dobbin, Frank, and Erin L. Kelly. 2007. "How to Stop Harassment: Professional Construction of Legal Compliance in Organizations." *American Journal of Sociology* 112(4): 1203–1243.

Edelman, Lauren B. 2016. *Working Law: Courts, Corporations, and Symbolic Civil Rights*. Chicago, IL: University of Chicago Press.

Edelman, Lauren B., Linda H. Krieger, Scott R. Eliason, Catherine R. Albiston, and Virginia Mellema. 2011. "When Organizations Rule: Judicial Deference to Institutionalized Employment Structures." *American Journal of Sociology* 117(3): 888–954.

Eitzen, D. Stanley. 2012. *Fair and Foul: Beyond the Myths and Paradoxes of Sport,* 5th edn. Lanham, MD: Rowman and Littlefield.

Elsbach, Kimberly D., and Greg Elofson. 2000. "How the Packaging of Decision Explanations Affects Perceptions of Trustworthiness." *Academy of Management Journal* 43(1): 80–89.

Emerson, Robert M., Rachel I. Fretz, and Linda L. Shaw. 1995. *Writing Ethnographic Fieldnotes*. Chicago, IL: University of Chicago Press.

Erwin, Patrick M. 2011. "Corporate Codes of Conduct: The Effects of Code Content and Quality on Ethical Performance." *Journal of Business Ethics* 99: 535–548.

Espeland, Wendy N., and Michael Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." *American Journal of Sociology* 113(1): 1–40.

Feldman, Martha S. 1992. "Social Limits to Discretion: An Organizational Perspective." In *The Uses of Discretion*, edited by Keith Hawkins, 163–184. Oxford: Clarendon Press.

Fourcade, Marion. 2011. "Cents and Sensibility: Economic Valuation and the Nature of 'Nature.'" *American Journal of Sociology* 116(6): 1721–1777.

Frey, James H., and D. Stanley Eitzen. 1991. "Sport and Society." *Annual Review of Sociology* 17: 503–522.

Gunts, Bucky, and David Michaels. 2002. "Figure Skating." XIX Olympic Winter Games NBC. February 11, 2002.

Guttmann, Allen. 2003. "Sport, Politics and the Engaged Historian." *Journal of Contemporary History* 38(3): 363–375.

Hafner-Burton, Emilie M., and Kiyoteru Tsutsui. 2005. "Human Rights in a Globalizing World: The Paradoxes of Empty Promises." *American Journal of Sociology* 110(5): 1373–1411.

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

Hines, James R. 2006. *Figure Skating: A History.* Urbana, IL: University of Illinois Press; Colorado Springs, CO: World Figure Skating Museum and Hall of Fame.

Hines, James R. 2011. *Historical Dictionary of Figure Skating*. Lanham, MD: Scarecrow Press.

IOC (International Olympic Committee). 2021. "International Olympic Committee – History, Principles & Financing." https://olympics.com/ioc/overview.

ISU (International Skating Union). 2004. "Program Components Overview."
ISU (International Skating Union). 2010. "Communication No. 1611: Scale of Values, Levels of Difficulty and Guidelines for Marking Grade of Execution for Single and Pair Skating."
ISU (International Skating Union). 2012. "Some Key Dates in ISU History."
ISU (International Skating Union). 2022a. "ISU Results: XXIV Olympic Winter Games 2022: Men." http://results.isu.org/events/owg2022_Men.htm.
ISU (International Skating Union). 2022b. "ISU FAQ – Program Components." https://isu.org/media-centre/press-releases/2022-4/29705-isu-faq-components/file.
ISU (International Skating Union). 2022c. "Communication No. 2475: Single & Pair Skating Scale of Values season 2022/23."
ISU (International Skating Union). 2023a. "ISU History." https://isu.org/inside-isu/about/history.
ISU (International Skating Union). 2023b. "Member Federations." https://isu.org/inside-isu/about/member-federations.
International Tchaikovsky Competition. 2010. "The Scores of Harmony Processing System."
Jackson, Jon. 2005. *On Edge: Backroom Dealing, Cocktail Scheming, Triple Axels, and How Top Skaters Get Screwed*. New York: Thunder's Mouth Press.
Jacobsson, Bengt. 2000. "Standardization and Expert Knowledge." In *A World of Standards*, edited by Nils Brunsson and Bengt Jacobsson, 40–49. New York: Oxford University Press.
Jick, Todd D. 1979. "Mixing Qualitative and Quantitative Methods: Triangulation in Action." *Administrative Science Quarterly* 24(4): 602–611.
Kornberger, Martin, Lise Justesen, Jan Mouritsen, and Anders K. Madsen (eds.) 2015. *Making Things Valuable.* Oxford: Oxford University Press.
Lamont, Michèle. 2009. *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge, MA: Harvard University Press.
Lamont, Michèle. 2012. "Toward a Comparative Sociology of Valuation and Evaluation." *Annual Review of Sociology* 38: 201–221.
Latham, Alison, and Piers Spencer. 2002. "Competitions in Music." In *The Oxford Companion to Music*, edited by Alison Latham. Oxford Music Online. http://www.oxfordmusiconline.com/subscriber/article/opr/t114/e1527.
Li, Jiatao, and Yi Tang. 2010. "CEO Hubris and Firm Risk Taking in China: The Moderating Role of Managerial Discretion." *Academy of Management Journal* 53(1): 45–68.
Lofland, John, and Lyn H. Lofland. 1994. *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis*, 3rd edn. Belmont, CA: Wadsworth.

Lom, Stacy E. 2016. "Changing Rules, Changing Practices: The Direct and Indirect Effects of Tight Coupling in Figure Skating." *Organization Science* 27(1): 36–52.

Maccoby, Michael. 2006. "Health Care Organizations as Collaborative Learning Communities." In *The Firm as a Collaborative Community*, edited by Charles Heckscher and Paul S. Adler, 259–280. New York: Oxford University Press.

McCormick, Lisa. 2009. "Higher, Faster, Louder: Representations of the International Music Competition." *Cultural Sociology* 3(1): 5–30.

Mastrofski, Stephen D. 2004. "Controlling Street-Level Police Discretion." *Annals of the American Academy of Political and Social Science* 593: 100–118.

Mau, Steffen. 2019. *The Metric Society: On the Quantification of the Social*. Medford, MA: Polity Press.

Meyer, John W., and Brian Rowan. 1977. "Institutionalized Organizations: Formal Structure as Myth and Ceremony." *American Journal of Sociology* 83(2): 340–363.

Morrison, Richard. 2011. "Off-Key Antics Failed to Derail the Tchaikovsky Competition." *The London Times* July 6: Section T2, p. 11.

Osnos, Peter. 1974. "Fodor: Tied for Silver, Not Gold." *The Washington Post* July 5.

Perrow, Charles. 1999. *Normal Accidents: Living with High-Risk Technologies*, with a New Afterword and a Postscript on the Y2K Problem. Princeton, NJ: Princeton University Press.

Peterson, Diane. 2011. "Sonoma Violinist Advances to Tchaikovsky Finals." *The Press Democrat* June 25.

Porter, Theodore M. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.

Power, Michael. 2003. "Auditing and the Production of Legitimacy." *Accounting, Organizations and Society* 28(4): 379–394.

Roberts, Selena. 2002a. "Canada to Appeal on Skating; Judge Pressured, French Say." *The New York Times* February 14: Section A, Column 1: 1.

Roberts, Selena. 2002b. "Skating Group Proposes a New System of Judging." *The New York Times* February 19: Section A, Column 2: 1.

Sauder, Michael. 2008. "Interlopers and Field Change: The Entry of U.S. News into the Field of Legal Education." *Administrative Science Quarterly* 53(2): 209–234.

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

Service, Tom. 2011. "Everything to Play for at the Tchaikovsky Competition." *The Guardian* September 20.

Sheehan, Mike, and David Michaels. 2018. "Figure Skating." XXIII Olympic Winter Games NBC. February 8, 2018.

Shen, Wei, and Theresa S. Cho. 2005. "Exploring Involuntary Executive Turnover through a Managerial Discretion Framework." *Academy of Management Review* 30(4): 843–854.

Stanley, Alessandra. 1994. "Musical Tradition of Acrimony." *The New York Times* July 2.

Star, Susan L., and Martha Lampland. 2009. "Reckoning with Standards." In *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*, edited by Martha Lampland and Susan L. Star, 3–24. Ithaca, NY: Cornell University Press.

Stevens, Mitchell L. 2007. *Creating a Class: College Admissions and the Education of Elites*. Cambridge, MA: Harvard University Press.

Stinchcombe, Arthur L. 2001. *When Formality Works: Authority and Abstraction in Law and Organizations*. Chicago, IL: University of Chicago Press.

Stults, Raymond. 2010. "Rodzinski to Revive Tchaikovsky Competition." *The Moscow Times* January 29.

Sunshine, Jason, and Tom R. Tyler. 2003. "The Role of Procedural Justice and Legitimacy in Shaping Public Support for Policing." *Law and Society Review* 37(3): 513–548.

Timmermans, Stefan, and Marc Berg. 2003. *The Gold Standard: The Challenge of Evidence-Based Medicine and Standardization in Health Care*. Philadelphia, PA: Temple University Press.

Tyler, Tom R. 1988. "What is Procedural Justice? Criteria Used by Citizens to Assess the Fairness of Legal Procedures." *Law and Society Review* 22(1): 103–136.

Tyler, Tom R. 2000. "Multiculturalism and the Willingness of Citizens to Defer to Law and to Legal Authorities." *Law and Social Inquiry* 25(4): 983–1019.

Tyler, Tom R. 2003. "Procedural Justice, Legitimacy, and the Effective Rule of Law." *Crime and Justice* 30: 283–357.

USFSA (United States Figure Skating Association). 2009a. "Agreement to Comply with Code of Ethics."

USFSA (United States Figure Skating Association). 2009b. "Technical Panel Code of Ethics."

Washington, Robert E., and David Karen. 2001. "Sport and Society." *Annual Review of Sociology* 27: 187–212.

Weaver, Gary R., Linda K. Trevino, and Philip L. Cochran. 1999. "Corporate Ethics Programs as Control Systems: Influences of Executive Commitment and Environmental Factors." *Academy of Management Journal* 42(1): 41–57.

Weber, Max. 1968. *Economy and Society: An Outline of Interpretive Sociology*. Edited by Guenther Roth and Claus Wittich. New York: Bedminster.

WFIMC (World Federation of International Music Competitions). 2021. "The Federation." https://www.wfimc.org/federation.

WFIMC (World Federation of International Music Competitions). 2022. "WFIMC Statutes 2022." https://www.wfimc.org/federation.

Stacy E. Lom is Assistant Professor of Sociology at the University of Central Arkansas. Her research interests include evaluative cultures, particularly how they develop and their effects in different contexts; rules and formalization; quantification; organizational decision making; and the politics of knowledge production and consumption. Her research focuses on rule systems and evaluation in sport, the arts, and education.

Valuation Studies 11(1) 2024: 138-161

Research note

Contested Commensuration: The case of a valuation instrument for historical buildings

Tineke van der Schoor, Harro van Lente, and Alexander Peine

Abstract
Environmental values are becoming increasingly important in restoration of historical buildings, while energy interventions can seriously damage historical qualities. Cultural-historical values and environmental values are often considered difficult to commensurate, with energy engineers and heritage experts adhering to widely differing values and relating to different discourses. Valuation instruments are devised to deal with such value conflicts in restoration projects. In this research note, we study what such instruments perform in the case of assessing historical buildings. We ask how these instruments work, and how they afford, support and guide valuation processes? Furthermore, we enquire what is achieved and what is lost in the reconciliation of values. Theoretically, we start from the notion of commensuration, which allows comparison of values through a shared metric. Empirically, this research note examines the history and use of DuMo, an instrument that aims to reconcile cultural – historical and environmental values and provides a range of sustainable restoration strategies. We find that DuMo indeed performs commensuration of these conflicting values, but also keeps intact the epistemic authority of the two professions. Our claim thus is that valuation instruments can successfully perform commensuration while at the same being contested by involved professionals.

Keywords: valuation instruments; commensuration; cultural heritage; energy efficiency; sustainability; assessment

Tineke van der Schoor is Associate Professor at Hanze University of Applied Sciences.
Harro van Lente is Professor of Science and Technologies at Maastricht University.
Alexander Peine is Professor of Culture, Innovation and Communication at the Open University of The Netherlands.

https://doi.org/10.3384/VS.2001-5992.2024.11.1.138-161
Hosted by Linköping University Electronic press
http://valuationstudies.liu.se

Introduction

Since the beginning of this century, European and national policies put increasing pressure on building owners to perform energy performance assessments and acquire an energy label. Although this is not obligatory for historical buildings, experts nevertheless perceived this as a threat to the cultural and aesthetic values of historical buildings because the special needs of historical buildings are not considered in energy assessments (RDMZ 2001; Cassar 2009, 2011; Grytli et al. 2012; Pankhurst and Harris 2013). In many countries, heritage professionals were confronted with increasing political pressure to improve the energy efficiency of historical buildings. Furthermore, demands regarding comfort and energy efficiency were raised. Heritage professionals feared that these developments could lead to ill-advised retrofit measures which would damage historical buildings.

Theoretically, valuation instruments have been characterized as tools for 'commensuration', which is defined as a social process that condenses the evaluated aspects and combines, or reconciles them in a shared metric (Espeland and Stevens 1998; Espeland and Sauder 2007). Typically, commensuration requires boundary work, to maintain and regulate boundaries between conflicting values and the epistemic authorities of professions. In the case of historical buildings, energy performance measurements and historical value assessments are performed by experts belonging to different professional groups. Energy assessments are usually carried out by energy engineers; heritage assessments are performed by trained assessors with a background in architectural or building history.

Environmental and cultural-historical values are considered difficult to commensurate, or to measure on a common scale. For historical buildings, a gain in environmental value can cause an irreducible loss in cultural-historical value, to quote Norrström (2013: 2624): 'exhaustive refurbishments with the energy measures undertaken [can lead] to the destruction of cultural, historic and architectural values'. Even stronger, energy intervention can threaten the survival of the building itself (Schellen 2002; Stappers 2008). Stephenson (2008) emphasizes that cultural-historical values are strongly related to cultural identity, communal identity and self-identity. Environmental values, on the other hand, refer to the protection of the environment and the mitigation of climate change.

To deal with conflicting values, several instruments have been proposed to find a common scale and to produce commensurability (Stubbs 2004; Landorf 2011; Liusman et al. 2013). In these instruments, topics such as heritage, environment, economy and social issues are brought together in one valuation method. In this research note we investigate how commensuration is achieved in a particular valuation instrument, the so-called DuMo instrument. We investigate

how the DuMo instrument was designed and how it works. Mobilizing the perspectives of energy engineers and architectural history experts, we assess what is gained and lost in the articulation and commensuration of values when using the DuMo instrument.

The research note is organized as follows. In the following two sections, we further elaborate on values and valuation practices in the case of assessing historical buildings. In the next we present our methods and materials, while the following section introduces our case study of the DuMo instrument, including its development, procedures and experts' views on its application. We will describe how different values are represented in the instrument, how they are weighted and how the results of the measurements are interpreted and translated into metrics that are easy to communicate. Clearly, the commensuration of values is an ongoing challenge, as will be discussed in the penultimate section. We conclude that while the valuation instrument combines historical and environmental values in one instrument, it also keeps the epistemic authority of the two professions intact. Our claim is that valuation instruments achieve commensuration while remaining to be contested by involved professionals.

Conservation and sustainability

Conservation of historical buildings connects philosophical ideas and human values with technical interventions, as Drury (2012) remarks:

> Building conservation is distinctly different from the physical processes of repair and adaptation. It is an attitude of mind, a philosophical approach, that seeks first to understand what people value about a historic building or place beyond its practical utility and then to use that understanding to ensure that any work undertaken does as little harm as possible to the characteristics that hold or express those values. (Drury 2012: 1).

Conservation requires constant monitoring, decision making and acting on the materials that make up the structure. Moreover, new demands regarding comfort or functionality often prompt changes to the form, material or layout of buildings. Indeed, buildings are constantly reconfigured: they need daily management, regular maintenance (Denis and Pontille 2015), repair (Graham and Thrift 2007) and sometimes restoration (Yaneva 2008). Such reconfiguration work can be in conflict with the conservation principle of minimal intervention, because even minor changes can seriously damage the cultural-historical qualities of historical buildings.

On an international level, codes for conservation have been established by the International Council on Monuments and Sites

(ICOMOS), under the auspices of the United Nations. Charters have been negotiated for specific building types, archaeological sites, immaterial heritage or specific cultures (Pickard 1996; ICOMOS 2003; Fredheim and Khalaf 2016). Authenticity, as affirmed in the Charter of Venice (ICOMOS 1964) and the Nara Document (ICOMOS 1994), is considered essential for the knowledge and protection of cultural-historical values. Based on these international charters, national heritage agencies published practical guidelines for conservation (Stovel and Smith 1996; English Heritage 2008; RCE 2009). Figure 1 depicts the categorization of interventions based on their impact on the heritage character of the building.

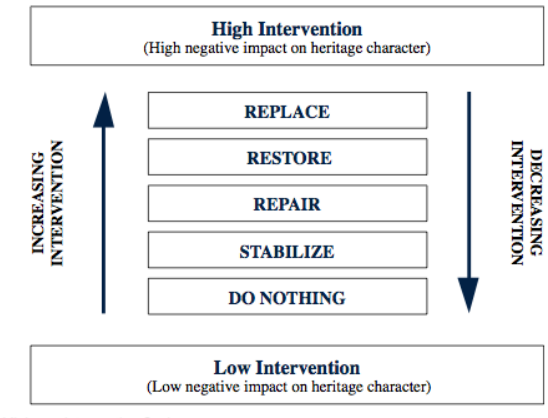

Figure 1: Minimum Intervention Scale, from FHBRO Code of Practice
Source: Stovel and Smith 1996: 18

Alignment of conservation with environmental values has been investigated before. Cluver and Randall (2010), Cassar (2011) and Godwin (2011) describe, for instance, technical results of energy efficient restorations. Cassar (2009) emphasizes the importance of monitoring the actual energy use and the effects of energy measures on the integrity and meaning of historical buildings. Several methods have been proposed to assess and manage the sustainability of historical buildings, cities, sites and landscapes, based on an integration of environmental, social and economic aspects. For example, Stubbs (2004) has developed a framework for the sustainability appraisal of the historic environment by distinguishing four topic areas: environmental, social/cultural, economic and sense of place. Landorf (2009) has crafted a model for sustainable management of industrial heritage sites with the two dimensions of long-term holistic management and the participation and empowerment of multiple stakeholders. Liusman et al. (2013) put forward a set of 'tailor-made' indicators for the assessment of heritage and applied these to a case study of a heritage building in Hong Kong. Eriksson et al. (2014) developed a software tool to support decisions on energy retrofit

measures. National heritage agencies have also published reports and guidelines on how to balance environment with conservation (RCE 2010; English Heritage 2013).

Nevertheless, commensuration of historical, environmental, social and economic values in building reconfigurations remains a contested space. These values cannot be reconciled easily, as is recognized by Wallace et al. (1999) and Pendlebury (2002). Moreover, Strange and Whitney (2003) argue for more research into the integration of sustainability in heritage management, especially as part of wider regeneration strategies.

Summarizing, the importance of balancing historical values with sustainability principles is increasingly recognized in the literature. Several generic frameworks have been proposed to balance historical and energy values, but theoretically this balancing is not fully understood. In the following section, we continue with the key question of this research note, how instruments of valuation afford and guide reconciliation of values.

Values, assessment instruments, commensuration and epistemic authority

The development of value assessment methods entails negotiations about the identification and relative importance of values, and how to bring them together in a shared framework. In this respect, Espeland and Stevens (1998) coined the notion of *commensuration*, which they define as a social process that brings various entities together in a common quantitative framework. Commensuration simplifies and reduces information, and subsequently imposes a shared metric on what remains. The metric is often used for ranking purposes, for example of schools (Espeland and Sauder 2007). Commensuration translates qualitative statements – on, say, how 'good' a school is – to quantities; in this respect Espeland and Sauder note that numbers create authority, circulate more easily and travel more easily to other contexts (2007: 17). Furthermore, they argue that 'Commensuration presupposes that widely disparate or even idiosyncratic values can be expressed in standardized ways and that these expressions do not alter meanings relevant to decisions' (2007: 12).

Professionals generally strive to gain 'epistemic authority', that is 'the legitimate power to define, describe and explain bounded domains of reality' (Gieryn 1999: 1). Boundary work is performed to construct and guard these professional domains, both regarding the division of labour and the definition and description of what is at stake. Gieryn (1995) points to the different epistemic authorities of professional groups and the concomitant incommensurabilities. When it comes to the experts involved with historic buildings, the different professional groups of architectural-history experts and energy engineers can be

characterized as 'two interdependent professions with more or less equal structural power and resources' (Gieryn 1995: 411). Since the nineteenth century, conflicts over boundaries between the domains of architects and engineers have been very common (Bruegmann 1978; Aibar and Bijker 1997; Saint 2007).

For commensuration of cultural-historical and environmental values these two types of values are brought together in a common framework. Commensuration entails boundary work to delineate epistemic authority and a division of labour. In our case study of the DuMo instrument this brings the question how the instrument affords and guides the reconciliation of values and how it depends on boundary work between professional groups.

Method and materials

The research was set up as a case study (Yin 1994), employing various materials to investigate the case of the development, goals and application of the DuMo instrument (see the case introduction section). To investigate experiences with DuMo assessments we held interviews with four members of the national steering group that was responsible for the development of DuMo. This includes architectural historians and building engineers, representing the main disciplinary perspectives in our study. The interviewees are identified by 'Exp.int.' professional background and a number. The interviews were transcribed and analysed in Atlas.ti. In the analysis of interviews we used an inductive approach (Charmaz 2014). We first identified and coded meaningful quotations in the interviews. We then performed a thematic analysis and compared the identified themes to the literature on cultural-historical valuation and sustainability. In the section Commensuration by an instrument, we discuss five themes: cultural-historical values, energy performance, intangible values, economic aspects and expert knowledge.

We also studied documentation on the DuMo instrument and reports on its application. First, we relied on the *Handboek Duurzame Monumentenzorg* [Handbook on Sustainable Conservation of Historical Buildings] (Van de Ven et al. 2011) (henceforth: the Handbook), which describes the assessment procedures and gives examples of finished projects. The Handbook also provides (online) assessment sheets. We investigated how experts are addressed in the Handbook and how specific professional values are transmitted through the instructions and energy improvement strategies. DuMo assessments are commissioned by the owners of historical buildings and performed by architectural experts and energy experts. A full report is typically between 75 and 100 pages long and includes detailed descriptions and illustrations of valuable features in the investigated building. DuMo reports remain the property of the commissioner and are not usually

publicly available. Here, we only refer to public sources. We examined DuMo reports (Dulski 2006, 2009, 2013), research notes about DuMo studies (van Bommel 2009, 2013; Nusselder 2009; de Jonge 2011) and archival materials of listed buildings across the Netherlands. Last, we compiled a comparative overview of 41 DuMo assessments which were published by NIBE.[1] The buildings investigated by NIBE were restored with a high energy ambition and revealed design strategies used to reconcile energy and historical values. This overview provided insights into the results of DuMo assessments, the applied restoration strategies and allowed comparison of the buildings to search for regularities, for example in building type, age and applied energy measures.

Case introduction: DuMo, an instrument for assessing historical and environmental values

In this section, we will first describe the background, development and design of the DuMo instrument. Then we introduce the different parts of the DuMo instrument and how the calculation for the DuMo label is constructed.

Development and design of the DuMo instrument

Since the 1990s, heritage professionals in the Netherlands have been faced with increasing political pressure to improve the energy efficiency of historical buildings. 'We realized that there was a threat, fear for political decisions on environmental standards, without recognizing that historical buildings are different' (Exp.int.1, historian). Furthermore, our interviewees stated that it was expected that owners of listed buildings would increase their demands regarding comfort and energy efficiency. They feared that these developments could lead to ill-advised retrofit measures and ultimately damage historical buildings. 'At that moment we already feared that the obligation for energy labels would be extended to historical buildings, or that new demands would be formulated for energy efficiency, we thought that in that case, we should be able to say, "You can", or "You can't"' (Exp.int.2, engineer).

In 2003 the Dutch Cultural Heritage Agency (RCE) initiated development of an instrument to value the sustainability of historical buildings. The method is called 'DuMo', which is an abbreviation of 'duurzame monumenten'.[2] Moreover, a knowledge base for historical

---

[1] NIBE is a Dutch consultancy specialized in sustainability. It was a member of the steering group responsible for the development of the DuMo instrument. The original list with projects can be obtained by the corresponding author. https://www.nibe-sustainability-experts.com/nl/dumo-duurzame-monumentenzorg

[2] Dutch for [sustainable historical buildings].

experts would be constructed, which could be used to assess proposals for energy efficiency. A steering group was set up, in which professionals from two backgrounds were represented: cultural heritage and sustainable energy. NIBE, a sustainable building consultancy, was commissioned to develop the new method. Architectural and building historians stressed that the instrument should be an expert model, to avoid lay persons from performing historical valuations. Users of the instrument would be architectural historians, energy engineers, civil servants and other heritage professionals.

First, an inventory was made of traditional sustainability features of historical buildings, such as rainwater cellars, window shutters, natural ventilation and insulation. Second, pilot buildings with recent energy efficiency measures were investigated, evaluating energy performance as well as loss of historical values. Assessment procedures, examples and strategies for improving the energy performance of historical buildings are described in the Handbook (2011). For practitioners, a code to download the DuMo calculation sheets is included. The Handbook also describes the pilot projects that formed the empirical basis of DuMo methodology.

Usually, DuMo assessments are carried out in preparation for a restoration project with a high energy efficiency ambition. To support the design process, the Handbook gives a broad range of appropriate technical strategies for energy efficiency and other sustainability measures. The restoration design is the basis for a second assessment of both cultural-historical and environmental values. The achievable gain in energy efficiency as well as the gains or losses of cultural-historical values are measured. The DuMo label (Figure 2) gives visual insight in the label-jump from the lowest level (G) to the highest level for existing buildings (A) that is achieved or expected after the energy efficient restoration.

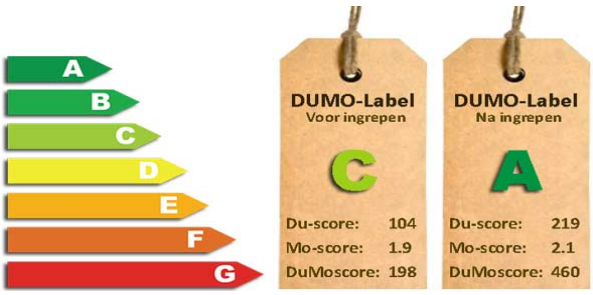

Figure 2: Example of DuMo Label Paushuize, Utrecht, the Netherlands
On the left, the EU-energy label categories ranging from A (most efficient) to G (least efficient). To the right, we see first DuMo Label C, which describes the situation before intervention. To the far-right DuMo Label A, after intervention. The lower part of the labels shows the scores for Sustainability (Du-scores), Historical values (Mo-scores) and the multiplied result (DuMo score).
Source: www.dumoprestatie.nl

Step 1: Historical values-coefficient

How does the DuMo instrument build this bridge between historical and environmental values? The reconciliation takes place in three steps: Step 1 is the assessment of historical values, leading to the Historical-Values Coefficient; step 2 is the assessment of sustainability performance, leading to the Sustainability Score. In step 3 the numerical results of these two assessments are multiplied, so the result is a merger of the two separate assessments. Importantly, valuation with DuMo is not an end in itself; it forms the basis for an energy efficient restoration plan.

Assessment of cultural-historical values in general requires a thorough investigation of building history, through examination in situ as well as by studying archival documents, local history and connections to important residents. A DuMo assessment requires considerable cultural-historical knowledge; therefore, it can only be carried out by a qualified evaluator.

Categories and items of historical value		
Architectural historical values (max. 60 points)	1	Building type and style
	2	Architectural quality
	3	Building quality
	4	Importance in oeuvre of architect
Cultural historical values (max. 27 points)	5	Importance with respect to historical themes
	6	Relation with local historical developments
	7	Relation with historical persons or events
Context values (max. 13 points)	8	Significance of environment for the building
	9	Significance of building for its environment
Completeness (factor 0,3-1)	10	How much of the historical material is preserved
	11	Technical state

Table 1: Categories and items of historical value DuMo instrument
Source: Based on Handbook Duurzame Monumentenzorg

The value-bearing features are noted on sketches or drawings of building elements, using the provided assessment sheets. Next, expert judgements are made on the value of each feature, which result in points. The main categories are architectural-historical values (max. 60

points), cultural-historical values (max. 27 points), context values (max. 13 points) and completeness (factor min. 0,3–max 1) (Table 1). Each category has two or more subcategories. For each subcategory the importance of the building is scored. All scores are transferred to an aggregate statement. In this part of the scoring procedure, values are translated in an ordinal scale, ranging from very positive (P), positive (Q), average (R), to negative (S). However, the underlying forms still contain the quantitative scores.

Total of the scores leads to the assignment of a 'Touchability' category to the building. This category combines two aspects: importance of the historical values and the vulnerability to interventions. DuMo defines four main touchability categories: A, B, C and X. Category A denotes 'museum quality', B stands for 'important historical value', C is characterized as 'flexible building with historical values' and X is reserved for buildings that are not listed[3] but do possess relevant cultural historical values. Furthermore, a numerical Historical Values Coefficient is given based on the total score. This coefficient later becomes the cultural-historical multiplier, as will be shown in Step 3 below. The DuMo report includes a comprehensive description of all identified values.

In this procedure we recognize first the certification of assessors, which must be architectural-historical experts. Second, a classification of values in four main categories with subcategories takes place. The large amount of information that is gathered for this procedure is significantly reduced and simplified, and scores are assigned for each feature and category. Next, a shared metric is applied which leads to the Historical Values Coefficient, a number. A further simplification is achieved by translating the scores to an ordinal scale (P, Q, R, S). The touchability category does not only express the importance but also the vulnerability of the building (A, B, C, X). Last, the procedure is validated by a second certified assessor.

Step 2: Sustainability Score

The Sustainability Score is based on GreenCalc+, a certified environmental assessment method[4] which was already being widely used when the DuMo instrument was developed. The sustainability sheets include three themes: water, materials and energy. The questions relate to the measurements of the building, technical installations, insulation, glazing, yearly energy use and so on. According to our

---

[3] 'Listed buildings' is commonly used in English to denote buildings that are placed on a national or municipal 'List of Buildings of Special Architectural or Historic Interest'.

[4] Audited and approved by the Bureau Veritas Certification and compliant with ISO 14040 and ISO 14044.

interviewees, these forms are not difficult to complete; any building professional could do it based on their education. The software performs calculations in the background and presents the viewer with the result. Figure 4 shows an example of the calculation sheet for sustainability, from the DuMo calculation package.

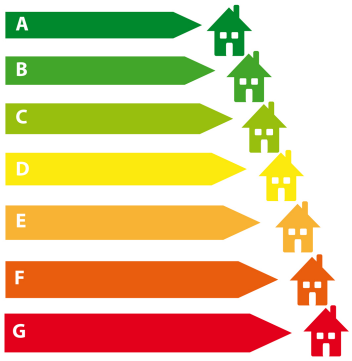

Figure 3: EU-labels for buildings
Source: European Parliament, https://www.europarl.europa.eu/resources/library/images/20171009PHT85660/20171009PHT85660_original.jpg

The resulting sustainability score can be translated to an energy label in the system of labels ranging from A (very efficient) to G (very inefficient). This labelling system has been used in the EU since 1994; it has regularly been updated and is very widely known. Figure 3 depicts the most recent EU-label system. Energy performance assessments are based on a benchmark, so new versions of the GreenCalc+ method reflect changes in energy efficiency technologies, national energy policies and building regulations. Because of more ambitious national energy goals and more efficient technologies, the rankings become more stringent as time progresses. This means that the building owner has to implement more energy measures to acquire a certain label, because the stakes are raised but the building remains the same. Presently, GreenCalc+ is merged with BREEAM-NL, a broad certification method for sustainable buildings managed by the Dutch Green Building Council (DGBC).

Figure 4: Example of sustainability assessment sheet used to calculate energy demand
In this part, the insulation values of the building and the installations for heat provision are described. The graph on the right shows the Sustainability Index (Du Index) for building part 1 (Bouwdeel 1)
Source: DuMo calculation sheets 2012, package provided with Handbook Duurzame Monumentenzorg

Calculation of the DuMo label

The aim of DuMo is to give a building a sustainability score that takes the special character of historical buildings into account. To this end, the result of the assessment of historical values (historical values coefficient) is multiplied by the calculated Sustainability Score. The resulting DuMo score is then translated into a traditional energy label ranging from A to G. In this way, both architectural historians and energy engineers find their respective expertise and valuation represented in the results of the assessment.

Take for example the scores of Paushuize, a house in Utrecht, built in 1517 for Adrianus VI, the only Dutch pope. The sustainability score is 104, which would lead to the assignment of label G, the lowest possible label. In DuMo, this score is multiplied by the historical values coefficient of 1.9, and this results in a score of 198. With this score, the building receives a DuMo label B (Dulski 2009). Scores can be checked in Figure 4.

We checked if energy performance is in any way related to historical characteristics. To that end, we examined the NIBE database to find relationships between historical values and potential improvements of energy performance. In these 41 cases we could find no relation between energy performance and building characteristics such as

touchability, age, type or function. For example, buildings with the highest energy-improvement can be found in all four 'touchability' categories; buildings date from the late Middle Ages to the first half of the twentieth century; building types vary from dwelling to factory to city hall; and functions (both original and new) vary from dwelling, museum, to office. We did not find any relation between historical values and energy performance. Thus, the assessment of these two parts of DuMO, as well as their multiplication, is not hampered by interdependencies.

High touchability of a building (category A, see previous subsection) relates to the importance of historical values present and its vulnerability to intervention. However, high touchability does not preclude energy improvement. Nevertheless, it does give an indication of the amount of care, creativity and architectural knowledge that will be required at the design stage of the restoration plan.

Commensuration by a valuation instrument

Certification of experts

The DuMo instrument was explicitly developed as an 'expert method' and is not designed to be used by laypersons. In the Handbook target groups are identified, such as contractors, heritage agencies, builders, architects or engineers. For each group an indication is given as to which parts of the DuMo instrument they can accomplish by themselves and for which parts they will need expert help (van de Ven et al. 2011: 11). In this framework, two professional groups are involved: energy engineers and architectural-historical experts.

To become a qualified valuator of historical values, one should either hold a masters in architectural history, or a postgraduate degree in building history and restoration. Furthermore, having assessed 150 historical buildings is an obligatory requirement for certification. Energy engineers do not receive cultural-historical training in their education. This inhibits engineers from recognizing historical values and conservation principles and can lead to ill-informed advice regarding insulation, glazing or appliances. As one of our interviewees states: 'If you arrive at a building with an experienced architectural-historical expert, he sees a hundred thousand things that I still overlook' (Exp.int.2, engineer). This lack of knowledge can also lead to failure to recognize the importance of authenticity. For example, producers or engineers offer new fixtures that 'look just the same' as historic items (Exp.int.2, engineer). However, according to international conservation principles (ICOMOS 1994) authentic material should have priority, because historical fabrics and materials

are a finite resource; once lost, they are irretrievable. This ties in with the principle of minimal intervention as relayed above (Figure 1) (Stovel and Smith 1996).

Categorisation of values and reduction of information

The DuMo procedures in Step 1 (design of the DuMo instrument) (i) identify what features are valuable; (ii) describe why these features are valuable; (iii) assess how valuable they are, which is congruent with the observations of Fredheim and Khalaf (2016).

In the DuMo procedure, architectural-historical and cultural-historical values are made commensurate by bringing scores of general design and specific features into clear and discrete categories and a judgement is provided which renders the categories comparable. The procedure reduces the large amount of gathered information to a numerical score for each subcategory. Within the division of historical values, a low score in one category can be compensated by a high score in another. The completeness of the building and its features is used as a factor, which may decrease the total score. The division in clear categories makes the valuation process more transparent, and numerical values are easier to communicate. The Historical Values Coefficient subsequently assigns a numerical value to the building, which is later used as multiplier in calculation of the DuMo label.

However, DuMo does not stop after assigning the Historical Value Coefficient. In the next step, buildings are graded according to a 'touchability' degree that further condenses the information about the building and makes buildings comparable along this metric. This concept was originally developed by one of our interviewees:

> At a certain moment, I just devised that concept, with the idea of investigating the different viewpoints for analysing a building, which could be a building-historical viewpoint, or a cultural-historical viewpoint, an important inhabitant who lived there, well, from these various perspectives, you can start the historical research of the building, with which you can underpin these stories, as well as explain much more clearly where the (historical) values actually reside. (Exp.int.1, historian).

According to this interviewee, during the development of DuMo various attempts were made to make touchability measurable, but in the end, it was decided to devise three simple grades (A, B, and C) and a non-grade (X).

> You have buildings that just are very untouchable, for example Hunting Lodge St. Hubertus; there you can do approximately nothing; then you have buildings where you can do something, but not too much, the Palace in Amsterdam is an example of that, you can do one thing and the other, but

> within boundaries; and then you have the average historical houses, where you have a lot more freedom. (Exp.int.1, historian).

The touchability degree thus is inversely proportional to degrees of freedom for the implementation of energy measures. With grade A you can do 'approximately nothing', with grade B you can do 'something, but not too much', and with grade C you 'have a lot more freedom'. For buildings that are characteristic, but not protected as a historical building, grade X is reserved.

Limitation of scope values

DuMo includes intangible values in assessments by identifying located stories, traditions, genealogies etc. (Stephenson 2008: 137). Values can be attached to intangible objects, such as stories, poems or music. Intangible values have been discussed in international venues under the auspices of UNESCO and are codified in the Burra Charter (Vecco 2010; ICOMOS 2013). Stories of buildings are part of the collected memories of buildings and places. Furthermore, such stories serve to interest local citizens and visitors and provide motivation for protection. 'Stories also make buildings sustainable' (Exp.int.3, historian). Nevertheless, as one of our interviewees says, energy measures are unlikely to damage the link with historical figures or important events. 'Take the Binnenhof as the [symbol of the] centre of government, and before that of the Graafschap (of Holland), I would say, if I apply double glazing it will still be this symbol' (Exp.int.1, historian).

In DuMo, values arising from relations to historical themes, local historical developments, historical persons or events are scored in the second part of the cultural-historical dimension (Table 1). Historical research can reveal for example if the building has played a significant part in local history or whether it is the birthplace of a local historical figure. However, DuMo does not take up lay values, collected memories or local traditions. We conclude that DuMo takes up a limited scope of cultural-historical values.

Energy performance in DuMo is calculated following existing certified methods for energy assessment. However, energy assessments generally do not take user behaviour into account. Likewise, DuMo does not consider user behaviour, although this has a large influence on energy use, especially in historical buildings. Indeed, according to our interviewees, there is a risk that unnecessary drastic interventions are proposed in the restoration plan.

Risks and limits of energy assessments

In the DuMo instrument, a standard energy assessment method is used to predict energy performance, however, this is not uncontested. Such methods are based on predicted heat loss, not on actual measurements. These predictions are based on premises that are impossible to ascertain in a historic building. 'You know nothing about existing buildings, except when you knock them down and build them up again, then you know!' (Exp.int.3, historian). Instead, our expert states, a simple yearly report of actual energy use would give a much more reliable indicator for energy performance. The difference between calculated and actual energy use ties in with the literature (see for example Aksoezen et al. 2015).

The second issue regarding energy performance is the influence of user behaviour. According to our interviewees, some users are quite happy to refrain from using certain draughty rooms in winter, if that is the price they have to pay to live in a historical building. They are also prepared to adapt their personal clothing and interior decoration, such as applying heavy curtains. 'Especially in historical buildings behaviour of users is very important for the actual energy use in a building' (Exp.int.2, engineer). If users do not show 'energy awareness', even in an energy neutral building the actual energy use can be much higher than expected. Therefore, monitoring actual use for heating and ventilation is necessary, both before and after restoration. 'First monitor what they actually do, where energy leaks away, or what it is used for' (Exp.int.3, historian).

The heavy influence of user behaviour together with the impossibility of rating a historical building is the reason one of the interviewees concludes that it is the user, not the building which should be labelled. In her view, this also lays the burden where it belongs, because it is actually the user that needs heating, not the building. 'Because the building doesn't mind if it is draughty, nobody cares, or if doors clatter, doesn't matter! (…) you just have to take care that it does not get wet, that is much more important. Don't get wet and keep it nicely draughty' (Exp.int.3, historian).

Last, according to our interviewees, it is not only user behaviour but also user knowledge that should be taken into account. Users have specific information about their building, they know where cold draughts exist and what spaces are especially cold or moist. Therefore, from this viewpoint, the starting point for energy advice should be the building itself and how to improve energy efficiency with minimal intervention.

Alignment of professional groups

In a DuMo assessment, these two professional groups are brought together. Working in DuMo projects teaches engineers to appreciate

their own limits of knowledge: 'In all these years I have learned what I don't know (laughing), that is the difference with the Energy Performance advisors, I think, they don't know what they don't know – and I do' (Exp.int.2, engineer).

Expert cultural-historical knowledge is often lacking in many municipal organizations. Therefore, before giving a building licence for restoration work, municipalities in the Netherlands typically employ experts from provincial agencies and rely on (compulsory) advice from RCE. Civil servants of small municipalities are not experienced valuators: 'And I exaggerate enormously, but such a civil servant does historical buildings on Monday, management of the swimming pool on Tuesday, and on Thursday he takes care of parking facilities. So, they have only limited time for heritage, which makes them very uncertain' (Exp.int.2, engineer). This uncertainty precludes civil servants from supporting owners by finding appropriate solutions for energy efficiency in historical buildings.

Owners of historical buildings are also laypersons, who although they often display great interest in their building, are usually not trained in architectural history. Some municipalities advise citizens to use freely available DIY checklists such as the Groene Menukaart to assess their historic buildings. However, according to our interviewees, it can still be difficult for laypersons to identify the valuable features of their building, because they do not recognize the historical styles. Owners can be overwhelmed by advisors that push expensive, heavy equipment that does not suit their building, says one of our interviewees. This development is exacerbated by the pressure for energy labels: 'What you often encounter is that people are overwhelmed by so many parties. There comes a contractor, there is the energy advisor, saying you should implement installation of so-and-so, and before you know it, they have done things that don't fit the building at all. Don't fit the use-pattern at all!' (Exp.int.3, historian). Laypersons are advised by energy engineers who clearly lack the knowledge of cultural-historical values, but nonetheless push their energy solutions, thereby potentially causing considerable damage. Nevertheless, the expertise of the energy engineer is indispensable in the development of a restoration plan. Here, energy interventions that are specifically suited to historical buildings are required, such as those that are described in the Handbook.

So, the DuMo instrument brings professional groups together in a restoration process. Cooperating in a DuMo project can help historians and engineers to acknowledge each other's expertise and their own limits of knowledge. DuMo standardizes and simplifies the process of valuation. However, DuMo does not make expert knowledge superfluous; in particular the assessment of historical values requires considerable cultural-historical knowledge. Furthermore, although the assessment of energy performance is

relatively straightforward, the design of energy interventions that are appropriate for historical buildings still requires expert knowledge.

Conclusion

The key question of this research note is how valuation instruments afford and guide the reconciliation of values that are difficult to commensurate. We studied the so-called DuMo instrument that seeks to commensurate heritage values with environmental criteria, and we traced in detail the steps that constitute the instrument. In our research note, we followed the development of the DuMo instrument and interviewed key persons working with DuMO. They shared their insights about the meanings of the number that is the outcome of the valuation process. Furthermore, the method forms the basis for restoration plans that include measures to improve the energy performance of the historical building. Therefore, we argue that assessing buildings with DuMo can be considered performative; not only does it give a valuation of the building in the form of an energy label in proportion to the identified historical values, but it also suggests pathways to improving energy performance. In our case study, we found that DuMo has also stimulated innovation; it brought about the development of new energy measures that are suitable for historical buildings. However, we noted that threats perceived by the heritage community in the early 2000s are still present. Both our case study and literature study suggest that if energy performance assessments were to become the basis for compulsory measures in historical buildings, the historical values of heritage would be seriously under threat. Therefore, the aims and procedures of instruments like DuMo are still very relevant today.

We studied how the DuMo instrument commensurates values from separate domains, e.g. energy and cultural history. Inspired by the conceptualization of Espeland and Stevens (1998) of commensuration processes, we suggest that understanding DuMo commensuration is achieved by virtue of the following six processes:

Certification of assessors. A precondition of DuMo is that in particular the assessment of cultural-historical values requires 'the right' valuators, as explained in Step 1. So, valuators are selected, and certification schemes are employed.

Categorization. Different values and aspects are divided into two domains or dimensions: cultural-historical values and sustainability values, including energy efficiency. Within these two domains further categorization takes place; different values are articulated and acknowledged as categories with questions or subcategories that can be scored.

Reduction. Not all types of values are included, leading to a reduction in information. A monument may have more than historical values and energy performance properties; it may, for instance, also be appreciated as a touristic highlight, as a meeting place for a community or as a token of spiritual value.

Simplification takes place by condensing multiple themes and aspects of cultural history in concise questions to be answered by expert judgement. The adoption of the EU-label system simplifies energy assessment.

Scoring of values takes place in both dimensions. Scoring for historical values is based on awarding points for valuable features, leading to qualifications on a scale from very positive to negative. The sum total leads to the historical value coefficient and the Touchability degree (A, B, C or X). The energy score is expressed in an interval scale.

Shared metric on what remains. DuMo produces a single number as outcome of the valuation process. The multiplication of the historical value-coefficient and the energy score makes the cultural-historical value have an effect on the energy label. Notably, if the cultural-historical value is low, there is no increase to the label. We note however that the shared metric in this case is basically an 'energy metric' with a cultural-historical multiplier. The resulting DuMo label acts as a translation of historical values in numbers, which is actionable in the different worlds of heritage experts and energy professionals.

The DuMo instrument did more, however. Efforts of commensuration also bring professional groups together and strengthen their identity. The ongoing discourse about energy and historical buildings is also a conflict over 'epistemic authority', that is 'the legitimate power to define, describe and explain bounded domains of reality' (Gieryn 1999: 2). The new guidelines for energy assessment of buildings have been interpreted as an infringement on the 'epistemic authority' of historical building professionals. They felt that their knowledge and experience were not taken into account, to the detriment of the historical buildings they care about. As explained earlier, development of the DuMo instrument was meant to settle or redraw the boundaries between the involved professions and to clearly state which tasks should be left exclusively to architectural historians. The demarcation of boundaries between the worlds of architectural historians and energy experts is reinforced with the separate assessment forms of the DuMo instrument, to be filled out by the respective professionals.

In other words, the DuMo instrument paradoxically brings reconciliation across borders by reinforcing boundaries; it keeps the epistemic authority of the two professions intact through separate

assessment forms. So, while environmental and cultural values are reconciled in one final score, their different assessment methods and results are fully acknowledged and maintained in separate forms. Our claim thus is that valuation instruments can perform commensuration and at the same time guard the boundaries of separate domains by acknowledging multiple professions.

Acknowledgments

We express our sincere thanks to the interviewees and to the reviewers who have commented on the previous version of this research note.

References

Aibar, Eduardo, and Wiebe E Bijker. 1997. "Constructing a City : The Cerdà Plan for the Extension of Barcelona." *Science, Technology & Human Values* 22(1): 3–30.

Aksoezen, Mehmet, Magdalena Daniel, Uta Hassler, and Niklaus Kohler. 2015. "Building Age as an Indicator for Energy Consumption." *Energy and Buildings* 87(0): 74–86.

Bommel, Bert van. 2009. *Duurzame Monumentenzorg: Een Spreekbeurt in Gent. Praktijkreeks Cultureel Erfgoed*. Vol. afl. 6, nr. Den Haag: Sdu.

Bommel, Bert Van. 2013. "Terugblik Op Een Geslaagd Project, de Restauratie van Het Koninklijk Paleis Amsterdam." *Bulletin KNOB* 112(2): 68–79.

Bruegmann, Robert. 1978. "Central Heating and Forced Ventilation: Origins and Effects on Architectural Design." *Journal of the Society of Architectural Historians* 37(3): 143–60.

Cassar, May. 2009. "Sustainable Heritage, Challenges and Strategies for the Twenty-First Century." *APT Bulletin* 40(1): 3–11.

Cassar, May. 2011. "Energy Reduction and the Conservation of Cultural Heritage: A Review of Past, Present and Forthcoming Initiatives." *International Preservation News* 55: 127–29.

Charmaz, Kathy. 2014. *Constructing Grounded Theory. Introducing Qualitative Methods*. 2nd ed. London: SAGE.

Cluver, John, and Brad Randall. 2010. "Saving Energy in Historic Buildings, Balancing Efficiency and Value." APT Bulletin 41 (1): 5–12.

Cultural Heritage Agency. 2009. *Erfgoedbalans 2009, Archeologie, Monumenten En Cultuurlandschap in Nederland*. Edited by E Beukers, M. De Boer, E Van As, H.J.P.M. Van den Berselaar, E Beukers, O. Brinkkemper, P.W.F. Brinkman, et al. Amersfoort: Rijksdienst voor het Cultureel Erfgoed. www.rijksdienstvoorhetcultureelerfgoed.nl.

Denis, Jerome, and David Pontille. 2015. "Material Ordering and the Care of Things." *Science, Technology, & Human Values* 40(3): 338–367.

Drury, Paul. 2012. "Conservation: An Evolving Concept." The Building Conservation Directory, 2012. http://www.buildingconservation.com/articles/conservation/conservation.htm.

Dulski, Birgit. 2006. "Restauratie Woonhuis 'De Vergulde Schoe' Duurzame Monumentenzorg Bij Een Historisch Woonhuis in Middelburg." Bussum.

Dulski, Birgit. 2009. "Duurzame Monumentenzorg Paushuize Utrecht." Bussum.

Dulski, Birgit. 2013. "Duurzaamheidsonderzoek Religieus Erfgoed Deventer; Verduurzamingsmogelijkheden Religieus Erfgoed Op Basis van Zes Casestudies." Bussum.

English Heritage. 2008. "Conservation Principles, Policies and Guidance." *Historic England* 25 January 2010. https://historicengland.org.uk/advice/constructive-conservation/conservation-principles, accessed 26 April 2024.

English Heritage. 2013. "Heritage Works, The Use of Historic Buildings in Regeneration, A Toolkit of Good Practice." *Historic England* 28 April 2017. https://historicengland.org.uk/images-books/publications/heritage-works/, accessed 26 April 2024.

Eriksson, Petra, Carsten Hermann, Sára Hrabovszky-Horváth, and Dennis Rodwell. 2014. "EFFESUS Methodology for Assessing the Impacts of Energy-Related Retrofit Measures on Heritage Significance." T*he Historic Environment: Policy & Practice* 5(2): 132–149.

Espeland, Wendy Nelson, and Michael Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." *American Journal of Sociology* 113(1): 1–40.

Espeland, Wendy Nelson, and Mitchell L. Stevens. 1998. "Commensuration as a Social Process." *Annual Review of Sociology* 24(1): 313–343.

Fredheim, L. Harald, and Manal Khalaf. 2016. "The Significance of Values: Heritage Value Typologies Re-Examined." *International Journal of Heritage Studies* 22(6): 466–481.

Gieryn, Thomas F. 1995. "Boundaries of Science." In *Handbook of Science and Technology Studies*, edited by Sheila Jasanoff, Gerald E. Markle, James C. Petersen, and Trevor Pinch, 1st ed., 393–443. Thousand Oaks, CA etc.: Sage Publications.

Gieryn, Thomas F. 1999. *Cultural Boundaries of Science : Credibility on the Line*. Chicago, IL: University of Chicago Press.

Godwin, P. J. 2011. "Building Conservation and Sustainability in the United Kingdom." *Procedia Engineering* 20(0): 12–21. Graham, Stephen, and

Nigel Thrift. 2007. "Out of Order, Understanding Repair and Maintenance." *Theory, Culture & Society* 24(3): 1–25.

Grytli, E, L Kværness, L S Rokseth, and K F Ygre. 2012. "The Impact of Energy Improvement Measures on Heritage Buildings." J*ournal of Architectural Conservation* 18(3): 89–106.

ICOMOS. 1964. "International Charter for the Conservation and Restoration of Monuments and Sites (the Venice Charter 1964)." Venice: International Council of Monuments and Sites.

ICOMOS. 1994. "The Nara Statement on Authenticity." Nara: ICOMOS. https://doi.org/10.1063/1.4748569.

ICOMOS. 2003. "Charter - Principles for the Analysis, Conservation and Structural Restoration of Architectural Heritage." Victoria Falls: ICOMOS.

ICOMOS. 2013. "Understanding and Assessing Cultural Significance, Practice Note Nr. 1." https://australia.icomos.org/wp-content/uploads/Practice-Note_Understanding-and-assessing-cultural-significance.pdf.

Jonge, Wessel de. 2011. "Report: Duurzaam Verbeterd! Bepaling Duurzaamheid Herbestemming Karakteristieke Gebouwen." https://www.climaticdesign.nl/images/Rapport_duurzaam_verbeterd_2012.pdf.

Landorf, Chris. 2009. "A Framework for Sustainable Heritage Management: A Study of UK Industrial Heritage Sites." *International Journal of Heritage Studies* 15(6): 494–510.

Landorf, Chris. 2011. "A Future for the Past: A New Theoretical Model for Sustainable Historic Urban Environments." *Planning Practice and Research* 26(2): 147–65.

Liusman, Ervi, Daniel C. W. Ho, and Jannet X. Ge. 2013. "Indicators for Heritage Buildings Sustainability." *CESB13–Central Europe towards Sustainable Building 2013*, 26-28 June 2013, Prague, 689–692.

Norrström, Heidi. 2013. "Sustainable and Balanced Energy Efficiency and Preservation in Our Built Heritage." *Sustainability* 5(6): 2623–2643.

Nusselder, E J. 2009. "De Juiste Balans Tussen Behoud En Verbetering." *Herenhuis*, 2009.

Pankhurst, Caroline, and Andrew Harris. 2013. "Conservation and Innovation – The Challenge of 'Eco' Renovation in Heritage Buildings." *Journal of Architectural Conservation* 19(1): 18–34.

Pendlebury, John. 2002. "Conservation and Regeneration: Complementary or Conflicting Processes? The Case of Grainger Town, Newcastle upon Tyne." *Planning Practice & Research* 17(2): 145–158.

Pickard, Rob. 1996. *Conservation in the Built Environment*. Edinburgh Gate: Longman.

RDMZ. 2001. "Duurzame Monumentenzorg." Zeist.

Rijksdienst voor het Cultureel Erfgoed. 2010. "Duurzaamheid Voor Het Oprapen." Amersfoort: Rijksdienst voor het Cultureel Erfgoed. https://english.cultureelerfgoed.nl/.

Saint, Andrew. 2007. *Architect and Engineer: A Study in Sibling Rivalry.* New Haven: Yale University Press.

Schellen, H L. 2002. "Heating Monumental Churches Indoor Climate and Preservation of Cultural Heritage." Technische Universiteit Eindhoven. http://alexandria.tue.nl/extra2/200213875.pdf.

Stappers, Marc. 2008. "Isoleren van Monumenten: De Keerzijde van de Medaille." In *Energieprestaties En Monumentale Constructies*, edited by H. De Clercq and N. Vernimme, Brussel: 59–74. Brussel: Koninklijk Instituut voor het Kunstpatrimonium en Vlaams Instituut voor het

Onroerend Erfgoed. https://docomomo.be/publication/nl-energieprestaties-en-monumentale-constructies/.
Stephenson, Janet. 2008. "The Cultural Values Model: An Integrated Approach to Values in Landscapes." *Landscape and Urban Planning* 84: 127–139.
Stovel, Herb, and Julian Smith. 1996. "FHBRO Code of Practice." Canada: Federal Heritage Buildings Review Office (FHBRO).
Strange, Ian, and David Whitney. 2003. "The Changing Roles and Purposes of Heritage Conservation in the UK." *Planning Practice & Research* 18 (2–3): 219–229.
Stubbs, Michael. 2004. "Heritage-Sustainability: Developing a Methodology for the Sustainable Appraisal of the Historic Environment." *Planning, Practice & Research* 19(3): 285–305.
Vecco, Marilena. 2010. "A Definition of Cultural Heritage: From the Tangible to the Intangible." *Journal of Cultural Heritage* 11(3): 321–324.
Ven, H Van de, E J Nusselder, M Haas, and B Dulski. 2011. *Handboek Duurzame Monumentenzorg. Theorie En Praktijk van Duurzaam Monumentenbeheer*. 2e ed. Rotterdam: SBR.
Wallace, Julia, Marilyn Higgins, and Jeremy Raemaekers. 1999. "Architectural Conservation and Environmental Sustainability: Conflict or Convergence?" *Journal of Architectural Conservation* 5(2): 56–71.
Yaneva, Albena. 2008. "How Buildings 'Surprise': The Renovation of the Alte Aula in Vienna." *Science Studies* 21(1): 8–28.
Yin, Robert K. 1994. *Case Study Research: Design and Methods*. 5th edition. Beverly Hills: SAGE Publications.

Tineke van der Schoor studied Andragogy at the University of Groningen, where she graduated on a thesis on the environment and technology. She also graduated in Dutch language and literature, with a specialization in Middle Dutch and Medieval Studies. After a career working for different NGOs on environment and development, as well as working as a politician in local and regional government, she became a senior researcher at NoorderRuimte, Research Centre for the Built Environment, Hanze University of Applied Sciences. In her PhD thesis she investigated strategies for energy reconfiguration in the built environment, including strategies for sustainable restoration of heritage buildings, focused on energy efficiency and cultural–historical values. Currently she is associate professor of Sustainable Heritage. Research interests are the digitalization of the built environment, especially the role of digital instruments for restoration, maintenance, and communication purposes. A second research theme is the relation between climate change and heritage, including the effects of climate change on heritage and climate adaptation measures for preservation.

Harro van Lente was trained in physics and philosophy at the University of Twente, the Netherlands, and graduated in both disciplines. His PhD research led to the rise of the Sociology of Expectations, which studies how representations of the future shape current socio-technical developments. He has published more than 100 academic articles and book chapters about the dynamics of technology-society interaction, technology assessment, foresight, and the politics of knowledge production. Between 1999 and 2014 he taught innovation studies at Utrecht University. Since 2014 he has been full professor of Science and Technology Studies and Head of Department at the Faculty of Arts and Social Sciences, Maastricht University. He is a member of the Dutch Health Council and Chair of the Board of the Netherlands Graduate Research School of Science, Technology & Modern Culture (WTMC). In 2018 he received the Freeman prize for the co-edited book *Emerging Technologies for Diagnosing Alzheimer's Disease: Innovating with Care*. In 2023 he published at Routledge the co-edited book *Technologies in Decline: Socio-Technical Approaches to Discontinuation and Destabilisation*.

Alexander Peine is Professor of Culture, Innovation and Communication at the Open University of the Netherlands. He is also the founding chair of the Socio-gerontechnology Network and the chair of the societal advisory board of the EU's Joint Programming Initiative "More Years, Better Lives" (JPI-MYBL). Before joining the Open University NL, Alexander was an associate professor at Utrecht University, the laureate of a prestigious Max Weber post-doctoral fellowship at the European University Institute in Florence, and a principal investigator at the Centre for Technology and Society at Berlin University of Technology. His research explores the many intersections between population ageing, including the challenges it allegedly poses for care and health systems, and technological change, including the push towards more interactive and 'smarter' technologies. His work has expanded the usual drive in this area to think of technologies as interventions with a unique line of research that thinks of ageing, care and health as being co-constituted with technology.