

Magicians, unicorns or data cleaners? Exploring the identity narratives and work experiences of data scientists

Occupational
identity of data
scientists

253

Lukas Goretzki

Stockholm School of Economics, Stockholm, Sweden, and

Martin Messner and Maria Wurm

*Department of Organization and Learning, Universität Innsbruck,
Innsbruck, Austria*

Received 13 January 2022
Revised 14 March 2023
Accepted 15 June 2023

Abstract

Purpose – Data science promises new opportunities for organizational decision-making. Data scientists arguably play an important role in this regard and one can even observe a certain “buzz” around this nascent occupation. This paper enquires into how data scientists construct their occupational identity and the challenges they experience when enacting it.

Design/methodology/approach – Based on semi-structured interviews with data scientists working in different industries, the authors explore how these actors draw on their educational background, work experiences and perception of the contemporary digitalization discourse to craft their occupational identities.

Findings – The authors identify three main components of data scientists’ occupational identity: a scientific mindset, an interest in sophisticated forms of data work and a problem-solving attitude. The authors demonstrate how enacting this identity is sometimes challenged through what data scientists perceive as either too low or too high expectations that managers form towards them. To address those expectations, they engage in outward-facing identity work by carrying out educational work within the organization and (paradoxically) stressing both prestigious and non-prestigious parts of their work to “tame” the ambiguity and hype they perceive in managers’ expectations. In addition, they act upon themselves to better appreciate managers’ perspectives and expectations.

Originality/value – This study contributes to research on data scientists as well as the accounting literature that often refers to data scientists as new competitors for accountants. It cautions scholars and practitioners alike to be careful when discussing the possibilities and limitations of data science concerning advancements in accounting and control.

Keywords Identity, Identity work, Data science, Digitalization

Paper type Research paper

Introduction

Inspired by advances in digital technologies and the exponential growth in the amount of data available, organizations increasingly deploy data science methods to translate “big data” into useful insights. Propelled by the “dream of perfect information and rational

© Lukas Goretzki, Martin Messner and Maria Wurm. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

The authors thank the interviewees for taking the time to share their experiences with them. Earlier versions of the paper were presented to the IPA conference 2021, to the EAA-VARS, and to research seminars at NHH Bergen, ESSEC Business School, Concordia University, the University of St. Gallen and the University of Innsbruck. The authors thank the participants and discussants at these events for their helpful comments. Particular thanks go to the two reviewers and the editor for their helpful guidance.



Accounting, Auditing &
Accountability Journal
Vol. 36 No. 9, 2023
pp. 253-280

Emerald Publishing Limited
0951-3574

DOI 10.1108/AAAJ-01-2022-5621

decision-making” (Quattrone, 2016, p. 118), data science has garnered enormous interest in recent years, also within the accounting domain, where data science applications promise new possibilities for control and decision-support (Al-Htaybat and von Alberti-Alhtaybat, 2017; Arnaboldi *et al.*, 2017; Moll and Yigitbasioglu, 2019).

The rise in data science goes along with the emergence of new types of experts. A prominent group hereby are data scientists (Avnoon, 2021) who have been portrayed as the “missing piece of the big data puzzle” (Carillo, 2017, p. 607), the “most wanted” experts (Waller and Fawcett, 2013), and even as “The Sexiest Job of the 21st Century” (Davenport and Patil, 2012). Given such accounts that reflect the current “buzz” around this nascent occupation, it is not surprising that indeed more and more organizations (seek to) employ such expert staff. While researchers have started to examine the work of data scientists (e.g. Avnoon, 2021; Barbour *et al.*, 2018; Carter and Sholler, 2016; Vaast and Pinsonneault, 2021), our empirical understanding of this occupational group and their work experiences is still in its infancy. Our paper seeks to advance the evolving conversation about data scientists by focusing on challenges relating specifically to their occupational identity.

Occupational identity relates to how people make sense of the central characteristics of their work and their position in the organization, thereby developing a sense of coherence, distinctiveness and direction for themselves (Kärreman and Alvesson, 2001). The case of data scientists is an interesting one in this respect, as the nascent state of the occupation (Avnoon, 2021) implies that data scientists cannot easily draw on established identity scripts or templates to develop an occupational identity (see Watson, 2008). Indeed, many organizations do not have much experience with data science yet, and so data scientists often enter an organizational context in which no established understanding of what data scientists (should) do is available. Different stakeholders will nevertheless form some expectations towards these new positions that, however, may not necessarily be aligned with the aspired identity of the data scientists themselves. In our paper, we empirically enquire into the way in which data scientists construct their occupational identity as well as the challenges of realizing such an identity within the organizations they work for.

Studying these issues is relevant also from an accounting perspective. Accounting scholars have emphasized the role of data scientists as a new breed of knowledge workers who can provide managers (and accountants) with new insights extracted from big data (e.g. Bhimani and Willcocks, 2014; Moll and Yigitbasioglu, 2019; Oesterreich and Teuteberg, 2019; Richins *et al.*, 2017). Data scientists are said to be well-equipped to fulfil managers’ contemporary information needs and “desires” (Vollmer, 2019, p. 28) and to become important decision-supporters who may not only complement (see Bhimani and Willcocks, 2014, p. 479) but potentially even outdo other information professions (see Abbott, 1988), not least accountants (see, e.g. Moll and Yigitbasioglu, 2019; Vollmer, 2019). This suggests that accountants need to understand the work of data scientists and how best to cooperate with them. By examining what data scientists do and how they see their own role and identity within the organization, our paper informs discussions about the relative positioning of, and cooperation between, accountants and data scientists.

The paper draws upon empirical material collected through an interview study with 25 data scientists working in different organizations and industries. We explore how these data scientists construct their occupational identity and how they experience and manage identity-related challenges when performing their work. We identify three main components of data scientists’ identity: a scientific mindset, an interest in sophisticated forms of data work and a problem-solving attitude. We then show how enacting this aspired identity is (to some extent) challenged in data scientists’ everyday work experiences. We show how data scientists respond to these challenges, not least by re-emphasizing parts of their aspired identities when interacting with managers and by carrying out educational work within the

organization. In particular, we show how, on the one hand, data scientists feel the need to emphasize the sophisticated parts of their activity (e.g. statistical modelling and exploring patterns in the data) to accentuate the status of their work vis-à-vis other – and in their view technically less advanced – data workers. This is important for them to counteract an ambiguous definition of data science that, in their view, sometimes encompasses job profiles that they do not want to be confused with. On the other hand, the hyped expectations confronting them also motivate data scientists to point towards the mundane reality of their work, which to a large extent consists of rather non-prestigious activities like data cleaning and preparation. Thus, unlike accountants and to some extent contradicting their attempts to emphasize the sophistication of their work, data scientists tend to not brush their dirty work under the carpet (see [Morales and Lambert, 2013](#)) but rather put it on display to tame others' hyped expectations about their craft.

Examining identity-related challenges that data scientists experience in and through their day-to-day work, our findings caution accounting scholars and practitioners alike to be careful when discussing the possibilities and limitations of data science concerning advancements in accounting and control. Data scientists appear more reserved when it comes to the implications of their work on decision-making than the promissory discourse on data science seems to suggest. Our study also emphasizes the importance of engaging more closely, and from a more micro-level perspective, with the work of data scientists and their role in establishing the use of digital tools in decision-making, accounting and control processes. The fact that data scientists try to tone down expectations towards the possibilities of their tools of the trade (i.e. data science methods) and emphasize the mundane and at times tedious aspects of their work indicates a need for more in-depth, empirical research on how data science is used in organizations as well as whether and, if so, how it affects the work of managers and other actors.

The remainder of the paper is structured as follows. We first provide a review of the literature on data scientists, before elaborating on the theoretical background of our study. The section thereafter explains our data collection and analysis. We then present our empirical findings, before offering a discussion and a conclusion.

Prior research on data scientists

With the emergence of data scientists as a new occupational group, researchers and practitioners have started to comment on the roles and responsibilities of these actors. Several authors have observed that there is quite some ambiguity around the term “data scientist”, with different job profiles being subsumed under it. [Harris *et al.* \(2013\)](#), for instance, write that data science results from a “buzzword meat grinder” and that the ambiguity around the term makes it difficult to “efficiently match talent to projects” (p. 3). Similarly, [Baškarada and Koronios \(2017\)](#) state that organizations would often “lack clear understanding of the required roles . . . and skills” (p. 66) of people working in the data science domain. Besides, data scientists are often presented as all-rounders with a broad set of competencies, including business knowledge, programming knowledge, statistical knowledge and soft skills ([Carillo, 2017](#); [Davenport and Patil, 2012](#)), and some authors question whether one individual can incorporate all required skills ([Carillo, 2017](#); [Waller and Fawcett, 2013](#)). [Baškarada and Koronios \(2017\)](#) hence talk about data scientists as “mythical creatures” and “unicorns”.

Given the occupation's novelty, empirical research on data scientists' work experiences in organizations is still rather scarce. Some studies, however, offer first insights. [Harris and Mehrotra \(2014\)](#) analyze differences between data scientists and data analysts and emphasize that data scientists “[e]xplore, discover, investigate and visualize”, while analysts “[r]eport, predict, prescribe and optimize” (p. 16). [Carter and Sholler \(2016\)](#) complement these findings

by pointing out that data scientists often have a natural “curiosity” that motivates them and that they enjoy the creativity that goes along with their work. Similarly, [Muller et al. \(2019\)](#) highlight that data scientists actively shape data and “create” findings by applying their expertise and creativity.

Prior research furthermore indicates the existence of organizational challenges related to the application of data science within organizations. [Harris and Mehrotra \(2014\)](#) show “that many organizations suffer from a lack of trust in the technologies, the data and ultimately the data scientists themselves” (p. 16). The authors also find that “[n]either data scientists nor managers are very good at speaking the other’s language, and executives compound the problem by the way they manage data scientists” (p. 16). [Barbour et al. \(2018\)](#) add to these findings by highlighting different obstacles for the use of analytics. They argue that working with data can be difficult for organizations not least because of the “existence of hierarchies, siloes and feelings of data ownership” (p. 269). Moreover, functional experts would often not know what they want to do with data. The authors further show how data analysts seek to build relationships with functional experts to have richer conversations not only about the data as such but also about the problems to be addressed. Focusing on such inter-functional relations in data science projects, [Pachidi et al. \(2021\)](#) show how tensions related to professional authority and power can emerge when introducing new digital technologies to the workplace.

Although researchers have started to examine the work practices of data scientists, this work is still in its infancy, however, and “too little is known about how analytics is practiced in organizations or its implications for how organizations use data” ([Barbour et al., 2018](#), p. 280). What seems especially called for are detailed and theoretically guided analyses of the challenges that data scientists – as a nascent technical occupation ([Avnoon, 2021](#)) – experience in relation to their expert role and identity. Important initial scholarly work was conducted by [Vaast and Pinsonneault \(2021\)](#) who emphasize the ambiguous role that technology plays in data scientists’ occupational identity. They demonstrate how technology serves as an identity referent for data scientists that affects how they develop their occupational identity and how this can lead to identity-related tensions. More specifically, the authors suggest that digital technologies do not only enable data scientists to carry out their work and develop a sense of the self but also challenge their sense of distinctiveness vis-à-vis other actors who might also use such technology in the future, as it becomes more widely accessible and user-friendly.

In another recent study, [Avnoon \(2021\)](#) looks at how data scientists carry out their identity work with respect to their skills and status as a nascent technical elite group. The author shows that rather than claiming an expert status through specializing in particular areas, data scientists tend to adopt a so-called “omnivorous” approach to skill acquisition. This means that they claim an expert or elite status in the organization through constructing an identity as generalists who possess a wide spectrum of both hard and soft skills and a strong, continuous self-learning attitude. [Avnoon \(2021\)](#) argues that when trying to gain occupational status by emphasizing generality over specialization, data scientists try to distinguish themselves from both “old-school” technical snobs and non-technical occupations” (pp. 342–343). The author furthermore demonstrates that while data scientists embrace both technical and social skills in their identity work, they create a symbolic hierarchy between different skills “wherein maths, computer skills and statistics are awarded the highest position” (p. 344).

With the present study, we aim to further our understanding of data scientists and their identity work within the organization. While [Vaast and Pinsonneault \(2021\)](#) focus on the ambiguous role of technology in data scientists’ identity work and [Avnoon \(2021\)](#) looks at the role of skill acquisition and the importance of presenting oneself as a generalist, we explore how data scientists develop an understanding of their role in the organization through

experiences of their day-to-day work and associated interactions with other organizational groups, especially managers.

From an accounting perspective, exploring the identity (work) of data scientists appears interesting since members of this supposedly “new technical elite” (Avnoon, 2021) have been described as part of “a new generation of data experts and information analysts who offer expertise in modelling, simulation and visualisation to stakeholders” (Vollmer, 2019, p. 28). Entering an organization’s “information service marketplace” (Vollmer, 2019), data scientists have been portrayed as being able to threaten more established information professions such as accountants by making “many of the more routine aspects of accounting redundant” or by satisfying managers’ desires for better information (Vollmer, 2019, p. 28; see also Quattrone, 2016). This view is underpinned by the largely positive and promissory character of the data science discourse that presents data scientists as well-positioned to compete with other information providers (see also Moll and Yigitbasioglu, 2019). Indeed, data scientists seem to benefit from a very positive image, being in short demand and sought-after by many organizations. One would therefore assume that they do not suffer from the types of identity threats that accountants sometimes face. We know that accountants at times struggle for recognition and to realize their aspired identity as valued supporters of management (or “business partners”) (see Goretzki and Messner, 2019; Morales and Lambert, 2013) and are confronted with negative or even stigmatized stereotypes such as that of being “bean counters” (see, e.g. Jeacle, 2008). Data scientists, on the contrary, do not appear to suffer from such negative images; they rather receive quite a lot of “advance praise” in the discourse surrounding their occupation. This would seem to put data scientists in an advantageous position to “take over” (some of) the work of accountants. Moreover, prior work has found that data scientists position themselves as having a broad skill set rather than being technical experts only (Avnoon, 2021). Such an “omnivorous strategy to construct their identity” (p. 333) may even increase the competitive power of data scientists compared to other occupational groups.

And, yet, it may well be that data scientists themselves face challenges in building and maintaining their aspirational identity. After all, many organizations do not have much experience with data science and may not always offer a context in which data scientists can thrive. This, in turn, may countervail the putative benefits that the positive discourse around their occupation creates for data scientists and affect their positioning in the “information marketplace” within their organizations. Before delving into the empirical examination of these issues, we will elaborate in more detail on the theoretical perspective underpinning our analysis.

Identity and identity work

In our analysis of data scientists’ work experiences, we build upon the concepts of identity and identity work (see Brown (2015) for a synthesis). The notion of *identity* concerns the question “Who am I?” or “Who are we?” and hence relates to “the self as reflexively understood by the person” (Giddens, 1991, p. 53). In addition to personal identities (e.g. as a mother, a religious person, a woman, etc.), working people develop work-related identities, of which occupational identities are particularly salient. Such identities relate to the type of occupational group one belongs to (such as accountant, marketing manager, academic, etc.) as well as the tasks, responsibilities and values that typically go along with this occupation.

Like all identities, occupational identities are fundamentally social in nature, i.e. they are influenced by other people’s understandings of this occupation and the corresponding role expectations they communicate to the focal actor (Berger and Luckmann, 1966; Jenkins, 2014). Such role expectations exist at the workplace, where superiors, colleagues, customers or other actors communicate their expectations through their interactions with the focal

person (Bechky, 2011). Identities are, however, also influenced by a broader discourse that extends beyond the workplace (Watson, 2008). Educational institutions, professional associations, books, movies, news or advertisements are some of the channels that transmit images of what it means to be an accountant, an entrepreneur, an academic, etc. (Alvesson and Willmott, 2002; Watson, 2008). This discourse has a direct impact on one's sense of the self and an indirect one through its influence on other people in the workplace and their respective perceptions and expectations.

While broader discourses can trigger and unfold disciplinary mechanisms within groups and individuals in organizations (see, e.g. Kuhn, 2006), they do not simply determine identities, however. Rather, individuals can exercise a certain degree of agency and selectively endorse some elements of this discourse, while resisting others (Brown, 2015). Put differently, actors engage in *identity work*, defined as the “forming, repairing, maintaining, strengthening or revising [of] the constructions that are productive of a sense of coherence and distinctiveness” (Sveningsson and Alvesson, 2003, p. 1165). In doing so, they are often guided by pursuing some sense of aspirational (or ideal version of the) self (Thornborrow and Brown, 2009). This is particularly visible when the realization of such aspirational identity is challenged, such as through competition from other occupations (Collinson, 2006), demeaning work tasks (Morales and Lambert, 2013), advances in technology (Nelson and Irwin, 2014) or stigmatization (Ashforth et al., 2007). In these instances, individuals perform identity work to protect themselves against (potential) identity threats, i.e. any “experiences appraised as indicating potential harm to the value, meanings, or enactment of an identity” (Petriglieri, 2011, p. 644). However, identity work can also take place in a more tacit way and on an ongoing basis, as part of the routine (reflexive) monitoring of one's actions in the world (Giddens, 1984).

Identity work appears in a *narrative* form when people reflect about their identities, talk about them or write them down in the form of autobiographies or diaries, for instance (Watson, 2008). At the same time, it also happens through *interactions* with others, such as in meetings (see, e.g. McInnes and Corlett, 2012), in the giving of instructions or the issuing of complaints at the workplace (Down and Reveley, 2009). More generally, when actors craft their job by changing tasks and relations to other actors (Wrzesniewski and Dutton, 2001), then such job crafting often performs and symbolizes a particular identity, both for oneself and others (Goretzki and Messner, 2019). Indeed, both narratives and interactions should be understood as having an inward- and outward orientation (Watson, 2008). They hence (potentially) influence one's self-understanding as well as others' perceptions of and expectations towards oneself, i.e. the identity ascribed by others.

Applying the above-outlined theoretical apparatus to explore the identity of data scientists, our study focuses on how members of this “new generation” of data workers (see Vollmer, 2019) construct and enact a sense of the self. Exploring these issues, it is important to distinguish “*individual-level* occupational identity” (p. 873) from “the perceived ‘essence’ of the occupation as a whole” and hence what can be referred to as “*macro-level* occupational identity” (Murphy and Kreiner, 2020, p. 873). What seems particularly interesting in our setting is that – in contrast to established occupations – individual data scientists cannot draw on discursively entrenched “scripts” to craft their identity (see Watson, 2008). As stated by Murphy and Kreiner (2020), “[i]n emerging occupations, individuals are given very little prepackaged identity ‘content’ – for example, occupational values, legitimating ideologies, clear goals, tasks, and/or routines – to help them build their individual-level occupational identities” (p. 871). Thus, as a generally accepted understanding of the essence of the data science occupation might not be established yet, individual role incumbents would need to come up with “creative ways of crafting a sense of legitimacy around their identities” (p. 888) and – more generally – a sense of the self.

Exploring these issues, and in line with the above-mentioned, we can, however, expect their identity to be affected not only by their day-to-day experiences at the workplace but, to a certain extent, also by contemporary discourses about data science (see [Alvesson and Willmott, 2002](#)). A particularly interesting aspect hereby is whether, and if so how, data scientists endorse (or resist) aspects of the broader discourse around their role in their efforts to establish a particular sense of the self and enact it vis-à-vis others in the organization. For example, given their novelty in the market for managers' attention (see [Vollmer, 2019](#)), data scientists might face challenges in crafting their identity narratives that are related to the very legitimacy of their role in the organization. Accordingly, (shaping) their own and others' perception that their role is indeed "desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions [here the organization]" ([Suchman, 1995](#), p. 574) can be an important element in how a data scientist constructs an identity narrative. In doing so, they might be able to draw on the above-mentioned discourse in which their craft is often hyped and created as a fashionable practice (see [Madsen and Stenheim, 2016](#); [Saunders, 2013](#)) as a reference point or storyable item ([Down and Reveley, 2009](#)) for their identity narrative. However, such attempts might be undermined by any negative experiences they make in their day-to-day work and that challenge the aspirational identity narrative they construct (see [Carter and Sholler, 2016](#); [Goretzki and Messner, 2019](#)). Precisely how they do this and what identity-related challenges emerge in this process is however an empirical question that will be explored in detail in the following.

Fieldwork and research design

Data collection

Most of our interviewees were recruited via LinkedIn. We first searched the LinkedIn website for profiles of "data scientists" in Austria and Germany, two countries in which the cultural context is particularly familiar to us and where we could conduct most interviews in the language of our informants, which in general facilitates the interview process. We then sent more than 50 contact requests in which we briefly explained our research study. Sixteen people accepted to be interviewed. In addition, we drew upon the idea of snowball sampling and asked our interviewees at the end of each interview whether they could refer us to other data scientists (in their own or other organizations) and, in this way, managed to enrol two further interviewees into our study. Finally, seven interviewees were recruited based on the personal contacts of one of the authors. Reflecting upon our interview transcripts, we could not identify any significant difference in the way the interviews from these three sources developed. All interviewees were quite communicative and open about how they perceive and experience their role as data scientist and the corresponding challenges in their work.

We conducted most of our interviews via video chat (using Skype or Zoom). Three interviews were conducted in person since the travel distance easily allowed for this. All but one interviewee agreed to have the interview being recorded. In the remaining case, we took extensive notes which we then corroborated through email exchange with the interviewee. All interviews were conducted from 2018 to 2020.

Most of the interviewees (17) work for business organizations that use data science for internal purposes (e.g. for customer analyses, financial forecasts, predictive maintenance, etc.). These organizations cover different industries, such as retail, manufacturing, insurance, telecom, banking or transportation. Two interviewees work for public sector or non-governmental organizations that likewise use data science methods internally. Five interviewees work for firms for which data science is an important part of their product offering (e.g. commercial software providers). One interviewee works for a consulting firm that offers (among other things) consulting services in data science and related fields. Several of our interviewees also had prior working experience in other organizations and partly

reflected upon these experiences in the interviews. The tenure of our interviewees, at the time of our interviews, ranged between a few months and several years, with a mix of junior and senior profiles. The list of interviewees is provided in [Appendix](#).

In conducting our semi-structured interviews, we were guided by an interview protocol which covered *a priori* defined areas of interest. Overall, our objective in the interviews was to learn about how the data scientists made sense of their role in the organization, what specific characteristics and goals they attach to it and how they distinguish themselves from members of other occupational groups. We further inquired about their work experiences, with a particular focus on the challenges that they encounter in their work. We were also interested in understanding where these challenges originated and how data scientists would (try to) deal with them. The interviews in this sense aimed to create more fine-grained insights into what challenges data scientists encounter, why they encounter these challenges and how they would respond to them. Accordingly, we had *a priori* outlined categories, which we sought to explore, further develop and expand in the interviews through incorporating interesting topics emerging from our conversations with the data scientists.

Data analysis

We conducted our data analysis by carefully reading through all interview transcripts (and notes) and identifying passages that somehow related to the *a priori* outlined areas of interest, as described above. We coded each instance on an aggregate level (using, e.g. the code “challenges encountered”) as well as on a more fine-grained level, using open codes that would reflect, for instance, the particular type of challenge encountered ([Strauss and Corbin, 1998](#)).

Having assigned codes to each selected interview passage, we then went through all these codes again and tried to consolidate them. This implied that we merged different items into the same code if we felt that they described the same phenomenon. When analyzing our interview material in this way, we realized that crafting their identity narratives, data scientists would draw on three main narrative elements, which we refer to as “a scientific mindset”, “sophisticated data work” and “a problem-solving attitude”. Most of our interviewees would draw on these narrative elements when reflecting on their role as a data scientist.

In addition to the different narrative elements, we focused on the challenges that data scientists experience when trying to enact their identity in the organization. We realized that our interviewees related many challenges to expectations they felt managers would have towards them and that do not always align with how they see themselves. The managers’ expectations that the data scientists reflected upon in our interviews, in turn, often related to – from their perspective – problematic ideas about data science featuring in, for example, the media. We thus noticed that instead of regarding the discourse about data science as a useful reference point for their identity, our interviewees would be rather critical about it and at times even try to distance themselves from it. In a further step, we therefore examined more closely how data scientists would deal with identity-related challenges through some form of “identity work” ([Brown, 2015](#); [Sveningsson and Alvesson, 2003](#)).

Empirical findings

Data scientists’ identity narratives

When people talk about themselves and about the occupational group they belong to, they typically emphasize “central characteristics” of this occupation, which they regard as particularly salient for defining and distinguishing themselves from others ([Kärreman and Alvesson, 2001](#), p. 63). Analyzing our interviews with the data scientists, we find that their

identity narratives feature three main components. Taken together, these can be seen as building blocks for our interviewees' identities.

A *scientific mindset*. Most of our interviewees have an educational background in the sciences, holding degrees in physics, statistics or computer science, sometimes even a Ph.D. from one of these fields. This background also transpires in the way in which they characterize their work and talk about their identity as data scientists. In particular, several of our interviewees would emphasize the “scientific” dimension of their work, in terms of investigating data like a researcher would do. Our interviewees consider understanding scientific ways of working as crucial to perform the work of a data scientist. For example, when being asked how she and her team would develop ideas when dealing with data, interviewee 4 explains,

That's where our scientific backgrounds help a lot. So having people who [...] needed to solve different things in academia, look through different datasets ... learn different statistical methods and of course still ongoing reading, trying to keep up with new methodologies that are coming out. So, yeah, some of it is knowing how to construct a good A/B-test. You can learn in academia, knowing ... learning different statistical methods. (...) [Y]ou definitely have the underlying way to have a grasp on it and together with (...) people who do research, quite creative and inquisitive people, keep trying to question: “How else could we do this? What else could we do? What else is out there?” Try to come up with their own ideas, try to validate them. So it's definitely R&D, it is still a research work. (4)

Associating the work of a data scientist with proper “research work”, some of our interviewees explicitly mentioned their research background and formal training as essential for becoming a data scientist, as interviewee 11 explains,

I then did my Ph.D. dissertation, which certainly helped me to get this job. I also had an offer from [another country] and, in almost all job ads there, a Ph.D. is a precondition for a job in the area of data science. Even though this is not the case [in our country], it does have advantages to have [a Ph.D.].

Interviewees also emphasize the creative aspects of the data science process and how a scientific mindset and corresponding urge to explore things are important for a data scientist. Interviewee 21, for example, mentions that what he likes about his work is to find “*hidden treasures in the data*” and to develop “*models that are stronger and better*” (21). Interviewee 1 suggests that a data scientist “*needs to know a lot of tools [and] how they work*”, but also requires “*a lot of intuition, how to tackle a new problem. There are no rules, it's a little bit of an art ... there is a lot of improvisation [in] the end and this is only gained from experience*”. One can argue that the way data scientists talk about their work – not least the explorative aspects of it – resembles how researchers would characterize their job.

Some of our interviewees, in this sense, emphasize the need to stay up to date in terms of new methods or algorithms being developed in the research community. To do so, they would engage with the academic literature:

I guess what I am missing to say here in this entire story is that there is also the research side, which is actually finding new models and understanding how they work. So, if I know a model, yes, I can go take it and apply it. But if I don't know a model or how it works, I will not be able to tune the model, to adjust certain parameters to make the model work better. And most importantly as well, staying up to date, there is research done and papers released almost daily. And therefore, we in the team we want to stay at the cutting-edge technology, therefore we need to read papers and also to understand these papers. That's also why I find a lot of joy in my job and I am very happy that they give me the opportunity to actually do research and find new models that we can apply. (21)

Overall, it transpires from the interviews that – while occupying positions as staff members in business organizations – data scientists attach significance to the “scientist” part of their job title. For many, the work of academic researchers, which for many of our interviewees relates to what they did in their previous roles as pre- or postdoctoral researchers, thereby seems to

serve as a reference point to make sense of their work and role in the organization. The academic background that characterized most of our interviewees thus arguably forms a “lingering” element of their identity narrative (see Wittman, 2019) affecting how they make sense of their organizational role as data scientist. As interviewee 4 puts it succinctly: “I love the word ‘science’ in general. So I am very proud of that being in my title (laughing)” (4).

Sophisticated data work. Closely related to the “scientific mindset” stressed by our interviewees is another component of their identity narratives. Interviewees would emphasize that they are engaged in *sophisticated* forms of data work. In doing so, they would often distinguish between what they consider a “real” data scientist and other job profiles, particularly that of a data analyst, who in their view performs less advanced tasks.

Several of our interviewees spontaneously describe their job by distinguishing themselves from data analysts whom they portray as “*much more superficial in their understanding of statistics. They would understand the descriptive statistics but not necessarily modelling or advanced methodologies and time decision-analysis and experimentations and so on, or machine learning*” (4). Analysts are engaged in “*dashboarding, visualization and the like*”, which is “*also really important, but probably not that sophisticated from a mathematical point of view*” (16). Another interviewee suggests that “*analysts often have a good understanding of data and can very well do queries, create dashboards, and draw implications from these. But only very rarely do they build statistical models*” (3). Accordingly, the data scientists we interviewed see their role as going beyond what data analysts (can) do. As one interviewee puts it in a formulaic way, “*A data scientist is a data analyst plus machine learning models or even reinforcement learning or artificial intelligence*” (2).

We noticed in our interviews that, when talking about this identity component, data scientists would also distinguish themselves from others through the freedom they are given in performing their work. This is illustrated in the quote below by interviewee 13. Note how the interviewee draws on the above-mentioned scientific part of the identity narrative (“it’s more like a laboratory”) and subtly emphasizes the superiority he sees in his role (“we do things better” and “it’s more up-to-date”) vis-à-vis other data workers before stressing the flexibility and freedom he has as key distinguishing aspects:

There are a lot of departments that work with similar tools and similar methodology. The main difference is that we’re not bound so much, we’re more free to choose what we can do. [S]o it’s more like a laboratory, I would say, it’s like an advanced analytics team. [. . .] There is a difference and it’s a substantial difference, but how to describe it . . . I mean, it’s saying that we do things better is not a good definition. That it’s more up-to-date, it’s also not a good definition. These people are very smart - other departments - and they can also do work on a very high quality. I would say we’re not bound to many regulations. We have much more time to innovate and to think how to optimise. That’s the main difference, I think. We’re given much more flexibility and much more time. We don’t have to deliver tomorrow or today, we have to be creative [. . .]. So, we kind of have more freedom. That’s the main difference, I would say. (13)

The self-understanding of someone who engages in advanced modelling is influenced not only by the educational background of the interviewees but also by their perception of popular descriptions of data science in the media or in organizational discourse. Interviewees tell us that the term “data science” is used in many ways, resulting in quite some ambiguity around it. Interviewee 4, for instance, explains that she knows “*people who think that if you are not doing machine learning you don’t do data science and people who think that if you are doing machine learning you are a computer scientist*”. Similarly, another interviewee suggests,

The spectrum of people who call themselves data scientists is really broad. You have the hardcore engineer who is only interested in how to merge data in the most efficient way. And on the other side, you have analysts who can barely deal with Excel or Tableau or whatever. (. . .). And to compare these two extremes, both of which call themselves data scientists, is very difficult. (15)

Interviewee 16 stresses in this context that “*there are too many different things subsumed*” under the notion of data scientist, which creates a lot of confusion.

Our interviewees suggest that ambiguity is also visible in job advertisements. Interviewee 3 feels that job advertisements are “*often ludicrous*”. When they talk about data scientists, they would range from “*people who only create dashboards to people who only do programming, and everything in between*”. Note how the data scientists in the illustrative quotes above emphasize that what they sometimes see in job ads would not match their understanding of a data scientist and rather correspond to other – less advanced – roles. Interviewee 15 mentions in this regard that one would often find job advertisement for data scientists which, when more closely reading through the tasks and requirements, would really describe the work of a “*data engineer*”.

The data scientists we have interviewed, to some extent, accept the ambiguity surrounding their job title as being related to the nascent state of their occupation. Some mentioned that they expect that, over time, more defined or specialized roles would crystalize, potentially leading to different types of data scientists. Yet, the above-quoted statements also show that interviewees are concerned about such ambiguity and, in response to it, re-emphasize what, in their eyes, a data scientist is or should mainly be doing, namely engaging in sophisticated forms of working with data and modelling. Thus, while interviewees often made statements about the level of sophistication of their work in factual terms (indicated by using the present tense), one should interpret those accounts as having an aspirational element (Thornborrow and Brown, 2009) emphasizing an arguably “elite” status of their emerging occupation among data workers (see Avnoon, 2021). In other words, the level of sophistication of their work can be understood as a desirable characteristic that data scientists would refer to in order to distinguish themselves from other “subordinate” data workers.

A “*problem-solving*” attitude. While the identity narratives of our interviewees revolve primarily around their scientific-technical expertise (as described above), this is not the only type of skill or mindset that is mentioned. Several of our interviewees also identify with the need to be oriented towards the “needs” of managers and help solve real-life problems. This includes data scientists’ ability to communicate effectively with managers who do not have the same level of technical expertise. Our interviewees feel that learning how to effectively communicate with managers is a crucial but often underdeveloped skill among data scientists. In other words, they see this as an aspirational element and something a good data scientist should be able to do. Interviewee 4 refers to it as an important “soft skill” that data scientists need to – but often fail to – develop:

I think still . . . in my experience (. . .) I think the biggest challenges in data sciences remain the more soft-skill-driven issues than the technical ones. I think there are a lot of people who have great technical skills and there are a lot of things we can do with methodologies. We can’t do everything, but problem-solving that requires more sophisticated codes or different ways to break down the numbers are in the natural sphere of most people who do data science. (. . .) A lot of the people who end up doing this work are more interested in figuring stuff out than maybe in how to be effective in an organization, in soft-skills and communication and business insights and so on. And I think for most data scientists (. . .), they are weaker on that side. (4)

The perceived importance of understanding organizational problems and effectively communicating with managers is sometimes taken as a trigger to distance oneself from an overly “technical” perspective on data scientists’ work, as in the following example:

Again, it is a complete mistake to take data science as IT or technological, algorithm, whatever, it doesn’t matter. I mean it’s about influencing, it’s about (. . .) a good data scientist is an influencer, it’s [sic] someone who understands the basics of the data and the potential of the algorithms and someone who is able to have some decision in the company and explain the roadmap to implement the strategy. (6)

Along the same lines, another interviewee suggests,

For me, a good data scientist is someone who does not only know many types of analysis and the technical possibilities, but who can also put oneself into the practical context, to find solutions for a particular area or generate new ideas. And who can present these results in a way which also a non-data-scientist can understand. (18)

Interviewees in this sense also relate their reasoning about what characterizes a data scientist and distinguishes them from other roles to a profound understanding of “why” they are performing their work. As the exemplary quote below from interviewee 11 illustrates, data scientists thereby refer to the practical relevance of their work. In that sense, they do not only distinguish themselves from other data workers through the sophisticated nature of their technical work (see above) but also through their orientation towards helping management solve problems:

Data science is characterised by the fact that the people who do it not only do the analysis of the data, but above all know why they are doing it. That is often not the case with data analysis. Because there, data sets are simply taken and analysed according to statistical methods. In data science, you go one step further to say that I have a concrete application, I have a business interest in analysing this data, or [. . .] I simply want to solve a concrete problem. And the person who wants to solve this problem must have an idea of both the subject matter and the analysis, and that is the big difference to a pure data analyst, who only knows how to evaluate a data set. [. . .] But without knowing now why you do it. You simply evaluate the data, then transfer it and someone else takes over the evaluation and interpretation of the results. And that’s the big difference as a data scientist, because you’re also responsible for interpreting the data. And I would go so far as to say that the essential part of the data science task is, because the analysis, you can put any statistician, but what’s important is that you say you also know that you’re drawing the right conclusions from it. And you are able to sell these conclusions to your stakeholders.

When reflecting on this component of their identity, some interviewees would refer to their own education or to the “typical training” of data scientists more generally. Interviewee 2, for instance, has both a higher degree in management and in data science, respectively, and suggests that the former has helped him communicate with managers in the organization and eventually distinguish himself from other technical experts:

I realize this when an engineer or statistician talks to a marketing person and that’s too technical for the marketing person, or when the marketing person, in turn, is too marketing-specific for the engineer. There are often these sorts of communication problems. The mindset is a different one, and in this respect, I have a small advantage.

Another interviewee, when being asked about the “ideal” skill set of a data scientist, suggests that, in addition to modelling/programming, data scientists should be exposed to different kinds of organizational problems in their education. This would enable future data scientists to learn how to help managers in solving problems.

And I think they should also know business cases, I think they are doing it quite well already because we just had an applicant from some university (. . .), they learn the theory, but they also learn their cases which are a little bit clean and prepared for the university, but they train it with pseudo real data and pseudo real problems. I think this is really good, I think this is really good. (. . .) [If you solved kind of business problems, you know ‘ah that’s what you do with this’ and then you are more creative to find new solutions within the company. ‘Oh, we could do that’, or things that they [i.e. managers] didn’t think [of]. (1)

Overall, and as we will see in more detail below, the “problem-solving attitude” component of the identity narratives is particularly triggered (and reproduced) by specific work experiences of the data scientists, which make them realize the importance of this component over and above their technical skills. However, like the level of sophistication of

their work compared to other data workers, interviewees presented the features connected to the “problem-solving attitude” – especially the knowledge about the operational departments they are working with – as aspirations that not all data scientists would (be able to) live up to.

Data scientists’ work experiences and identity work

In the next step, we discuss how the data scientists’ work experiences influence their sense of identity. In particular, we can see how enacting their aspired identity as a problem-solving-oriented scientist engaging in sophisticated analysis and modelling can become challenging when other actors in the organization do not confirm such a sense of the self. We can see how this leads to identity work on the part of the data scientists in the form of affirming their aspired identity or of expanding it in some way.

Ambiguity around the data scientist’s role and space for crafting an identity. Data scientists revealed in our interviews that they feel that managers do not always form clear-cut expectations regarding their work. They perceive ambiguity when it comes to job descriptions or job profiles and the corresponding expectations among managers and other internal stakeholders regarding what precisely a data scientist should be doing. Interviewees reflected in this context that the unclear expectations they see themselves confronted with mirror a broader concern with ambiguity in the contemporary data science discourse. While challenging at times, such ambiguity also provides incumbents of a data scientist role with leeway to actively position themselves as suitable for the role. One interviewee reflects in this context on how he obtained his current position:

[T]he fact that I had a post-doc in machine learning opened me the door. Because on the company level, they don’t know in detail what I was doing and they just knew that [I knew about] machine learning. ‘Ah, you must know the[se] things’. They didn’t have a clear idea of what they knew and what they didn’t. And I thought that I was capable of doing the rest even though I had no experience because I had learned it in online courses, for instance, and I knew I can do that. And that’s why they gave me the position. (1)

Vague job descriptions then also allow (and require) data scientists to actively craft their job in line with their own idea of what it means to be a data scientist. Interviewee 23 offers an illustrative account of such activities when being asked about his job description:

... the job description was actually rather vague [...] with certain buzzwords [...]. Artificial intelligence, machine learning, and so on. Definitely. But the pure data science activity, which I really experience and practice in my everyday job, differs somewhat and is also designed in such a way that I have it in my own hands and can ultimately redefine it every day. [...] Therefore, this is ultimately an activity that I believe is still being re-formed over and over again and that has to find new ways and ultimately has to fit in where the need is. (23)

To say that one’s activity “is still being reformed over and over again” reflects the ongoing process of working upon one’s occupational identity. Such identity work (Sveningsson and Alvesson, 2003) can be observed in many contexts, but is particularly important when the organization provides no or only vague “role scripts” to the employee, as in this case. The main problem, in our context, does not seem to be that employers *per se* fail to produce clear role descriptions. There is no generally accepted view (yet) on what the “essence” of the data scientist occupation is that organizations could draw on when crafting job ads or descriptions (see Murphy and Kreiner, 2020). In many organizations, the data scientist is still a novel role occupied by only a few or even just one person. Indeed, data scientists need to, themselves, hold or develop an understanding of what characterizes (or should characterize) their position in the organization, how their role is distinct from others and what goals to strive for (see Kärreman and Alvesson, 2001; Vaivio *et al.*, 2021). Data scientists seem, in turn, however, not to experience too strong restraints through (external) scripts imposed on what can be referred

to as their “inward-facing’ identity work” (Watson, 2009). This eventually enables them to incorporate their individual aspirations and understanding of “this is what I want to be like’ and ‘that is what I don’t want to be like” (Watson, 2009, p. 432) into their identity narratives.

Dealing with work assignments that are not in line with the aspired identity. While the above-described ambiguity that data scientists perceive around their role can be seen as providing them with an opportunity to craft their role in line with their aspirations, our interviewees also experience some downsides. For example, the lack of clearly defined “scripts” can render data scientists’ job challenging to the extent that they constantly need to make sense of what managers expect from them. Interviewee 9, for example, works for a retail firm and explains that managers wanted to initiate a so-called “shopping basket analysis”. But “they did not have a fixed idea of what this could be. They just called it ‘shopping basket analysis by one mouse-click’ (. . .) No idea what exactly a ‘shopping basket analysis’ could look like”. Interviewee 3 states in this context that it is very difficult to do a good job “if [internal] customers come without a specific question”, because “if you don’t know the business area, then you never identify the things that the customer is really interested in”. Unclear expectations can, in this sense, make it more difficult for data scientists to enact the “problem-solving” element of their identity narrative.

An even bigger concern on the data scientists’ part is that managers might form expectations which are at odds with their aspirational identity (Morales and Lambert, 2013; Thornborrow and Brown, 2009) and the core characteristics they attach to their role (Kärreman and Alvesson, 2001). Several interviewees report in this regard that they (at times) would be confronted with tasks that they do not see as in line with their aspired identity. This is specifically so when they are asked to carry out activities that they would see as belonging to the remit of other (in their view) less sophisticated data workers. Interviewee 9, for instance, recounts a meeting with the accounting department to discuss topics that the accountant had collected and wanted the data scientists to work on. But when they went through the list of topics, the data scientists realized that “about 80% of them were about BI [Business Intelligence]” and “nothing was really an analytics topic”. Drawing on his experience, the interviewee feels that there is indeed a challenge within the organization to clarify “what distinguishes a data science topic from standard reporting”. Similarly, interviewee 11 remembers a situation where he and his team were approached by another team, but that request was quickly “outsourced to the reporting department” as they felt that it was not within their area of responsibility. Such examples illustrate that data scientists often see themselves confronted with expectations that do not match their understanding of who they are and what they can and should do.

Seeing themselves consistently confronted with expectations that do not align with how they make sense of their role appears particularly frustrating for those data scientists who experience a lack of freedom for job crafting. As described above, such freedom features prominently in data scientists’ identity narrative as an important enabler for their “scientific” and “sophisticated” “problem-solving” work. Talking about his previous job, interviewee 16 in this context remembers his boss asking him to visualize data in a dashboard with different charts. “I was a bit annoyed by this, did not like it. And so I would say that this is not my strength, or it’s not data science, not my role”. Similarly, another interviewee reflects on how he was asked to carry out tasks that he did not see as part of a data scientist’s responsibility: “Sometimes it is only business analytics. And then I try to refuse this, since I want to focus on machine learning, in terms of my career and so on. But sometimes this does not work and sometimes it also belongs together and then I have to do it” (2). Here, the data scientist would enact his aspired identity of a sophisticated data modeller and, against this background, respond to the manager’s request. Data scientists seem to accept doing work that is not in line with their aspirations merely within certain boundaries, namely only if it remains the exception rather than the rule. As interviewee 23 suggests, accepting those tasks as a rule

would lead to a kind of “estrangement” from their aspired identity and potentially threaten it over time:

Pure dashboarding or building dashboards is of course not a data science task, and yes, I sometimes like to do it, but, as I said, if it were one hundred percent, then I wouldn't be a data scientist. [...] it always depends on whether you will be labelled with it and whether you will then be tempted to only carry out such activities. [...] As I said, I see it more like this: as long as it doesn't get out of hand, it's not a problem for me [...] but if it were to be too much, then I would also say, sorry, I am not a business analyst. I can do that when there is an emergency or when it is desired, but not one hundred percent. (23)

Our interviews indicate that a broader role encompassing tasks that data scientists would not consider as part of their occupational core is more likely when an employee works alone in that role within their organization or department. That is, without being embedded in a larger team of data workers. The bigger the team, the more specialized job profiles tend to become.

Our interviews indicate that a persistent mismatch between their self-understanding and aspirations, on the one hand, and the work tasks assigned to them by their internal stakeholders, on the other, can even motivate data scientists to use an exit strategy and change their employer. As interviewee 21 explains, he enjoyed his first job for the first two years, but then: “*you start realizing again, ‘OK, I am still not doing what I was supposed to be doing here or what I wanted to do here, so I continue and look for it [elsewhere].’*” Changing positions thus sometimes appears as the only reasonable alternative for data scientists to protect their preferred occupational identity.

Facing unrealistic expectations by managers. Our interviewees furthermore reflected on work experiences where managers would form what they perceived as unrealistic expectations regarding the output or process of data scientists' work. One of our interviewees lamented in this context that managers would often see the data scientist as a kind of “*MacGyver*” [1] who can easily address all kinds of difficult problems. Such expectations can challenge their aspired identity insofar as they may lead to disappointment among managers who might ultimately start regarding data science as “not keeping its promises” and as a sheer fashion that is expected to pass away eventually. Interviewee 5 puts this succinctly when suggesting that there are currently only extreme opinions about data science: “*What data scientists do is either seen as superfluous and a hype that will pass away, or data science is regarded as a sort of *deus ex machina* which can solve everything*”. Both positions arguably put data scientists in a challenging position in which they need to provide ongoing evidence of their importance for the organization.

A key challenge that our interviewees mention in this context is that the data science process is a “black box” to most managers and that much of the work they carry out is invisible to others. As a result, managers may develop expectations that are misaligned with the work realities that data scientists experience and the limitations they see in their craft. Interviewee 8 explains in this context that managers would often say: “*Well, the data are available anyway, you can do [the analysis] straight away*”, without understanding the effort it takes to aggregate and prepare the data and to make sure that this happens without incurring any errors. Similarly, interviewee 1 states that some managers “*just think we have to collect data and that's the only thing that matters. We just collect it and then analyze it. Like there is a black box that does it by itself. But it has to come in the right form, so lots of time is invested in that.*” Sharing a similar view, interviewee 21 does “*not think people realize how much time I am spending on just cleaning the data*”. Managers would sometimes even get impatient expressing that data scientists “*have been working on this for three weeks and there is still no result, so what are they actually doing the whole day?*” Interviewees mention that the main problem is that managers lack knowledge to appreciate the underlying task “*of making sure that the [data] quality is right to be able to draw the right conclusions*” (3).

We can see in these examples how, from the interviewees' perspective, managers' ignorance about the details of the data science process can lead to expectations which put undue pressure on them, and how this eventually presents a potential challenge for their occupational identity. Other interviewees offer further examples of cases where they felt managers had unrealistic expectations because they lacked knowledge about the steps that a data scientist must go through to produce meaningful results. Interviewee 11 explains that there was a new source of data in the organization, and before the data scientists had even seen the data, "*salespeople already wanted to get started and use the data to sell products. So we told them: 'That's all well and good, but we can't do this, we first have to look at it and this will take us two months. Nothing will happen before that'*". Having made similar experiences, interviewee 16 reflects on a recent case in which a sales-related key performance indicators (KPIs) had been declining for several months and suddenly "*the CEO panics and wants to know the reason for the decline within three days*" – a timeframe that the interviewee believes is not in line with the realities of data science work.

Our interviewees also provide several examples of managers having unclear or unrealistic expectations regarding the *outcomes* of data science projects and eventually what data scientists can deliver. Interviewee 10 explains in this respect that the "*whole machine learning hype*" is "*somewhat damaging*" to his efforts in the organization because managers would sometimes have "*obscure ideas regarding what you can do with models*". He thus regards it as his responsibility to "*explain to them what is possible and what is not*". Similarly, another interviewee feels that "*suddenly a lot of people talk about this topic [data science]. Unfortunately, also many people who have no clue about it and who then initiate completely unrealistic things which can only fail*" (12). He also expresses his wish that "*top management should undergo a training in data science*" so that they know what is possible and what is not. Similarly, interviewee 17 talks about "*many misinterpretations and wrong assumptions*" that circulate regarding the possibilities of data science. In particular, he sees little awareness for the importance of having the "*right data in the right quality*" to arrive at meaningful models. Interviewee 6 recounts that managers "*thought that data science would bring magic*", ignoring all the practical challenges and limitations that a data science project would involve. He complains that managers in his firm would "*just give you the data and they expect you to do something out of it and it's not realistic*". Talking about a particular example, he explains that he "*told them very soon that the project was not possible, the goals are not very well defined and unless we redefine our goals to something . . . more attainable in the short-term, we are going to waste our time doing research and not coming out with any product*" (6). Projects that are "doomed to fail" due to unrealistic expectations are hence problematic for data scientists as they may undermine their aspired identity as "scientifically minded problem-solvers".

The examples presented above already illustrate how data scientists try to manage expectations regarding both the process and outcomes of data science projects by communicating extensively with managers. This involves attempts to manage the expectations managers have regarding the outputs of a project and make them acknowledge the limitations of data science. This form of engagement seems important for data scientists to counteract not only the ambiguity but also the hype that surrounds their role and renders their work challenging. Interviewee 11 puts this succinctly when suggesting that "*the first thing to dispel is the idea that we can do everything and that the results come in the next day*". Similarly, interviewee 17 explains that when managers suggest an analysis that is not feasible, then he would "*communicate, short and sweet, that what they want is not possible. And I will not try to show things that are simply not to be found in the data*". Another data scientist confirms that "*it is important, from the outset to communicate realistic expectations regarding what a model can deliver and what it can't*" (14). And interviewee 21 even emphasizes that "*[f]irst of all, when someone requests something, I always push back, because I know I will need the time*".

Such forms of managing the expectations of internal stakeholders do not only facilitate the work of data scientists from a pragmatic perspective (e.g. in terms of having enough time to finish a project, etc.) but on a more substantial level constitute an important form of outward-facing identity work (Watson, 2009). Indeed, the way data scientists respond to their work experiences affirms the key features of their identity narrative outlined above. By pushing back managers' unrealistic demands, data scientists both enact and protect their identity as "scientists" who perform sophisticated work that requires going thoroughly through different steps to ensure a credible outcome of the project that can provide solutions to a concrete problem. One can argue that, in doing so, they also try to establish an understanding among managers that, as with any form of research, data science projects feature a good amount of uncertainty about their outcomes.

At the same time, managing expectations is about emphasizing those parts of the data science process which are not particularly sophisticated but crucial to prepare for the technically advanced parts of their work. That is, preparing and cleaning data. The fact that most of their time is allocated to these tasks stands in some contrast to their idea of carrying out "sophisticated data work" on which they draw to distinguish themselves from others. This contrast becomes particularly visible when our interviewees reflect on the "typical" data science project and what they spend their time on there. They would speak of the "data modelling" stage as the one that they enjoy the most and where they can really leverage their skills. Interviewees thereby emphasized that this stage would ultimately also make a difference in terms of the insights gained for the business. At the same time, several interviewees stress that this stage is only a small part of their work and that they would spend more time on rather mundane preparatory tasks. As one data scientist puts it aptly, "*There are many people who say that the real work of data scientists is modelling. Yes, this is part of it, but, overall, it accounts for perhaps 40% [2]. The rest is about organization and data preparation*" (11). Emphasizing this part of their work becomes particularly necessary for data scientists when managers form unrealistic expectations about the process of data science.

Taken together, our empirical material demonstrates how work experiences lead data scientists to constantly reaffirm their identity as "researchers". Doing so, they would try to uphold a narrative of performing sophisticated work while emphasizing those parts of their aspired identity which are grounded in solid and indeed somewhat mundane but crucial forms of data work.

Having difficulties in understanding each other. Remember that we had identified having a "problem-solving attitude" as one recurring component of data scientists' identity narratives. This part of their aspirational identity was particularly fueled (and reproduced) by experiences made in daily organizational life. In particular, our interviewees would make frequent reference to challenges that they encountered in communicating with managers and understanding each other's perspectives. Interviewees would stress in this context that such misunderstandings would already start with the definition of the problem to be solved and the data to be used, as explained by interviewee 1:

I would say that one of the most difficult parts is to get on the same page with the people who call me, telling me there is a problem. There is a lot of time invested in understanding each other because they come without really knowing how to even start. They say, (...) 'We have data, we think there is a solution out there'. And there is a lot of time invested in understanding data, what is this thing you are giving me. How did you label data? A lot of time is invested there. (...) The problem is to get on the same page with them and to communicate what we want and the way we want it, the way we need it.

Similarly, interviewee 3 explains that

... it may be the case that [an idea] does not work out, because I did not fully understand the problem or because the data are insufficient (...). Then I either try other things or I talk to the [operational

managers] and tell them: 'It does not work, I can't answer the question with these data. Do you have any other ideas?' Or I do have an idea and I check with them whether we should go in that direction.

Difficulties in creating a joint understanding with operational managers regarding what the problem to be solved is (about) can, in this sense, be seen as an issue for data scientists for it renders enacting their aspired identity as problem-solvers more difficult. Data scientists hence see communicating frequently with their internal stakeholders as crucial for addressing such challenges of misunderstanding and ensuring that they can perform their work in a way that managers would appreciate. While this can often solve an issue, some of our interviewees would ideally want their internal counterparts to be more tech-savvy as this would make conversations easier in their view:

I mean you need to have someone who understands. I mean you need to have some technical people. The CEO or the CFO, (. . .) these guys they have to know what it is about. You have to have some technical understanding, you have to have some technical people that you can talk with, otherwise it's just, they get into lots of bullsh*, lots of confusion in the air and they don't get the right decisions, they have to be technical persons, ok. They have to have some level of technical understanding, otherwise they take . . . they don't, they are not going to take good decisions. (6)

To the extent that such expertise is limited in the organization, data scientists feel a sense of responsibility to "educate" managers on what can (not) be done with data science, as interviewee 4 explains,

And there are probably areas or individuals that have less of that experience and with them it's more of a . . . there is more probably educational work as well, which is part of my work in the organization as well to actually make us more data-driven and to influence what people can do and want to do with that. So part of making this a data-driven company and training the culture of data. That is definitely part of my priorities and my department in general. So, with those there is more of an educational part, where we don't just come and say: "Ok, let's do this." It is more like: "Here is why we should do this!"

We can see here how the interviewee defines "educational work" as belonging to her duty as data scientist. The extent to which individual data scientists identify with such responsibilities appears to depend on the size of the data science team though. As interviewee 1 explains, he is more focused on the operational data science work, while his boss (who acts as "project manager" and liaison to the data scientist team's internal stakeholders) would be the one who proactively reaches out to other departments:

There is a project manager who knows (. . .) the company very well, all departments, the people you should ask, and he is also in charge in self-promoting our department. In order for people to know, there is this team, we could ask for some project with them. [. . .] Personally, the departments or people in charge need to think "ah, we didn't know but this can be optimized". Because there is this new thing called data science, they can solve these kinds of problems, or they can improve this process we have never thought about it.

The quote above further illustrates that data scientists' educational work vis-à-vis their internal stakeholders has an "internal marketing" perspective. In other words, in communicating what data science is about, data scientists try to promote their craft and the value it can generate for the organization. Thus, convincing others of the value they can generate, they try to attract more internal projects from their internal clients.

Importantly, data scientists do not only see the need for educational work vis-à-vis managers but also acknowledge that they must learn how to communicate more effectively. Some of our interviewees in this context recognize the need to speak in the language of their addressees. Interviewee 15, for instance, explains that "*when [the data scientist] starts to use technical terms or concepts which only [s]he knows, then it usually becomes problematic*". Rather, what is necessary is to "*communicate through gestures, you have to visualize a lot,*

that's really important, you have to work a lot with examples". Interviewee 17 confirms that when leaving the university, he was quite focused on technical questions and research and, at the beginning of his job, *"used a lot of technical terms"*, with the effect that he lost everyone after three days. *"You have to talk less about the methods because they are important for me but not for the others and focus on the questions and themes that could be of interest to them. And if you start in this way, then you can get into a lively discussion"*. Interviewee 8 confirms that speaking the same language is crucial, as *"simply because of language it can come to a lot of misunderstandings"*. Interviewee 5 tells us in this respect that the first thing he does when joining a new company is to learn the language of the people there. *"The challenge is to learn the words, the language of the industry and the language of the department and then to give an answer in that language. It's like a foreign language"*.

What these examples highlight is how experiences in working with other people in an organization increase data scientists' identification with a job profile that goes beyond the technical core of data science work. Rather, they acknowledge the importance of understanding the perspectives of managers and other internal stakeholders, speaking their language, and being able to translate their technical expertise into that language. This can be interpreted as a form of identity work where particular skills and competencies are emphasized as being important for successfully carrying out one's work within the organization (Goretzki and Messner, 2019; Ibarra and Barbulescu, 2010). As mentioned above, data scientists hereby also reflect on their own skills and emphasize the importance of thinking one's way into how managers reason and talk to better understand the problems they want data scientists to work on. This eventually leads to a form of self-disciplining where data scientists internalize the importance of the non-technical aspects of their work and act on themselves to learn and be able to adopt their internal stakeholders' "language" and view to perform their role in a way that would facilitate positive affirmation. However, the data scientists' reflection that challenges often arise around different languages and "thought worlds" (Dougherty, 1992) in the organization indicates that the "problem-solving attitude" component in their aspired identity does not easily materialize in their daily work.

Discussion

Data scientists' critical distance to the discourse around their occupation

This paper aims to improve our understanding of the workplace experiences of data scientists and especially how members of this nascent technical occupation (Avnoon, 2021) define and work upon their occupational identity. We demonstrate that data scientists draw upon three main components or "identity pegs" (Goffman, 1963, p. 57) when constructing their identity narratives. These are the "scientific mindset" that they adopt in their work; the "sophisticated forms of data work" that they carry out; and the "problem-solving attitude", and corresponding orientation towards working together with managers, that a data scientist should have.

Elaborating on these identity components, we show how data scientists draw not only on their educational background and work experiences but also on their perception of the data science discourse. Reflecting on the discourse in media and organizations, they stress both the hype and ambiguity around their work and occupation. While one could assume that data scientists would utilize positive features of the discourse to bolster their occupational status, we demonstrate that they tend to distance themselves from what they perceive as a too broad, ambiguous or even misguided understanding of data science. We reason that the hype and ambiguity that data scientists experience can be linked to the "fashionable" status (Abrahamson, 1991, 1996) of their occupation in the contemporary digitalization discourse, where both being a data scientist and building data science teams in organizations is often presented as being in vogue. Fashions are often characterized by ambiguity insofar as when

many people start referring to a hyped practice, they likely do so in different ways (Giroux, 2006). While this helps the practice spread even further, it also leads to ambiguity. Data science seems to be a case in point: While more and more people refer to data science and how important it is for organizations, they tend to associate very different things with this (and related) term(s).

Hence, we suggest that the fashionable nature of data science can, somewhat paradoxically, threaten the realization of data scientists' aspirational identities. Indeed, our study demonstrates how data scientists experience that the above-mentioned features of the data science discourse frequently translate into the workplace and render it challenging for them to enact their identity narratives. The data scientists we interviewed appeared to have a relatively clear understanding of the "essence" of their role (see Murphy and Kreiner, 2020). Hence, the novelty of their occupation does not seem to present the main challenge for them. Lacking generally established role scripts (see Watson, 2008), data scientists, for example, draw on their research background ("scientific mindset") and education as a reference in their identity work. This argument can be substantiated when looking at our interviewees' rather undivided identity narratives and the core characteristics they associate their role with (Kärreman and Alvesson, 2001). The main challenge they face relates to how *others* (i.e. managers and other internal stakeholders) make sense of what a data scientist is and what role they (can and should) play in the organization. The fashionable nature of their role and craft can eventually result in expectations that data scientists regard as too low (i.e. tasks that they feel do not require their "scientific mindset" and are not "sophisticated" enough to land on the data scientist's desk) or unrealistically high (i.e. unsolvable assignments condemned to failure). Both manifestations of managerial expectations confront data scientists with potential identity threats in the form of either demeaning tasks or unsolvable problems making it difficult to receive positive confirmation for their work.

Data scientists provide a fascinating case in this respect since "fashionable" occupations experiencing the hype and ambiguity around their role on the micro-level are rarely the focus of studies that are concerned with identity struggles. Much work in this area is about occupations that are perceived as tainted, dirty or stigmatized (Ashforth *et al.*, 2007) or that are being challenged by other occupations or through managerial or technological change (Kahl *et al.*, 2016; Nelson and Irwin, 2014; Reay *et al.*, 2017). Data scientists who are stylized as, for example, the "missing piece of the big data puzzle" (Carillo, 2017, p. 607) or "the latest idealized subject" in a data-driven world (Gehl, 2015, p. 420) seem to differ significantly from previously studied occupations experiencing identity struggles, not least accountants.

Comparing the workplace experiences of data scientists with those of accountants indeed allows for interesting observations. Accountants often struggle with negative stereotypes (see Jeacle, 2008) or face threats of becoming rationalized through computerization (see Frey and Osborne, 2017) or a strive towards productivity (Goretzki and Pfister, 2022). Despite "counter-discursive" attempts by, for example, accounting associations to reposition the accounting profession as more exciting and strategically important (see, e.g. Baldvinsdottir *et al.*, 2009; Goretzki *et al.*, 2022), negative stereotypes (see Jeacle, 2008) seem to constantly linger in the background, making it often difficult for accountants to enact aspirational identities like the "business partner" (Goretzki and Messner, 2019; Goretzki and Pfister, 2022; Morales and Lambert, 2013). Data scientists, in contrast, enjoy a rather positive discursive framing as "sexy" and important. Reputation thus seems to precede data scientists, sometimes even resulting in expectations that they consider unrealistically high. Legitimization and justification of their role or working against a negative image are thus not key concerns for data scientists, demonstrating a stark contrast to what accountants often experience (see Goretzki *et al.*, 2022; Messner *et al.*, 2008).

And yet, our analysis demonstrates how the fashionable framing of their occupation creates its own identity challenges for data scientists. It is thus not only digital technologies

per se (Vaast and Pinsonneault, 2021) but also the hype and ambiguity in the discourse about such technologies that have ambivalent consequences for data scientists. Our study thus complements previous research on data scientists (Avnoon, 2021; Vaast and Pinsonneault, 2021) by exhibiting how conditions that one might assume would support their aspirational identity can contribute to its fragility. As such, our findings also speak to the management fashion literature that has mainly focused on factors affecting the diffusion, implementation and rejection of new tools, techniques or practices (see, e.g. Abrahamson, 1991, 1996; Baskerville and Myers, 2009; David and Strang, 2006; Kieser, 1997), without discussing the effects of such fashionable practices on employees' identities. It will be interesting to observe, in this context, how the general image of this nascent occupation will develop in the future when managers start to compare hopes and promises associated with data science with the outcomes they see (or miss) on the organizational level.

Managing expectations as a vital component of data scientists' "paradoxical" identity work

As elaborated above, a concern raised by our interviewees is that the hype and ambiguity around data science often spill over to the expectations that managers develop vis-à-vis the role of data scientists and eventually the work they assign to them. Data scientists in this sense lament that managers sometimes construe their role in terms that are not in line with their sense of the self. Prior research shows that, when faced with a gap between a preferred identity and assigned tasks, actors would either try to get rid of or reinterpret these tasks (Morales and Lambert, 2013) or would, alternatively, attempt to realign their sense of self to make it fit these tasks (Pratt *et al.*, 2006). Some of our interviewees indeed mentioned that they would seek to avoid tasks which they consider belonging to positions that they perceive as technically inferior. In other situations, they would find themselves struggling with assignments that, in their view, are rooted in managers' unrealistically high expectations regarding the possibilities of data science or a misguided understanding of the data science process (Zbaracki, 1998).

Our analysis thus indicates that data scientists try to balance two (partly competing) forms of identity work. On the one hand, reacting to the ambiguity in their job title, data scientists accentuate the "sophisticated" parts of their work that they consider as important pillars of their aspired identity (especially data modelling). By doing so, they present themselves as an "elite group" in the ecosystem of data workers. On the other hand, they try to create awareness among managers for the less prestigious parts of their job. In emphasizing how much time they must allocate to mundane tasks like data collection, preparation and cleaning, data scientists engage in what can be called "de-hyping". Trying to tone down managers' expectations, they highlight that they are not "data wizards" performing "magic" and stress the intricacies and limitations of their craft. They eventually try to draw a picture of their role that – though less "heroic" – they regard as more representative. This situation creates a kind of identity work paradox for data scientists that can be seen as a specific feature of a fashionable occupation facing both ambiguity and hype around their work: While emphasizing the prestigious parts of their work to present themselves as "elite" data workers, data scientists, at the same time, stress the less prestigious aspects and limitations of their work to respond to managers' unrealistic expectations. The fact that data science is still a new occupation for which a generally accepted understanding of its essence (see Murphy and Kreimer, 2020) has not been established yet arguably intensifies this issue.

As we demonstrate in this paper, the expectations that data scientists see themselves confronted with create specific challenges that make them deal with the more and less prestigious aspects of their work in ways that differ from what we know about accountants. Due to the fashionable nature of their nascent occupation, data scientists do not experience the need to (pro-)actively create a positive image of being important and exciting (cf. Goretzki and Messner, 2019). They also do not consider it necessary to hide tasks that are not in line

with their aspired identity from others (cf. [Morales and Lambert, 2013](#)). Quite the contrary, data scientists deliberately choose to render less prestigious tasks visible to their organizational counterparts as part of their outward-facing identity work ([Watson, 2009](#)). They thus seem to have a pragmatic attitude towards those tasks (see [Carter and Sholler, 2016](#)). Putting on display that mundane tasks like data cleaning form an important though non-prestigious part of their work, data scientists try to manage their counterparts' (hyped) expectations. This seems particularly necessary in situations where hiding those tasks might, in the long run, lead to fiercer identity threats, such as when managers start to realize that data scientists cannot live up to their expectations.

What transpires from our study is that expectations management is an important element in data scientists' identity work and attempts to enact their identity narratives. While data scientists at times experience managers' expectations as either too low or too high, they would often respond to such challenges through "educational work", trying to create awareness among managers of what data science is and what it can(not) do. Our study, however, also demonstrates that data scientists try to address this challenge by acting upon themselves. In other words, interviewees acknowledge the need to immerse themselves more strongly in the organization. This encompasses endeavours to put themselves in the managers' shoes to understand their perspective and the challenges they face. Interviewees mentioned in this context that they would try to learn to speak the managers' language to be able to better communicate with them. Such attempts are arguably important for data scientists to understand and effectively respond to managers' expectations and hence to effectually enact the problem-solving attitude component of their identity.

Our interviews thus revealed that instead of merely identifying differences and similarities between themselves and the managers they are working with, data scientists try to *create* similarities concerning the understanding of the possibilities and limitations of data science. This translates into both outward- and inward-facing identity work ([Watson, 2009](#)). Outward-facing, data scientists try to influence managers' understanding of data science and how they see the role of the data scientist. Part of this is to counter too strong an "action orientation" ([Jörden et al., 2022](#)) on the part of the managers. To do so, data scientists enact their "scientific mindset" and particularly emphasize the time and diligence it takes for their "sophisticated work" to be conducted. This involves emphasizing some of the more mundane aspects of their data work such as data cleaning even though these tasks stand in some contrast to their emphasis on the sophisticated parts of their work.

Inward-facing, data scientists act upon themselves to better understand the managers' "thought worlds" ([Dougherty, 1992](#)) and the problems they face in their daily work and incorporate this evolving understanding into their occupational identity. Such reactions can be interpreted as manifestations of data scientists' attempts to support the "problem-solving attitude" component of their identity narrative. Ambiguous expectations can make it challenging for data scientists to figure out what managers consider appropriate deliverables. This can, over time, challenge the sense of continuity and coherence of their identity narrative, affecting eventually how they develop a sense of direction and orientation for their work (e.g. setting appropriate goals for oneself) or challenge their understanding of what characterizes and makes them distinct from others ([Kärreman and Alvesson, 2001](#), p. 63).

What follows from this is that data scientists try to enlarge their sense of self from its "technical" core to encompass domain-related skills and activities to better comprehend and manage their internal stakeholders' expectations. Thus, while being generally framed as a technical occupation (see [Avnoon, 2021](#)), based on their workplace experiences and perceived need to manage their internal stakeholders' expectations, data scientists develop a "hybrid identity". That is, they see themselves as both technical specialists and general problem-solvers who (strive to) understand the various challenges that management is facing.

This observation might also help to explain Avnoon's (2021) conclusion that data scientists try to claim an expert role by presenting themselves as generalists rather than specialists. Focusing too strongly on a role as technical specialists might make it more difficult for data scientists to integrate their work into other organizational practices and enact the "problem-solving attitude" part of their identity. It might furthermore signal that what data scientists do is so distinct from managers' work that they do not need to engage with the intricacies of data science. This might spur data scientists' identity challenges that result from managers' lack of knowledge about their craft.

Conclusion

There is an increased interest in questions of big data and data science within the accounting literature (e.g. Bhimani and Willcocks, 2014; Moll and Yigitbasioglu, 2019; Möller *et al.*, 2020; Oesterreich and Teuteberg, 2019; Richins *et al.*, 2017). This discussion has also raised the question of whether accountants may eventually be replaced by data scientists or how the relationship between these occupational groups will develop (see Möller *et al.*, 2020). While our paper has not specifically focused on the relationship between accountants and data scientists – the declared focus was on inquiring more generally into the work experiences and identity-related challenges of data scientists – it is nevertheless informative about these questions. Two conclusions appear plausible in this respect.

First, our empirical material indicates that many data scientists see themselves as rather specialized in sophisticated types of data analysis and do not want to carry out tasks that fall into the domain of reporting, business intelligence or (in their view) simpler forms of data analysis. Many suggested that their work would neither affect nor overlap with what accountants do. This would suggest a rather clear separation in job profiles between data scientists and accountants who typically perform such tasks. While data scientists might take over some activities from accountants (those which require sophisticated forms of data science), it is unlikely that they want to do the full spectrum of tasks that accountants typically perform. Rather, we believe that there will be some consolidation in job profiles revolving around data analysis and business intelligence with that of (traditional) accountants or finance managers.

Second, our findings also suggest however that accounting and finance staff – like other functional managers – need to be able to cooperate with data scientists if the organization shall benefit from the possibilities of data science. Interestingly, while some of our interviewees have worked on accounting-related tasks (e.g. forecasting revenues), overall, they had rather little interaction with accounting and finance staff and worked primarily on topics that would fall into the area of marketing, sales, manufacturing or risk management – presumably, because these topic areas had more priority when it came to applying data science. Our findings hence corroborate that data scientists and (management) accountants currently work closely together in relatively few organizations (see Möller *et al.*, 2020). In addition, our study adds empirical nuance to prior work conceptually discussing a potential competition between data scientists and accountants (see, e.g. Moll and Yigitbasioglu, 2019; Vollmer, 2019). Still, there are organizations in which applying data science in the area of accounting and finance is high on the agenda and where accountants might, in the future, need to cooperate (and in some areas even potentially compete) with data scientists. This requires that accountants have some knowledge of what data science can and cannot do; and that they appreciate the work process that data scientists typically go through when working with data.

Our paper has some limitations that can serve as stepping stones for future research. As our analysis is focused on the perspectives of data scientists, we identified other actors' expectations through the eyes of our interviewees. Although this makes sense for what

matters from an identity perspective is how the data scientists perceive such expectations, interviewing managers who work with data scientists might offer additional insights into why such expectations emerge and how managers perceive the collaboration with this novel type of expert staff. We thus encourage future qualitative work in this area that also zooms in on the dynamics of data science in individual organizations (e.g. [Barbour et al., 2018](#)) or “on the ground” ([Carter and Sholler, 2016](#)). A focus on specific workplaces ([Bechky, 2011](#)) may thereby reveal organizational context factors that facilitate (or challenge) the identity formation processes of data scientists and increase our understanding of how new technologies, and the expert staff deploying them, shape organizational life.

Going beyond the interactional dynamics between data scientists and managers, [Vaast \(2020\)](#) shows that data scientists carry out their identity work not only within their organizations but also through social media. Future research could hence explore how those different identity work “arenas” intersect and influence each other in data scientists’ identity work. Another area for further (multi-level) research relates to the future development of the data scientist as an occupation. Some of our interviewees second-guessed that rather than seeing “unicorn data scientists” flourish, we might encounter more specialized experts (cf. [Avnoon, 2021](#)) with more fine-grained job titles and that the title “data scientist” might even disappear at some point (see [Vaast and Pinsonneault, 2021](#)). Indeed, we have conducted our study at a time when the data scientist has been “on the rise” so to speak. It is therefore too early to assess how this occupational role will develop in the long term, but future research could examine the development of data science, and the data science occupation, in a more longitudinal fashion.

Notes

1. MacGyver is the protagonist of a TV series from the 1980s, who works for a government organization and solves tricky problems with inventive use of everyday items, thereby mobilizing his engineering and physics skills.
2. While those numbers present just rough estimations that data scientists made to give a general account of their work, it still seems interesting to mention that some interviewees estimated that modelling would account for only approx. 20% of their work.

References

- Abbott, A. (1988), *The System of Professions: An Essay on the Expert Division of Labor*, Chicago UP, Chicago.
- Abrahamson, E. (1991), “Managerial fads and fashions: the diffusion and rejection of innovations”, *Academy of Management Review*, Vol. 16 No. 3, pp. 586-612.
- Abrahamson, E. (1996), “Management fashion”, *Academy of Management Review*, Vol. 21 No. 1, pp. 254-285.
- Al-Htaybat, K. and von Alberti-Alhtaybat, L. (2017), “Big Data and corporate reporting: impacts and paradoxes”, *Accounting, Auditing & Accountability Journal*, Vol. 30 No. 4, pp. 850-873.
- Alvesson, M. and Willmott, H. (2002), “Identity regulation as organizational control: producing the appropriate individual”, *Journal of Management Studies*, Vol. 39 No. 5, pp. 619-644.
- Arnaboldi, M., Busco, C. and Cuganesan, S. (2017), “Accounting, accountability, social media and big data: revolution or hype?”, *Accounting, Auditing & Accountability Journal*, Vol. 30 No. 4, pp. 762-776.
- Ashforth, B., Kreiner, G., Clark, M. and Fugate, M. (2007), “Normalizing dirty work: managerial tactics for countering occupational taint”, *Academy of Management Journal*, Vol. 50 No. 1, pp. 149-174.
- Avnoon, N. (2021), “Data scientists’ identity work: omnivorous symbolic boundaries in skills acquisition”, *Work, Employment and Society*, Vol. 35 No. 2, pp. 332-349.

-
- Baldvinsdottir, G., Burns, J., Nørreklit, H. and Scapens, R.W. (2009), "The image of accountants: from bean counters to extreme accountants", *Accounting, Auditing & Accountability Journal*, Vol. 22 No. 6, pp. 858-882.
- Barbour, J.B., Treem, J.W. and Kolar, B. (2018), "Analytics and expert collaboration: how individuals navigate relationships when working with organizational data", *Human Relations*, Vol. 71 No. 2, pp. 256-284.
- Başkarada, S. and Koronios, A. (2017), "Unicorn data scientist: the rarest of breeds", *Program*, Vol. 51 No. 1, pp. 65-74.
- Baskerville and Myers (2009), "Fashion waves in information systems research and practice", *MIS Quarterly*, Vol. 33 No. 4, pp. 647-662.
- Bechky, B.A. (2011), "Making Organizational theory work: institutions, occupations, and negotiated orders", *Organization Science*, Vol. 22 No. 5, pp. 1157-1167.
- Berger, P.L. and Luckmann, T. (1966), *The Social Constructions of Reality: A Treatise in the Sociology of Knowledge*, Anchor, Garden City, NY.
- Bhimani, A. and Willcocks, L. (2014), "Digitisation, 'Big Data' and the transformation of accounting information", *Accounting and Business Research*, Vol. 44 No. 4, pp. 469-490.
- Brown, A.D. (2015), "Identities and identity work in organizations: identities and identity work", *International Journal of Management Reviews*, Vol. 17 No. 1, pp. 20-40.
- Carillo, K.D.A. (2017), "Let's stop trying to be 'sexy' – preparing managers for the (big) data-driven business era", *Business Process Management Journal*, Vol. 23 No. 3, pp. 598-622.
- Carter, D. and Sholler, D. (2016), "Data science on the ground: hype, criticism, and everyday work", *Journal of the Association for Information Science and Technology*, Vol. 67 No. 10, pp. 2309-2319.
- Collinson, J.A. (2006), "Just 'non-academics'? Research administrators and contested occupational identity", *Work, Employment and Society*, Vol. 20 No. 2, pp. 267-288.
- Davenport, T.H. and Patil, D.J. (2012), "The sexiest job of the 21st century", *Harvard Business Review*, Vol. 90 No. 5, pp. 70-76.
- David, R.J. and Strang, D. (2006), "When fashion is fleeting: transitory collective beliefs and the dynamics of TQM consulting", *Academy of Management Journal*, Vol. 49 No. 2, pp. 215-233.
- Dougherty, D. (1992), "Interpretive barriers to successful product innovation in large firms", *Organization Science*, Vol. 3 No. 2, pp. 179-202.
- Down, S. and Reveley, J. (2009), "Between narration and interaction: situating first-line supervisor identity work", *Human Relations*, Vol. 62 No. 3, pp. 379-401.
- Frey, C.B. and Osborne, M.A. (2017), "The future of employment: how susceptible are jobs to computerisation?", *Technological Forecasting and Social Change*, No. 114, pp. 254-280.
- Gehl, R.W. (2015), "Sharing, knowledge management and big data: a partial genealogy of the data scientist", *European Journal of Cultural Studies*, Vol. 18 Nos 4-5, pp. 413-428.
- Giddens, A. (1984), *The Constitution of Society: Outline of the Theory of Structuration*, University of California Press, Berkeley, Los Angeles.
- Giddens, A. (1991), *Modernity and Self-Identity: Self and Society in the Late Modern Age*, Stanford University Press, Stanford.
- Giroux, H. (2006), "'It was such a handy term': management fashions and pragmatic ambiguity", *Journal of Management Studies*, Vol. 43 No. 6, pp. 1227-1260.
- Goffman, E. (1963), *Stigma: Notes on a Spoiled Identity*, Prentice-Hall, Englewood Cliffs.
- Goretzki, L. and Messner, M. (2019), "Backstage and frontstage interactions in management accountants' identity work", *Accounting, Organizations and Society*, Vol. 74, pp. 1-20.
- Goretzki, L. and Pfister, J.A. (2022), "The productive accountant as (un-)wanted self: realizing the ambivalent role of productivity measures in accountants' identity work", *Critical Perspectives on Accounting*, 102504.

- Goretzki, L., Löhlein, L., Schäffer, U., Schmidt, A. and Strauss, E. (2022), "Exploring the role of metaphors in social-identity construction: the case of the German controller", *European Accounting Review*, Vol. 31 No. 4, pp. 877-903 (in press).
- Harris, J.G. and Mehrotra, V. (2014), "Getting value from your data scientists", *MIT Sloan*, Vol. 56 No. 1, pp. 15-18.
- Harris, H., Murphy, S. and Vaisman, M. (2013), *Analyzing the Analyzers: An Introspective Survey of Data Scientists and Their Work*, O'Reily Media, Sebastopol, CA.
- Ibarra, H. and Barbulescu, R. (2010), "Identity as narrative: prevalence, effectiveness, and consequences of narrative identity work in macro work role transitions", *Academy of Management Review*, Vol. 35 No. 1, pp. 135-154.
- Jeacle, I. (2008), "Beyond the boring grey: the construction of the colourful accountant", *Critical Perspectives on Accounting*, Vol. 19 No. 8, pp. 1296-1320.
- Jenkins, R. (2014), *Social Identity*, Routledge, Milton Park.
- Jörden, N.M., Sage, D. and Trusson, C. (2022), "'It's so fake': identity performances and cynicism within a people analytics team", *Human Resource Management Journal*, Vol. 32 No. 3, pp. 524-539.
- Kärreman, D. and Alvesson, M. (2001), "Making newsmakers: conversational identity at work", *Organization Studies*, Vol. 22 No. 1, pp. 59-89.
- Kahl, S.J., King, B.G. and Liegel, G. (2016), "Occupational survival through field-Level task integration: systems men, production planners, and the computer, 1940s–1990s", *Organization Science*, Vol. 27 No. 5, pp. 1084-1107.
- Kieser, A. (1997), "Rhetoric and myth in management fashion", *Organization*, Vol. 4 No. 1, pp. 49-74.
- Kuhn, T. (2006), "A 'demented work ethic' and a 'lifestyle firm': discourse, identity, and workplace time commitments", *Organization Studies*, Vol. 27 No. 9, pp. 1339-1358.
- Madsen, D.Ø. and Stenheim, T. (2016), "Big Data viewed through the lens of management fashion theory", *Cogent Business and Management*, Vol. 3 No. 1, 1165072.
- McInnes, P. and Corlett, S. (2012), "Conversational identity work in everyday interaction", *Scandinavian Journal of Management*, Vol. 28 No. 1, pp. 27-38.
- Messner, M., Becker, C., Schäffer, U. and Binder, C. (2008), "Legitimacy and identity in Germanic management accounting research", *European Accounting Review*, Vol. 17 No. 1, pp. 129-159.
- Möller, K., Schäffer, U. and Verbeeten, F. (2020), "Digitalization in management accounting and control: an editorial", *Journal of Management Control*, Vol. 31 Nos 1-2, pp. 1-8.
- Moll, J. and Yigitbasioglu, O. (2019), "The role of internet-related technologies in shaping the work of accountants: new directions for accounting research", *The British Accounting Review*, Vol. 51 No. 6, 100833.
- Morales, J. and Lambert, C. (2013), "Dirty work and the construction of identity. An ethnographic study of management accounting practices", *Accounting, Organizations and Society*, Vol. 38 No. 3, pp. 228-244.
- Muller, M., Lange, I., Wang, D., Piorowski, D., Tsay, J., Q, V.L., Dugan, C. and Erickson, T. (2019), "How data science workers work with data: discovery, capture, curation, design, creation", *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*.
- Murphy, C. and Kreiner, G.E. (2020), "Occupational boundary play: crafting a sense of identity legitimacy in an emerging occupation", *Journal of Organizational Behavior*, Vol. 41 No. 9, pp. 871-894.
- Nelson, A.J. and Irwin, J. (2014), "'Defining what we do—all over again': occupational identity, technological change, and the librarian/internet-search relationship", *Academy of Management Journal*, Vol. 57 No. 3, pp. 892-928.
- Oesterreich, T.D. and Teuteberg, F. (2019), "The role of business analytics in the controllers and management accountants' competence profiles: an exploratory study on individual-level data", *Journal of Accounting and Organizational Change*, Vol. 15 No. 2, pp. 330-356.

- Pachidi, S., Berends, H., Faraj, S. and Huysman, M. (2021), "Make way for the algorithms: symbolic actions and change in a regime of knowing", *Organization Science*, Vol. 32 No. 1, pp. 18-41.
- Petriglieri, J.L. (2011), "Under threat: responses to and the consequences of threats to individuals' identities", *Academy of Management Review*, Vol. 36 No. 4, pp. 641-662.
- Pratt, M.G., Rockmann, K.W. and Kaufmann, J.B. (2006), "Constructing professional identity: the role of work and identity learning cycles in the customization of identity among medical residents", *Academy of Management Journal*, Vol. 49 No. 2, pp. 235-262.
- Quattrone, P. (2016), "Management accounting goes digital: will the move make it wiser?", *Management Accounting Research*, Vol. 31, pp. 118-122.
- Reay, T., Goodrick, E., Waldorff, S.B. and Casebeer, A. (2017), "Getting leopards to change their spots: Co-creating a new professional role identity", *Academy of Management Journal*, Vol. 60 No. 3, pp. 1043-1070.
- Richins, G., Stapleton, A., Stratopoulos, T.C. and Wong, C. (2017), "Big data analytics: opportunity or threat for the accounting profession?", *Journal of Information Systems*, Vol. 31 No. 3, pp. 63-79.
- Saunders, T. (2013), "Data science and data scientists: what's in a name?", *Information Management*, Vols 1-3, November 2011.
- Strauss, A. and Corbin, J. (1998), *Basics of qualitative research. Techniques and procedures for developing Grounded Theory (2nd edition)*, Sage, London.
- Suchman, M.C. (1995), "Managing legitimacy: strategic and institutional approaches", *Academy of Management Review*, Vol. 20 No. 3, pp. 571-610.
- Sveningsson, S. and Alvesson, M. (2003), "Managing managerial identities: organizational fragmentation, discourse and identity struggle", *Human Relations*, Vol. 56 No. 10, pp. 1163-1193.
- Thornborrow, T. and Brown, A.D. (2009), "Being regimented: aspiration, discipline and identity work in the British parachute regiment", *Organization Studies*, Vol. 30 No. 4, pp. 355-376.
- Vaast, E. (2020), "A seat at the table and a room of their own: interconnected processes of social media use at the intersection of gender and occupation", *Organization Studies*, Vol. 41 No. 12, pp. 1673-1695.
- Vaast, E. and Pinsonneault, A. (2021), "When digital technologies enable and threaten occupational identity: the delicate balancing act of data scientists", *MIS Quarterly*, Vol. 45 No. 3, pp. 1087-1112.
- Vaivio, J., Järvenpää, M. and Rautiainen, A. (2021), "Accounting in identity regulation: producing the appropriate worker", *European Accounting Review* (in press).
- Vollmer, H. (2019), "Accounting for tacit coordination: the passing of accounts and the broader case for accounting theory", *Accounting, Organizations and Society*, Vol. 73, pp. 15-34.
- Waller, M.A. and Fawcett, S.E. (2013), "Click here for a data scientist: big data, predictive analytics, and theory development in the era of a maker movement supply chain", *Journal of Business Logistics*, Vol. 34 No. 4, pp. 249-252.
- Watson, T.J. (2008), "Managing identity: identity work, personal predicaments and structural circumstances", *Organization*, Vol. 15 No. 1, pp. 121-143.
- Watson, T.J. (2009), "Narrative, life story and manager identity: a case study in autobiographical identity work", *Human Relations*, Vol. 62 No. 3, pp. 425-452.
- Wittman, S. (2019), "Lingering identities", *Academy of Management Review*, Vol. 44 No. 4, pp. 724-745.
- Wrzesniewski, A. and Dutton, J.E. (2001), "Crafting a job: revisioning employees as active crafters of their work", *Academy of Management Review*, Vol. 44 No. 4, pp. 724-745.
- Zbaracki, M.J. (1998), "The rhetoric and reality of total quality management", *Administrative Science Quarterly*, Vol. 43 No. 3, pp. 602-636.

#	Job title	Sector
1	Junior Data Scientist	Manufacturing
2	Lead Data Scientist	Transportation
3	Senior Data Scientist	Consulting
4	Data Science Team Lead	Software services
5	Head of DS/Chief Data Scientist	Manufacturing
6	Senior Data Scientist	Services
7	Data Scientist	Retail
8	Data Scientist	Utilities
9	Data Scientist	Retail (same firm as #7)
10	Head of Predictive Analytics and Machine Learning	Software services (same firm as #4)
11	Data Scientist	Services
12	Data Scientist	Services
13	Data Scientist	Services
14	Data Scientist	Services (same firm as #12)
15	Senior Business Intelligence Developer and Full Stack Data Scientist	NGO
16	Data Scientist	Software services
17	Data Scientist	Manufacturing
18	Data Scientist	Manufacturing
19	Data Scientist	Public agency
20	Predictive Data Analyst	Services
21	Data Scientist	Software services
22	Head of Data Science	Software services
23	Data Scientist	Services
24	Senior Software Engineer	Manufacturing (same firm as #1)
25	Data Scientist	Manufacturing

Table A1.
List of interviewees

Corresponding author

Martin Messner can be contacted at: martin.messner@uibk.ac.at

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com