FASHIONABLE PRACTICES AND OCCUPATIONAL IDENTITIES:

HYPE AND AMBIGUITY AS CHALLENGES FOR DATA SCIENTISTS

Lukas Goretzki [1]

Martin Messner [2, *]

Maria Wurm [2]

[1] Stockholm School of Economics, Sweden

[2] University of Innsbruck, Austria

* Corresponding author, martin.messner@uibk.ac.at

Acknowledgements: We thank our interview partners for taking their time to share their experiences with us. 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, and Concordia University. We thank the participants and discussants at these events for their helpful comments.

FASHIONABLE PRACTICES AND OCCUPATIONAL IDENTITIES: HYPE AND AMBIGUITY AS CHALLENGES FOR DATA SCIENTISTS

Abstract:

The paper examines how the fashionable nature of the data science discourse, in terms of its hype and ambiguity, affects the occupational identity of data scientists. Drawing upon interviews with data scientists, we show how these data workers carry out narrative identity work to distance themselves from the hype and ambiguity that they perceive in the discourse on data science. The fashionable nature of the data science discourse also translates into data scientists' work in organizations, where it creates challenges for successfully performing their work and realizing their aspirational identity. We show how data scientists try to address these challenges by exercising interactive identity work vis-à-vis significant others. Overall, our paper discusses the relationship between fashionable ideas and discourses, on the one hand, and occupational identities, on the other hand, in light of the contemporary digitalization megatrend. It furthermore contributes to our understanding of the work realities of data scientists as a nascent occupational group.

FASHIONABLE PRACTICES AND OCCUPATIONAL IDENTITIES: HYPE AND AMBIGUITY AS CHALLENGES FOR DATA SCIENTISTS

INTRODUCTION

Based on 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 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 (Moll & Yigitbasioglu, 2019). Indeed, and notwithstanding its true transformational potential, several observers refer to the “fashionable” nature (see Madsen & Stenheim, 2016) of the data science discourse (see Carter & Sholler, 2016; Gandomi & Haider, 2015), sometimes even referring to an “obsession” with big data and data science (Gehl, 2015, p. 419).

The fashionable discourse around data science does not only cover new practices and systems but also new types of experts (Avnoon, 2021). The most prominent group of knowledge workers creating and applying data science models are data scientists. Data scientists have been portrayed as the “missing piece of the big data puzzle” (Carillo, 2017, p. 607), the “most wanted” experts (Waller & Fawcett, 2013), and even as “The Sexiest Job of the 21st Century” (Davenport & Patil, 2012). Accounting scholars have also 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 & Willcocks, 2014; Moll & Yigitbasioglu, 2019; Oesterreich & Teuteberg, 2019; Richins, Stapleton, Stratopoulos, & Wong, 2017). Given these accounts that reflect the current “buzz” around this new expert group, it is not surprising that indeed more and more organizations (seek to) employ such expert staff and that an increasing number of individuals are interested in a career as a data scientist.

This paper empirically enquires into the work experiences and challenges of data scientists in organizations. For, despite the positive discourse surrounding this occupation, we still know rather little about how data scientists experience and enact their role in organizations. Research has started to explore the work experiences of data scientists (e.g., Barbour, Treem, & Kolar, 2018; Carter & Sholler, 2016; Harris & Mehrotra, 2014) and we aim to advance this evolving conversation by focusing on challenges that relate specifically to the data scientists’ occupational *identity* (Avnoon, 2021; Vaast & Pinsonneault, 2021). This is particularly relevant since, due to the nascent state of the occupation, data scientists cannot draw on established (ready-made) identity scripts or templates to develop a sense of self (see Watson, 2008).

Like prior research, we take the current interest in, and enthusiasm for, data science as an empirical motivation for our study (e.g., Avnoon, 2021; Barbour et al., 2018; Carter & Sholler, 2016; Harris & Mehrotra, 2014). Yet, we also consider this enthusiasm (or hype) to be theoretically relevant for understanding the identity work of data scientists as members of a nascent occupation born out of the fashionable discourse on data science. On the one hand, the fashionable nature of their occupation makes data scientists less concerned about struggles for legitimation that other occupations, such as accountants (e.g., Jeacle, 2008), have historically been involved in. On the other hand, such a fashionable nature can also pose a challenge for data scientists. When a set of practices (like data science) becomes fashionable, managers and other actors may develop unrealistic expectations about it and interpret it in an ambiguous way (Abrahamson, 1991, 1996; Giroux, 2006; Zbaracki, 1998). This, we argue, can create challenges for the identity of those actors (here data scientists) who deploy these fashionable practices (here data science). Indeed, we argue that it is precisely the hype around data science, and the accompanying ambiguity surrounding that term, which can create complications for data scientists in the workplace, not least concerning how they construct their occupational identity. This is because managers at times develop vague or unrealistic expectations regarding the possibilities of data science, which data scientists have to cope with in their day-to-day work. Hence, complementing insights on the delicate nature of data scientists' identity (Vaast & Pinsonneault, 2021), we demonstrate that the fashionable nature of data science is not only beneficial for data scientists (in terms of job opportunities and good salaries, for instance) but can also render their occupational identity fragile.

To develop these findings, we draw 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 in light of the fashionable discourse on data science and how they experience and manage identity-related challenges when performing their work. In so doing, we extend prior research which has mostly focused on data scientists' activities and interactions with managers, without examining in detail their identity (work) (e.g., Barbour et al., 2018; Carter & Sholler, 2016; Harris & Mehrotra, 2014).

Research in accounting has started to explore the implications of digital business practices and datafication for decision-making and control (e.g., Al-Htaybat & von Alberti-Alhtaybat, 2017; Arnaboldi, Busco, & Cuganesan, 2017; Chapman, Chua, & Fiedler, 2021; Van den Bussche & Dambrin, 2020). We certainly agree that there is vast potential for data science applications to improve many organizational processes, not least in the area of accounting and control. Data scientists play an important role in this respect, as experts who liaise with accountants and functional managers to analyze and learn from data. Our findings suggest however that there are also challenges on the way there. In particular, unclear or unrealistic expectations on the part of managers can make it difficult for data scientists to successfully leverage their expertise. There is a need for managers to take some critical distance from the hype around data science and to stay 'grounded'. And there is a need for data scientists to overcome language barriers with managers so as to successfully bring together the technical and the business dimensions.

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 concluding discussion.

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, Murphy, and Vaisman (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 allrounders with a broad set of competencies, including business knowledge, programming knowledge, statistical knowledge, and soft skills (Carillo, 2017; Davenport & Patil, 2012), and some authors question whether one individual can incorporate all required skills (Carillo, 2017; Waller & Fawcett, 2013). Baškarada and Koronios (2017) hence talk about data scientists as ‘mythical creatures’ and ‘unicorns’ and emphasize the at times unrealistically high expectations vis-à-vis actors filling those positions.

Given the occupation’s novelty, empirical research on data scientists’ work experiences in organizations is still rather scarce. Some studies, however, have offered 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 the problems to be addressed. Focusing on such inter-functional relations in data science projects, Pachidi, Berends, Faraj, and Huysman (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 role and identity. Important initial scholarly work on data scientists’ occupational identity was conducted by Vaast and Pinsonneault (2021). 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 find their identity, 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 elite or "expert" group. Based on interviews with data scientists the author shows that rather than claiming an expert status through specializing and focusing on specific areas, data scientists tend to adopt a so-called "omnivorous" approach to skill acquisition. This means that data scientists 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" (p. 342-343). Although 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 the data scientist and their identity work within the organization. While Vaast and Pinnsoneault (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 examine the impact of the fashionable nature of the data science discourse on data scientists' identity work. In doing so, we aim to contribute to an empirically grounded and theoretically substantiated understanding of this nascent occupation. We thereby pay particular attention to the tensions the fashionable data science discourse creates for data scientists in establishing a sense of self in the organization. We argue that this focus is important not least because data scientists are often assigned a salient expert role in the discourse around digital technologies. And, as mentioned above, data scientists are sometimes presented in the popular discourse as "digital saviors" who can help managers and accountants to make the "dream of perfect information and rational decision-making" (Quattrone, 2016, p. 118) come true. The fact that data scientists currently populate more and more organizations and seem to enjoy a good amount of praise in the general discourse begs the question how the fashionable nature of their occupation is implicated in their identity work "on the ground" (see Carter & Sholler, 2016). The next section outlines the theoretical background for our analysis.

THEORETICAL BACKGROUND

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, pp., 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 that are communicated to the focal actor (Berger & 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 & 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.

Such external influence does not simply determine identities, however. Rather, individuals exercise agency and selectively endorse some elements of this discourse, while resisting others (Brown, 2015). In other words, actors engage in *identity work*, defined as the "forming, repairing, maintaining, strengthening or revising the constructions that are productive of a sense of coherence and distinctiveness" (Sveningsson & Alvesson, 2003, p. 1165). Identity work is guided by some sense of aspirational self (Thornborrow & Brown, 2009) that individuals pursue. 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 & Lambert, 2013), advances in technology (Nelson & Irwin, 2014), or stigmatization (Ashforth, Kreiner, Clark, & Fugate, 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 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, in the giving of instructions, or the issuing of complaints at the workplace (Down & Reveley, 2009). More generally, when actors craft their job by changing tasks and relations to other actors (Wrzesniewski & Dutton, 2001), then such job crafting often performs and symbolizes a particular identity, both for oneself and others (Goretzki & Messner, 2019). Indeed, both narratives and interactions should be understood as having an inward- and outward orientation (Watson, 2008), in the sense that they (potentially) influence one's self-understanding as well as others' expectations towards oneself, i.e., the identity ascribed by others.

Turning towards data scientists, what is particularly interesting concerning the identity work of the members of this nascent (technical) occupation (Avnoon, 2021), is that – in contrast to established occupations – they cannot draw on well-established and discursively available "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) but need "to figure out creative ways of crafting a sense of legitimacy around their identities" (p. 888). We thus expect the occupational identities of data scientists to be shaped by the broader discourse on data science, as transmitted through various channels, but also by experiences at the workplace. We propose in this respect that, in crafting their identity, data scientists endorse some aspects of the general discourse around their role and craft and resist others in their efforts to establish a particular sense of self. Precisely how they do this and what identity-related challenges emerge in this process is however an empirical question. Before entering a detailed analysis of this process, we will first provide a theoretical foundation for the nature of the contemporary data science discourse confronting data scientists as well as their organizational counterparts. We will do so by reverting to the notion of "fashion" and discuss its relevance for our analysis.

Fashionable practices and occupations

As previously elaborated, data science can be seen as currently "in fashion" and as featuring many of the characteristics that management fashions exhibit more generally (Abrahamson, 1991, 1996). The enthusiasm for data science is not least driven by a fashion-setting community that consists of "consulting firms, software vendors, management gurus, conference/seminar organizers, business school academics, business media, analysts, as well as professional organizations" (Madsen & Stenheim, 2016, p. 3). These push data science as an important building block of the future of almost any organization, thus making it difficult for managers to ignore this trend. And, indeed, the popular discourse on data science (see Carter & Sholler, 2016) also shapes organizational discourses. Authors refer in this context to a growing "data imperative" or "fetishization of data" affecting all kinds of organizations (Saifer & Dacin, 2021). This also affects the sphere of accounting where data science has become a fashionable topic associated with modern tools and systems that build on the latest technological advancements to which companies and individuals – not least managers and accountants – are expected to adjust to be "future-ready".

It is noteworthy that to call something a 'fashion' does not mean that it is not substantial in terms of its economic potential or effects. It simply means that there is a lot of attention being paid to this phenomenon, to the effect that many actors will believe this to be important for them – irrespective of whether it is or not. We refer to this pronounced attention as *hype* (see Carter & Sholler, 2016; Eccles, Nohria, & Berkley, 1992; Fenn & Raskino, 2008). For a concept or practice to become a fashion, it "must claim that it is innovative and that its application is inevitable" (Benders, van den Berg, & van Bijsterveld, 1998, p. 201). As Eccles et al. (1992) write, "stressing change and newness has an undeniable rhetorical power" (p. 4), and it is based on such rhetorical excess that the hyped nature of a concept or practice (like data science) is established.

Fashions typically feature another key characteristic, and that is *ambiguity*. Ambiguity means that the concept or practice "allow[s] for different interpretations of its content" (Benders et al., 1998, p. 201). Indeed, it has been repeatedly argued that such ambiguity, vagueness, or interpretive viability allows fashionable practices to become popular in the first place, as ambiguity allows different actors to enact the practice in the way that meets their interests (Ansari, Fiss, & Zajac, 2010; Benders & Van Veen, 2001; Giroux, 2006). Ambiguity, in turn, increases as a practice becomes more fashionable, as variable enactments by organizations feed back into the discourse on the practice (Giroux, 2006).

In the same way in which data science can be seen as a fashion, data scientists can be regarded as a "fashionable occupation", that is, an occupation that relatively quickly gained a strong popularity and coverage in the wider management discourse (Jung & Kieser, 2012, p. 329). The fashionable status of the job profile 'data scientist' is apparent from descriptions such as "The Sexiest Job of the 21st Century" (Davenport & Patil, 2012) or in rankings of the data scientist job as one of the best jobs in America in 2020[1]. A fashion on the demand side obviously creates opportunities for those who are on the supply side, i.e., the data scientists. And yet, the fashionable nature of the data science discourse may not only have benefits for data scientists. Indeed, as we shall demonstrate below, our empirics evidence that the hyped and ambiguous nature of the data science discourse creates a set of challenges for data scientists in the workplace, particularly concerning their identity. Discourses shape identities (Alvesson & Willmott, 2002), and the focus in our analysis below is on examining how the fashionable nature of the data science discourse affects the identity (work) of data scientists as members of a nascent technical occupation (Avnoon, 2021). In doing so, we follow the call for future research by Carter and Sholler (2016) who argue that more research looking at the individual level is called for to better understand what data scientists do "on the ground" and how they cope with the overly positive popular discourse on their craft.

RESEARCH METHOD

The empirical study is a cross-sectional interview study with 25 data scientists from 21 different organizations. The initial objective of carrying out a cross-sectional study (as opposed to a single case study, for instance) was to identify patterns in the work experiences and perspectives of data scientists across different organizations and sectors. By choosing a cross-sectional *qualitative* study, we sought to identify such patterns while at the same time providing a level of empirical detail that could illuminate in a reasonably authentic way the self-understandings, work practices, and challenges of data scientists (similarly for other occupational groups: Ashcraft, 2007; Collinson, 2006; Reed & Thomas, 2021; Wright, Nyberg, & Grant, 2012).

Data collection

Most interview partners 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. 16 people accepted to be interviewed. In addition, we asked our interview

[1] https://www.glassdoor.com/List/Best-Jobs-in-America-LST_KQ0,20.htm (retrieved January 11, 2022)

partners 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 enroll two further interviewees into our study. Finally, 7 interview partners were recruited based upon 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 open and communicative.

Most of the interviews were conducted 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 validated 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.). Two 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 list of interviewees is provided in the Appendix.

In conducting the 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 the work experiences of data scientists, with a particular focus on the challenges that they encounter in their work. We were interested in understanding where these challenges originated and how data scientists would deal with them. Accordingly, we had pre-specified three main deductive categories, which we sought to explore in more depth in the interviews. In other words, the interviews 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.

Data analysis

We conducted our data analysis by carefully reading through all interview transcripts (and notes) and identifying passages that somehow related to the deductively defined areas of interest, as described above. We coded each instance on the 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 & Corbin, 1998).

Having assigned codes to each selected interview passage, we then went through all these codes again and tried to consolidate. 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 many challenges related to expectations held by managers that did not align with how the data scientists saw themselves and their work. These expectations, in turn, often related to ideas about data science in the broader discourse on the topic. In a second step, we, therefore, focused more closely on those interview codes where data scientists reflected upon the discourse on data science and positioned themselves vis-à-vis this discourse – instances of identity work (Sveningsson & Alvesson, 2003). Two themes were dominant here: the hype around data science and the ambiguity of the term. Both themes triggered identity constructions among data scientists and related to experiences in the workplace. We selected those work challenges that pertained to hype and ambiguity, respectively, and looked at how data scientists would respond to such challenges.

Table 1 summarizes our findings. We suggest that the fashionable nature of the phenomenon surfaces in terms of ambiguity around the term data science as well in the form of a hype. Ambiguity and hype motivate narrative identity work among data scientists, through which they position themselves vis-à-vis this discourse. Ambiguity and hype are also visible within organizations, where they translate into practical challenges for data scientists and can pose threats to their identity. Data scientists respond to these challenges through interactive identity work.

– Insert table 1 about here –

The first part of our empirical section examines the first two columns; the second section is then dedicated to discussing columns three and four.

EMPIRICAL FINDINGS

Identity narratives in light of ambiguity and hype

Our interviews show that data scientists do not take their work and occupational identity for granted, but reflect about who they are and what they do. These reflections are triggered in particular by two perceptions on the part of the data scientists. On the one hand, our interviewees perceive *ambiguity* in the notion of a data scientist and its use, which they contemplate on; on the other hand, they also reflect upon the *hype* that accompanies the data science phenomenon. We find that our interviewees take a critical distance both towards ambiguity and hype and we interpret this rhetorical distancing as a form of identity work, whereby data scientists enact, and communicate, a particular self-understanding about who they are or want to be (Murphy & Kreiner, 2020; Sveningsson & Alvesson, 2003).

Ambiguity: More than a data analyst

It becomes apparent from our interviews that data scientists perceive the term 'data science' to be used in various ways, resulting in much ambiguity of that term. As explained in our theory section, such ambiguity is quite common for emerging occupations (Murphy & Kreiner, 2020) and so it is perhaps not surprising to see it materialize in our empirical context as well, given that data scientists have only emerged recently as an occupational group. What is interesting to us is how data scientists position themselves vis-à-vis this ambiguity.

Our interviewees suggest that there are different understandings of what data science is and, hence, what a data scientist should be doing. Interviewee 4, for instance, tells us 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 confirms 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*". Interviewee 15 confirms 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*". He predicts that the job title 'data scientist' as such will disappear in time, giving way to more specialist titles that better capture differences in activity.

To some extent, the data scientists interviewed accept this ambiguity surrounding their job profile as being related to the nascent state of their occupation. Yet, they also voice concerns related to their own identity as a data scientist. The perceived ambiguity in the term 'data scientist' creates an identity threat (Petriglieri, 2011) for data scientists insofar as it may create understandings of their work that are not desirable. As a consequence, ambiguity triggers identity work among data scientists, whereby the latter try to differentiate themselves from part of the job profiles that are subsumed under this term (Murphy & Kreiner, 2020). In particular, they emphasize the need to distinguish between a 'real' data scientist and what they refer to as a 'mere' data analyst. Several of our interviewees suggest such a separation, without being prompted by us. They describe data analysts 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 see their role as having to go beyond what data analysts 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).

While such statements are often phrased in factual terms and in the present tense, one should interpret them as having an aspirational element (Thornborrow & Brown, 2009) whereby data scientists enact an identity and image which they feel is desirable. Key to this aspirational identity seems to be the advanced modelling part of their job. Data scientists like to find "*hidden treasures in the data*" and to develop "*models that are stronger and better*" (21). Data scientists believe that good models are the basis for making a real change to the organization, and this is where they see their value-added. Interviewee 6 contrasts having such an influence with what he calls "*cosmetic data science*", where someone "*wants to improve a little bit here, a little bit there*". According to him, "*data science is not about cosmetics",* but implies "*changing the way they do business"* by promoting *"data-driven decisions".* Such statements reflect an aspirational identity that highlights the importance of data scientists and distinguishes them from what is seen as more traditional or incremental ways of managing data. Such an identity goes along with expertise and some of our interviewees explicitly highlight the creative aspects of their work. Interviewee 1, for instance, 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*".

Overall, we can see how the perceived ambiguity of their occupational identity triggers identity work on a rhetorical level, whereby data scientists emphasize the modelling part of their role, in distinction to what they regard as less sophisticated ways of handling data.

Hype: No unicorns or magicians

A second recurrent theme in our interviewees' reflections about their work and occupational identity was the hype that they perceive around data science. Like in the case of ambiguity, interviewees are critical about this hype, even though they also understand that they benefit from it, in terms of a high demand for data science and the job opportunities that emerge from it. Still, too much hype around their work seems to make them uncomfortable, creating again an identity threat (Petriglieri, 2011). In contrast to ambiguity, hype is not a key feature of emerging occupations more generally, as there are also several emerging occupations which do not experience a hype, but go rather unnoticed or even have to battle for legitimacy (Murphy & Kreiner, 2020). We see the hype around data scientists as resulting from the all-present 'digitalization imperative' that organizations and managers nowadays face.

The hype around data science materializes in the circulation of a lot of "*buzzwords*" (7), such as big data or machine learning, which are used without always being clear about what these precisely mean. Our interviewees feel that there is a sense that big data "*solves all your problems*", which is not true, however, as the applicability of data science methods is also limited (7). Interviewee 12 tells us that "*sometimes you get the impression that, because something in the company is complicated or problematic, you just build an artificial intelligence [solution] and jam in all the data you have and then it gives us the answer* [but] *this is an illusion*". This hype around data science translates into hyped images of the data scientist. One of the interviewees feels like having to be a "*unicorn*" (16) and refers to marketing departments in particular which would expect: "*Now we have data and a data scientist, so he will be able to do this somehow and to derive some insights*", while in reality, things are often not that straightforward.

It is apparent from the above accounts that data scientists also feel that the enthusiasm about data science is accompanied by a certain level of ignorance about it, both in organizations and public discourse more generally. Unwarranted beliefs in the 'magic' of data science are in this sense not just evidence of an overly optimistic attitude (hype) about data science, but also originate from a lack of knowledge among managers (or commentators) regarding the actual possibilities (and challenges) that it brings about. Interviewee 23 recounts in this context that

> *"… of course, everyone has heard the words data science, machine learning, artificial intelligence and so on. Many people like to talk about it regardless of whether they […] really know what they mean or not. And, of course, these buzzwords are often used in the wrong context. […] Buzzwords are just very popular in this context and you often read them in newspapers or [they appear] on the television or radio […] and of course also in the company. So especially when things go up at board level, these words are also very popular to ultimately describe your own company or to describe the future of your own company"* (23).

The close relationship between hype and ignorance is not surprising. Prior literature suggests that enthusiasm for new practices often goes along with ignorance about them and that such ignorance "leads managers to hold unrealistic and simplistic expectations" (Zbaracki, 1998, p. 616) – something which we will elaborate further below.

The hype around data science is potentially a problem for data scientists in two directions. On the one hand, it leads to unrealistic expectations regarding the possibilities of data science practices. On the other hand, a hyped discourse may lead to skepticism among managers regarding the 'true' value of data science, to such an extent that it is seen as a fashion that is expected to pass away eventually. Interviewee 5 puts this succinctly when explaining that there are only extreme opinions at the moment. *"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 deux ex machina which can solve everything"*. He further recounts that the data scientist is often seen as a kind of *"MacGyver"*[2] who can easily address all kinds of difficult problems.

When our interviewees refer to expectations like being a 'unicorn', 'magician' or acting like 'MacGyver', it is clear that they take a critical distance towards such hyped understandings of their work. It seems as if they feel the need to temper this hype and move the discussion about what data science – and they as data scientists – can and cannot do to a more realistic level. This is interesting, as it stands in contrast to the identity work that we described in the previous section, where data scientists would emphasize their difference to less sophisticated job profiles that they don't want to be amalgamated with. Indeed, this contrast becomes very visible when our interviewees speak about the typical data science project and what they spend their time on there. They would speak of the 'data modelling' stage as the one where they can really use their skills and the stage that would ultimately make the difference in terms of insights gained. At the same time, several of our interviewees emphasize 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%. The rest is about organization and data preparation*" (11).

---

[2] MacGyver is the protagonist of a TV series from the 1980s, who works for a government organization and solves tricky problems with an inventive use of everyday items, thereby mobilizing his engineering and physics skills.

Interviewees state that those tasks are time-consuming and "*can be cumbersome*" (16). Talking about a particular project, interviewee 8 remembers that "*it took [her] much time to check which data are available, which ones can be used and how they are prepared*". Data might also be "*distributed among two databases and [the question then is] how to combine them? You have to transfer them to somewhere else and then you need a server for this or you have it on the laptop. It can be cumbersome*" (16). The data collection stage is sometimes also prolonged because of time constraints of other departments or due to bureaucratic barriers. Interviewee 13 tells us that getting data from the IT department is difficult because it is hard for them sometimes to "*dedicate resources just to send [the] data*". And this means that he has to "*fight for data*". Similarly, interviewee 1 explains that getting data is difficult "*because of all the security issues inside and it has to go through so many people"* until he can get the data.

Data preparation and cleaning also consume quite a lot of time, according to the data scientists we interviewed. Talking about one of her projects, interviewee 8 explains that she and her team "*spent a lot of time on understanding the data and preparing them, to have a basic database on which we can run analyses*". Another interviewee similarly tells us that the "*quality of the data*" is often a problem in his projects. Databases are "*often not correctly populated and then you have to use a lot of time to solve these problems*" (2). Interviewee 6 even says that, in one recent case, he took out "*like 90% of the data… I threw it out of the window because I say there is no information here, you are just confusing the model"*. Overall, as one of the data scientists put it, the work of data cleaning and preparation "*does not advance the prognosis, but is a foundation that you have to create to do [the prognosis].*"

We would see these accounts from data scientists not only as factual descriptions of their work at the coalface, but also as constructions of their identity as grounded in 'solid' and indeed somewhat mundane data work. By engaging in such identity constructions, data scientists oppose the hyped characterizations of their occupation as magicians or wonder boys/girls that at times circulate in the popular discourse or in organizations.

Managing identity challenges in the workplace

It became evident in our interviews that hype and ambiguity do not only feature in the media and public sphere. They also find their way into organizations, where they create work-related challenges for the data scientists. These challenges materialize not least in the form of identity challenges, threatening a particular way of seeing themselves and being seen by others. In response to such identity threats at the workplace (Petriglieri, 2011), data scientists re-emphasize their aspired identity, while also undertaking actions to mitigate such challenges.

Ambiguous or wrong expectations by managers

The ambiguity in the use of the 'data scientist' term translates into organizations in two ways. First, interviewees experience ambiguity in terms of their job descriptions or job profiles and in the corresponding expectations among managers what precisely a data scientist should be doing. Such ambiguity is not a huge problem, but it requires data scientists to actively craft their job in line with their aspired identity. 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, that 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 role and identity in an organization. Such identity work (Sveningsson & Alvesson, 2003) can be observed in many contexts, but is particularly important when the organization provides only vague job descriptions and role models to the employee, as in this case.

A second, and more problematic implication of such ambiguity is that managers sometimes specify the data science role in a way that is at odds with the data scientists' aspirational identity (Thornborrow & Brown, 2009), leaving only little freedom for the latter to craft their job themselves. In particular, several interviewees report that they (at times or regularly) are 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 data analysts or similar data workers.

Interviewee 9 recounts a recent 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*". Accordingly, the interviewee feels that there is really 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 department, but that request was quickly "*outsourced to the reporting department"* as they felt that it was not within their area of responsibility.

Such examples show that organizations at times still struggle with defining the role of data scientists in such a way that it matches those employees' skills and aspired identity. When data scientists are consistently confronted with expectations that do not align with how they make sense of their role and when there is little freedom for job crafting, they may become frustrated, as interviewee 16 explains. Talking about his previous job, he 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*".

The examples provided above show how data scientists, when confronted with requests that they consider inappropriate, address such a challenge by communicating with managers and managing their expectations. To be sure, some of our data scientists acknowledge that not all of their activities are 'pure' data science and that it is alright to sometimes engage in such other activities, as long as they remain the exception rather than the rule. As interviewee 23 suggested:

> *"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)

A somewhat broader job profile is more likely in those cases where data scientists work alone in that role within their organization or department, without being embedded in a larger team of data scientists. The bigger the team, the more specialized job profiles tend to become.

A persistent mismatch between aspirations and actual work tasks can motivate data scientists to 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]'.*"

Hyped or unrealistic expectations by managers

We also see evidence that the general hype around data science translates into organizations, in the sense that managers form expectations which, from the perspective of data scientists, are unrealistic. This can become a challenge for their identity as such overly optimistic expectations can lead to disappointment among managers about the possibilities of data science, which in turn can threaten the data scientists' image and standing. Unrealistic expectations relate both to the process of data science and to its expected outcomes.

In terms of the process, a key challenge appears to be that the data science process is a 'black box' to most managers, and much of the work that data scientists carry out is invisible to others. As a result, managers may develop expectations that are misaligned with the work realities of the data scientists, especially so if the hype around data science makes managers want to see quick results. 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 that it is not that easy to aggregate and prepare the data and to make sure that this happens without incurring any errors. Similarly, interviewee 1 tells us 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."* And interviewee 21 confirms, "*I don't think people realize how much time I am spending on just cleaning the data*". Managers would sometimes get impatient and feel 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 problem is that managers do not understand the underlying task "*of making sure that the [data] quality is right to be able to draw the right conclusions*" (3).

We can hence see here how ignorance about the details of the data science activity can lead to expectations among managers which put undue pressure on data scientists, and, therefore, present a potential challenge for their identity. Two other interviewees offer further examples of cases where managers had unrealistic expectations because they did not know all the steps that the data scientist has to go through. Interviewee 11 explains that there was a new source of data in the organization, but data scientists had not even seen the data. "*But 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'*". And interviewee 16 tells us of a recent case in which a sales-related KPI 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 is not in line with the realities of data science work. Rushing to see results is an expression of the hype around data science, and when coupled with ignorance about the details of the data science process, can lead to unrealistic expectations.

The hype is even more visible when it comes to the envisaged outcomes of data science projects. Our interviewees provide several examples of managers having unclear or unrealistic expectations regarding the objectives of such projects. Interviewee 10 explains in this respect that the "*whole machine learning hype*" is "*somewhat damaging*" to his efforts in the organization because internal stakeholders would sometimes have "*obscure ideas regarding what you can do with models*". He regards it as his job 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 the 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).

While, from the interviewed data scientists' perspective, objectives are sometimes unrealistic, in other cases, they feel that there is no clear idea behind a request. Interviewee 9 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*". And interviewee 3 confirms 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*".

To be sure, not all organizations feature the same amount of hype. One of our interviewees (25) explains that, when he was starting his position as a data scientist, the head of the IT department would explain to him that it was not clear how the topic of data science would develop in the organization, i.e., whether it would become more important or eventually wither. This shows that managers in organizations may sometimes actively temper expectations, such that hype does not emerge within the organization.

We could already see from the examples above how data scientists try to manage expectations regarding the process and outcome of data science projects by communicating extensively with managers. This involves endeavors to set realistic expectations regarding the outputs of a project and acknowledging the limitations of data science to counteract the hype that surrounds this phenomenon. 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 says, "*First of all, when someone requests something, I always push back, because I know I will need the time*".

Importantly, data scientists do not only see such communication as an instance of "*educational work*" (5) vis-à-vis managers. They also acknowledge that their technical language may make it difficult for managers to understand what data science can(not) do and that sometimes the problem lies in their understanding of the managers' expectations. Accordingly, data scientists emphasize the need for dialogue and exchange of perspectives. Interviewee 1 suggests in this respect 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*". 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".*

Understanding each other's expectations requires speaking the same language. Interviewee 15 explains that "*when [the data scientist] starts to use technical terms or concepts which only 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*".

Data scientists acknowledge that it is at least partly up to them to bridge the link to the business users. Interviewee 5 tells us 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*". Some interviewees mention that having a business background is an advantage as it facilitates communication with internal stakeholders: "*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*" (2).

We can therefore see that data scientists, to some extent, critically reflect upon a 'pure' technical identity, rooted in informatics and statistics, and emphasize the need to learn the language of the business and immerse oneself into the business context. We suggest interpreting this 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 & Messner, 2019; Ibarra & Barbulescu, 2010).

CONCLUDING DISCUSSION

Fashionable practices and identity challenges

Data science and data scientists feature quite prominently in the general discourse around digitalization. Accounting scholars have also highlighted the transformative potential of this nascent technical occupation (e.g., Moll & Yigitbasioglu, 2019). The main contribution of this paper is to show how the fashionable nature of the data science discourse translates into identity challenges for data scientists in the workplace and how they deal with those challenges. We illustrate how data scientists perceive the discourse on data science as featuring both hype and ambiguity – two essential components of management fashions (Abrahamson, 1991, 1996) – and how they take a critical distance from this discourse when crafting their identity narrative.

Notwithstanding the narrative identity work through which data scientists try to cope with the hyped and ambiguous nature of the data science discourse, our interviews demonstrate that the features of this discourse frequently translate into the workplace, where they create challenges for data scientists to perform their work and enact their aspirational identity. Ambiguity implies that managers, at times, define the data scientist role in the organization in terms that are not in line with the role incumbents' aspirational identities. Prior literature shows that, when faced with such a gap between identity and assigned work tasks, actors try to get rid of or reinterpret these tasks (Morales & Lambert, 2013) or attempt to re-align their sense of self to make it fit these activities (Pratt, Rockmann, & Kaufmann, 2006). Some of our interviewees indeed mentioned that they would seek to avoid tasks which they consider belonging to, for example, data analysts or reporting specialists (positions that they perceive as inferior), by voicing their expectations to managers. However, our empirical material also indicates that a continuing mismatch between an aspirational identity and assigned tasks can lead to frustration or even to data scientists eventually leaving their organizations.

Similarly, we could see how hype materializes in the form of managers' unrealistic expectations regarding the possibilities of data science and a misguided understanding of the data science process and its intricacies (Zbaracki, 1998), eventually forming a challenge for data scientists. The latter often respond to this challenge through 'educational work', trying to create awareness among managers of what data science is and what it can and cannot do. Our interviewees acknowledge in this respect the need to immerse themselves more strongly into the business context. They hence try to enlarge their identity from its 'technical' core to encompass business-related skills and activities to better understand and manage their counterparts' expectations. This indicates that, in addition to what Vaast and Pinsonneault (2021) showed for digital technologies, managers serve as identity referents (see Goodrick & Reay, 2010) for data scientists. 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 specialist role might make it more difficult for data scientists to integrate their work into other organizational practices. Framing themselves as specialists might furthermore signal to management that what data scientists do is so distinct from managers' work that they do not need to engage with the details and intricacies of data science. This might spur the challenges that data scientists experience in their everyday work.

Our interviews revealed in this context that instead of merely identifying differences and similarities between themselves and the managers they are working with, data scientists try to actively *create* similarities with respect to the understanding of the possibilities and limitations of data science. This translates into both inward- and outward-facing identity work (Watson, 2008). Inward-facing, data scientists act upon themselves to better understand the managers' 'thought worlds' (Dougherty, 1992) and incorporate this evolving understanding into the construction of their own occupational identity. The outcome of this seems to be the above-mentioned self-understanding of being more of a generalist rather than a specialist (see Avnoon, 2021). Outward-facing, data scientists try to influence managers' understanding of data science and how they see the role of the data scientist. They regard this as crucial to manage the expectations managers develop towards them and their "tools of the trade". Both forms of identity work can be seen as critical with respect to coping with the ambiguity and hype around data science and the respective workplace challenges that these features of the fashionable discourse translate into.

Overall, we can see how the fashionable nature of data science has a direct influence on data scientists' identity narratives as well as an indirect one through its impact on managers' expectations, which can threaten data scientists' aspirational identities. These observations are interesting insofar as 'fashionable' occupations 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, stigmatized (Ashforth et al., 2007) or that are being challenged by other occupations or through managerial or technological change (Kahl, King, & Liegel, 2016; Nelson & Irwin, 2014; Reay, Goodrick, Waldorff, & Casebeer, 2017). From a discursive perspective, 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 (knowledge work) occupations that experience identity struggles; not least accountants (Goretzki & Messner, 2019; Morales & Lambert, 2013). And yet, our analysis demonstrates that the hype around their occupation, together with the ambiguity around data science, creates its own sets of identity challenges for data scientists. We can therefore suggest that it is not only digital technologies per se (Vaast & Pinsonneault, 2021) but also the hype and ambiguity in the discourse about such technologies and associated practices that have ambivalent consequences for the identity of data scientists. Our study complements previous research on the occupational identity of data scientists (Avnoon, 2021; Vaast & Pinsonneault, 2021) by demonstrating how conditions that one might assume would support their aspirational identity (here a fashionable discourse) paradoxically contribute to its fragility and call for careful identity work. As such, our findings also speak to the management fashion literature that has mainly focused on exploring and explaining the factors affecting the diffusion, implementation, and rejection of new tools, techniques, or practices (see e.g., Abrahamson, 1991, 1996; Baskerville & Myers, 2009; David & Strang, 2006; Kieser, 1997), without discussing the effects of such fashionable practices on employees' identities, however.

Balancing prestigious and mundane aspects of identity

Concerning the identity work carried out by data scientists, our analysis indicates that data scientists try to balance between different (and partly competing) ways of constructing their occupational identity. We could hereby observe two orientations in their identity: On the one hand, data scientists would emphasize the prestigious parts of their work (i.e., data modelling) when they react to the ambiguity in their job title and amalgamations of different types of job profiles, some more 'sophisticated' than others. In doing so, they present themselves as an 'elite group' among data workers. At the same time, and somewhat paradoxically, they also try to create awareness for the more mundane parts of their job. They do so in response to the hyped expectations that circulate in the broader discourse and to the corresponding unrealistic expectations of managers who would often lack knowledge about data science. In emphasizing their involvement in data collection, preparation, and cleaning, they engage in what can be referred to as 'de-hyping', i.e., they draw a picture of their work that they believe is – though less 'heroic' – more representative of what they (can) do. In doing so, they highlight that they are not 'data wizards' performing 'magic', but indeed have to carry out rather mundane tasks that – although forming a large part of their work – often remain invisible for managers. In addition, they emphasize the limitations of their craft and try to tone down managers' expectations towards their work and the contributions it can make to organizational processes.

These two manifestations of data scientists' identity work are motivated by different concerns with the fashionable discourse (i.e., with hype and ambiguity, respectively) and, to some extent, compete in the sense that they point into different directions for their occupational identity. One can argue that this situation creates some kind of identity paradox for data scientists who try to present themselves as the 'top of the class' among the data workers, while at the same time feeling a need to emphasize the less prestigious aspects and limitations of their work to manage the managers' unrealistic expectations about their work and its outcomes. We propose that this paradoxical form of identity work is a specific feature of a fashionable occupation that faces both ambiguity and hype in the discourse around their work.

In this sense, data scientists seem to experience a situation that is different from other occupations like accountants that lack such a fashionable aura. Studies show, for example, that in some geographical contexts, controllers had to legitimize their existence as a distinct occupational group and to prove that they can make an original contribution to the management of the company (Goretzki, Löhlein, Schäffer, Schmidt, & Strauss, 2021). In contrast to that, due to the fashionable nature of the data science discourse, reputation seems to precede data scientists. Similarly, while accountants, as an established occupation, sometimes need to work against negative images or stigmatizing stereotypes (see e.g., Jeacle, 2008) and face threats of becoming rationalized through computerization (see Frey & Osborne, 2017), data scientists – despite being a nascent occupation – enjoy a reputation of being 'sexy' and important. Thus, legitimization and justification of their work and role seems less of an issue for them.

As we demonstrate in this paper, this situation however creates specific challenges for data scientists 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 discourse, data scientists typically do not need to (pro-)actively create an image of being important and exciting. They hence do not try to hide tasks that are not in line with their aspired identity from others (cf. Morales & Lambert, 2013). Rather, data scientists deliberately choose to render these tasks visible to their organizational counterparts as part of their outward-facing identity work. They seem hereby to have a pragmatic attitude towards those tasks (see Carter & Sholler, 2016). Putting on display that tasks like data cleaning form an important though non-prestigious part of their work is a form of identity work through which 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 stronger identity threats, such as when managers start to realize that the data scientists are not able to live up to the hype.

Limitations and future research

The analysis presented in this paper is focused on the perspectives of data scientists and we identified other actors' expectations through the eyes of our interviewees. Although this makes sense since 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 & Sholler, 2016). A focus on specific workplaces (Bechky, 2011) may thereby reveal organizational context factors that facilitate (or challenge) 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 & 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.

References

Abrahamson, E. (1991). Managerial fads and fashions: The diffusion and rejection of innovations. *Academy of Management Review, 16*(3), 586-612.

Abrahamson, E. (1996). Management fashion. *Academy of Management Review, 21*(1), 254-285.

Al-Htaybat, K., & von Alberti-Alhtaybat, L. (2017). Big Data and corporate reporting: impacts and paradoxes. *Accounting, Auditing & Accountability Journal, 30*(4), 850-873.

Alvesson, M., & Willmott, H. (2002). Identity regulation as organizational control: Producing the appropriate individual. *Journal of Management Studies, 39*(5), 619-644.

Ansari, S. M., Fiss, P. C., & Zajac, E. J. (2010). Made to fit: How practices vary as they diffuse. *Academy of Management Review, 35*(1), 67-92.

Arnaboldi, M., Busco, C., & Cuganesan, S. (2017). Accounting, accountability, social media and big data: revolution or hype? *Accounting, Auditing & Accountability Journal, 30*(4), 762-776.

Ashcraft, K. L. (2007). Appreciating the 'work'of discourse: Occupational identity and difference as organizing mechanisms in the case of commercial airline pilots. *Discourse & Communication, 1*(1), 9-36.

Ashforth, B. E., Kreiner, G., E. , Clark, M. A., & Fugate, M. (2007). Normalizing dirty work: Managerial tactics for countering occupational taint. *Academy of Management Journal, 50*(1), 149-174.

Avnoon, N. (2021). Data Scientists' Identity Work: Omnivorous Symbolic Boundaries in Skills Acquisition. *Work, Employment and Society, 35*(2), 332-349.

Barbour, J. B., Treem, J. W., & Kolar, B. (2018). Analytics and expert collaboration: How individuals navigate relationships when working with organizational data. *Human Relations, 71*(2), 256-284.

Baškarada, S., & Koronios, A. (2017). Unicorn data scientist: the rarest of breeds. *Program, 51*(1), 65-74.

Baskerville, R. L., & Myers, M. D. (2009). Fashion waves in information systems research and practice. *MIS Quarterly, 33*(4), 647-662.

Bechky, B. A. (2011). Making organizational theory work: Institutions, occupations, and negotiated orders. *Organization Science, 22*(5), 1157-1167.

Benders, J., van den Berg, R. J., & van Bijsterveld, M. (1998). Hitch-hiking on a hype: Dutch consultants engineering re-engineering. *Journal of Organizational Change Management, 11*(3), 201-215.

Benders, J., & Van Veen, K. (2001). What's in a fashion? Interpretative viability and management fashions. *Organization, 8*(1), 33-53.

Berger, P. L., & Luckmann, T. (1966). *The social construction of reality: A treatise in the sociology of knowledge*: Anchor.

Bhimani, A., & Willcocks, L. (2014). Digitisation,'Big Data'and the transformation of accounting information. *Accounting and Business Research, 44*(4), 469-490.

Brown, A. D. (2015). Identities and identity work in organizations. *International Journal of Management Reviews, 17*(1), 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, 23*(3), 598-622.

Carter, D., & Sholler, D. (2016). Data science on the ground: Hype, criticism, and everyday work. *Journal of the Association for Information Science and Technology, 67*(10), 2309-2319.
Chapman, C., Chua, W. F., & Fiedler, T. (2021). Seduction as control: Gamification at Foursquare. *Management Accounting Research*, in press.
Collinson, J. A. (2006). Just 'non-academics'? Research administrators and contested occupational identity. *Work, Employment and Society, 20*(2), 267-288.
Davenport, T. H., & Patil, D. (2012). Data scientist. *Harvard Business Review, 90*(5), 70-76.
David, R. J., & Strang, D. (2006). When fashion is fleeting: Transitory collective beliefs and the dynamics of TQM consulting. *Academy of Management Journal, 49*(2), 215-233.
Dougherty, D. (1992). Interpretive barriers to successful product innovation in large firms. *Organization Science, 3*(2), 179-202.
Down, S., & Reveley, J. (2009). Between narration and interaction: Situating first-line supervisor identity work. *Human Relations, 62*(3), 379-401.
Eccles, R. G., Nohria, N., & Berkley, J. D. (1992). *Beyond the hype: Rediscovering the essence of management*: Beard Books.
Fenn, J., & Raskino, M. (2008). *Mastering the hype cycle: how to choose the right innovation at the right time*: Harvard Business Press.
Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change, 114*, 254-280.
Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137-144.
Gehl, R. W. (2015). Sharing, knowledge management and big data: A partial genealogy of the data scientist. *European Journal of Cultural Studies, 18*(4-5), 413-428.
Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*: Univ of California Press.
Giddens, A. (1991). *Modernity and self-identity: Self and society in the late modern age*: Stanford university press.
Giroux, H. (2006). 'It was such a handy term': Management fashions and pragmatic ambiguity. *Journal of Management Studies, 43*(6), 1227-1260.
Goodrick, E., & Reay, T. (2010). Florence Nightingale endures: Legitimizing a new professional role identity. *Journal of Management Studies, 47*(1), 55-84.
Goretzki, L., Löhlein, L., Schäffer, U., Schmidt, A., & Strauss, E. (2021). Exploring the Role of Metaphors in Social-Identity Construction: The Case of the German Controller. *European Accounting Review*, in press.
Goretzki, L., & Messner, M. (2019). Backstage and frontstage interactions in management accountants' identity work. *Accounting, Organizations and Society, 74*, 1-20.
Harris, H., Murphy, S., & Vaisman, M. (2013). *Analyzing the analyzers: An introspective survey of data scientists and their work*: O'Reilly.
Harris, J. G., & Mehrotra, V. (2014). Getting value from your data scientists. *MIT Sloan Management Review, Fall 2014*, 15-18.
Jeacle, I. (2008). Beyond the boring grey: The construction of the colourful accountant. *Critical Perspectives on Accounting, 19*(8), 1296-1320.
Jenkins, R. (2014). *Social identity*: Routledge.

Jung, N., & Kieser, A. (2012). Consultants in the management fashion arena. In M. Kipping & T. Clark (Eds.), *The Oxford handbook of management consulting* (pp. 327-346). New York: Oxford University Press.

Kahl, S. J., King, B. G., & Liegel, G. (2016). Occupational survival through field-level task integration: Systems men, production planners, and the computer, 1940s–1990s. *Organization Science, 27*(5), 1084-1107.

Kieser, A. (1997). Rhetoric and myth in management fashion. *Organization, 4*(1), 49-74.

Madsen, D. Ø., & Stenheim, T. (2016). Big Data viewed through the lens of management fashion theory. *Cogent Business & Management, 3*(1), 1165072.

Moll, J., & Yigitbasioglu, O. (2019). The role of internet-related technologies in shaping the work of accountants: New directions for accounting research. *The British Accounting Review, 51*(6), 100833.

Morales, J., & Lambert, C. (2013). Dirty work and the construction of identity. An ethnographic study of management accounting practices. *Accounting, Organizations and Society, 38*(3), 228-244.

Muller, M., Lange, I., Wang, D., Piorkowski, D., Tsay, J., Liao, Q. V., . . . Erickson, T. (2019). *How data science workers work with data: Discovery, capture, curation, design, creation.* Paper presented at the Proceedings of the 2019 CHI conference on human factors in computing systems.

Murphy, C., & Kreiner, G. E. (2020). Occupational boundary play: Crafting a sense of identity legitimacy in an emerging occupation. *Journal of Organizational Behavior, 41*(9), 871-894.

Nelson, A. J., & Irwin, J. (2014). "Defining what we do—all over again": Occupational identity, technological change, and the librarian/Internet-search relationship. *Academy of Management Journal, 57*(3), 892-928.

Oesterreich, T. D., & 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 & Organizational Change, 15*(2), 330-356.

Pachidi, S., Berends, H., Faraj, S., & Huysman, M. (2021). Make way for the algorithms: Symbolic actions and change in a regime of knowing. *Organization Science, 32*(1), 18-41.

Petriglieri, J. L. (2011). Under threat: Responses to and the consequences of threats to individuals' identities. *Academy of Management Review, 36*(4), 641-662.

Pratt, M. G., Rockmann, K. W., & 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, 49*(2), 235-262.

Quattrone, P. (2016). Management accounting goes digital: Will the move make it wiser? *Management Accounting Research, 31*, 118-122.

Reay, T., Goodrick, E., Waldorff, S. B., & Casebeer, A. (2017). Getting leopards to change their spots: Co-creating a new professional role identity. *Academy of Management Journal, 60*(3), 1043-1070.

Reed, C., & Thomas, R. (2021). Embracing indeterminacy: on being a liminal professional. *British Journal of Management, 32*(1), 219-234.

Richins, G., Stapleton, A., Stratopoulos, T. C., & Wong, C. (2017). Big data analytics: Opportunity or threat for the accounting profession? *Journal of Information Systems, 31*(3), 63-79.

Saifer, A., & Dacin, M. T. (2021). Data and Organization Studies: Aesthetics, emotions, discourse and our everyday encounters with data. *Organization Studies*, in press.
Strauss, A., & Corbin, J. (1998). *Basics of qualitative research techniques*: Citeseer.
Sveningsson, S., & Alvesson, M. (2003). Managing managerial identities: Organizational fragmentation, discourse and identity struggle. *Human Relations, 56*(10), 1163-1193.
Thornborrow, T., & Brown, A. D. (2009). Being regimented': Aspiration, discipline and identity work in the British parachute regiment. *Organization Studies, 30*(4), 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, 41*(12), 1673-1695.
Vaast, E., & Pinsonneault, A. (2021). When Digital Technologies Enable and Threaten Occupational Identity: The Delicate Balancing Act of Data Scientists. *MIS Quarterly, 45*(3), 1087-1112.
Van den Bussche, P., & Dambrin, C. (2020). Peer-to-peer evaluations as narcissistic devices: fabricating an entrepreneurial community. *Accounting, Auditing & Accountability Journal*.
Waller, M. A., & 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 (Vol. 34, pp. 249-252): Wiley Online Library.
Watson, T. J. (2008). Managing identity: Identity work, personal predicaments and structural circumstances. *Organization, 15*(1), 121-143.
Wright, C., Nyberg, D., & Grant, D. (2012). “Hippies on the third floor”: Climate change, narrative identity and the micro-politics of corporate environmentalism. *Organization Studies, 33*(11), 1451-1475.
Wrzesniewski, A., & Dutton, J. E. (2001). Crafting a job: Revisioning employees as active crafters of their work. *Academy of Management Review, 26*(2), 179-201.
Zbaracki, M. J. (1998). The rhetoric and reality of total quality management. *Administrative Science Quarterly, 43*(3), 602-636.

Tables

Table 1: Synthesis of our findings

Components of the fashionable nature of data science	**Narrative identity work**	**Challenges faced in the organization**	**Interactive identity work**
Ambiguity of data science	“More than a data analyst” • Emphasizing the modelling part of their job • Highlighting the need for sophisticated skills and for experience and creativity	Ambiguous or wrong expectations • Vague or ambiguous job descriptions • Being given tasks that are outside their remit (dashboarding etc.)	Defining one’s role • Communicating with managers to set right expectations • Active choice of projects • Leaving the company if persistent mismatch
Hype around data science	“Not MacGyver, not a magician” • Emphasizing the mundane parts of their job (data cleaning etc.)	Unrealistic expectations • Too high expectations regarding outcome • Ignorance about the process of data science	Aligning expectations with managers • Educational work • Learning the language of business • Frequent dialogue

Appendix: List of interviewees

#	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 & 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 & 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