Check for updates

Article

Current Sociology
2024, Vol. 72(6) 1107–1125
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00113921231166147
journals.sagepub.com/home/csi
SAGE

Replace, absorb, serve: Data scientists talk about their aspired jurisdiction

Netta Avnoon
Tel Aviv University, Israel; Columbia University, USA

Abstract
How do data scientists frame their relations with domain experts? This study focuses on data scientists' aspired professional jurisdiction and their multiple narratives regarding data science's relations to other fields of expertise. Based on the analysis of 60 open-ended, in-depth interviews with data scientists, data science professors, and managers in Israel, the findings show that data scientists institutionalize three narratives regarding their relations with domain experts: (a) replace experts, (b) absorb experts' knowledge, and (c) provide a service to experts. These three narratives construct data scientists' expertise as universal and omnivorous; namely, they are relevant to many domains and allow data scientists to be flexible in their claim for authority.

Keywords
Algorithms, artificial intelligence, data science, expertise, interprofessional relations, professional jurisdiction, professions

Introduction

Artificial intelligence (AI) is a new technology that has now entered many areas of life. One important facet of the rise of AI is its impact on expert labor in professions such as finance, medicine, the social sciences, and law (Capogna, 2020; Hansen, 2021; Susskind and Susskind, 2015). However, despite their autonomous image, algorithms do not build themselves (Gillespie, 2014; Hansen, 2021); rather, a new profession is forming to construct them (Avnoon, 2021; Brandt, 2016; Dorschel, 2021; Ribes, 2019). Data science is

Corresponding author:
Netta Avnoon, Coller School of Management, Gershon H. Gordon Faculty of Social Sciences, Tel Aviv University, Ramat Aviv, Tel Aviv 6997801, Israel.
Email: nettaa@tauex.tau.ac.il

a nascent profession focused on the computational processing of big data, and along with other emergent titles for practitioners in this field (e.g. ‘machine learning architect’, ‘algorithms developer’, and ‘AI researcher’), it has been integrated into the system of professions and has been carving a jurisdiction for its expertise since the early 2000s (Brandt, 2016; Dorschel and Brandt, 2021). Data science is being institutionalized by repeatedly and habitually constructing identities, establishing a community, and structuring a wide network of interprofessional relations with what data scientists call ‘domain experts’[1] (Ribes, 2019). Unlike prior efforts of quantification (Bruno et al., 2016), data scientists view their expertise as ‘domain agnostic’[2]; therefore, they do not limit their expertise to one disciplinary field (Brandt, 2016).

This article investigates how data scientists as members of a nascent profession construct and negotiate their aspired professional jurisdiction. Based on 60 open-ended, in-depth interviews with data scientists and data scientists’ professors and managers in Israel, I outline three narratives data scientists use to frame their relations with other experts, examine data scientists’ dynamic discursive alternation between these narratives, and discuss the role of this maneuvering in professional positioning. These narratives are as follows: (a) to replace other experts, (b) to absorb experts’ knowledge and expertise, and (c) to provide a service for other experts. I show how all three narratives construct data science expertise as a universal expertise that is relevant to many domains (Avnoon, 2021; Ribes, 2019), while also allowing flexibility in relations with members of existing professions in terms of authority. Thus, this multiplicity supports data scientists’ efforts to widen their aspired professional jurisdiction.

Interprofessional relations, jurisdictions, and the construction of expertise

Traditionally, jurisdiction has been defined in the sociology of professions as the link between a profession and a set of tasks and problems (Abbott, 1988: 20). However, establishing jurisdiction for a group of experts involves more than coming up with unique and esoteric knowledge on tasks and problems. Forming relations with other professional groups and clients in the process of attributing the ability to perform tasks and solve problems to a certain professional group is the main challenge in a jurisdictional claim (Anteby et al., 2016; Eyal, 2013; Freidson, 1970). This means that to establish a viable profession, expertise must be imputed to its members by other actors in the social world such as clients, managers, and other professional groups (Alvesson, 2001). Thus, persuasion and legitimization by forming relations are prominent efforts for the creation of jurisdiction by professions (Kahl et al., 2016).

In the past, the sociology of professions tended to highlight aspects of dominance and monopoly as a form of professional social control and the only path to a viable jurisdiction (Abbott, 1988; Freidson, 1970; Larson, 1977). However, in recent years, sociologists point to new social constructions of expertise, focusing more on its networked and relational aspects (Eyal, 2013; Muzio et al., 2011; Noordegraaf, 2020). For example, Huising (2015) showed that attentiveness and keeping in close contact with professional clients by doing scut work allowed health physicists to claim *relational authority* over their professional clients. At the same time, Huising found that bio-safety officers’ withdrawn and removed relational strategy resulted in their loss of connection with their professional clients and loss of jurisdiction. Kahl et al. (2016) found that an integrative strategy to professional jurisdiction – defining their task domains with relations and dependencies on other professional groups – led production managers to occupational survival. At the same time, boundary work and the inward focus of system men led to their disappearance.

In this 'relational turn' (Anteby et al., 2016), it might seem that cooperation, coproduction, and generosity have replaced competition, monopoly, and dominance as the main features of a successful professional project. However, others argue that contemporary professions can simultaneously enact protective, competitive, and dominant strategies for monopolizing expertise, while also connecting with other professions and clients, generously coproducing with them, and even complying with their fellow professions' or clients' needs (Eyal, 2013; Noordegraaf, 2020). These nuanced perspectives on experts and expertise are complicating the study of social relations in the realm of expertise. In this regard, the study of data science, a nascent profession that aims to establish a wide network of relations with many different stakeholders, is especially interesting.

The problem of a wide jurisdiction

In his discussion of 'The Chaos of Disciplines', Abbott (2001) argued that *interstitial professions* – those that are 'not very good at excluding' problems from their expertise – are especially vulnerable to encroachment (p. 5). This means that applying expertise widely and, thus, forming an extensive network of interprofessional relations present special risks for a profession. For example, Abbott (1988) discussed the statistics profession, which, like data science, also focuses on quantitative data analysis. However, Abbott argued that statistics lends its methods to a wide array of professions by embedding it in software or teaching it to other professions; thus, its intellectual jurisdiction becomes vulnerable to other professions claiming authority over the practice of statistics (Abbott, 1988: 236). Recent research has confirmed Abbott's hypothesis of the vulnerability of wide jurisdictions. Most prominently, Eyal (2013) argued that establishing a network of expertise – the structuration of a cooperative network of connections – although allowing the expertise to flourish, may leave the experts vulnerable to the authority of other experts (see also Kahl et al., 2016).

Allegedly, data science should have been as vulnerable as statistics to the authority of other experts. As Ribes (2019) argued, data scientists frame their expertise as a '*universal science*, essential to all future sciences and beyond' (p. 516). In a similar vein, Avnoon (2021) showed that data scientists construct their expertise based on symbolic boundaries of expansion rather than specialization. A question then arises as to what happens when a group of experts, such as data scientists, claim an interstitial, universal, omnivorous jurisdiction? How can authority be maintained when expertise is applied widely? Inspired by Goffman (1959) and sociological notions of agent–structure duality (Giddens, 1984), and the study of speech as organizing the social world (Lamont and Swidler, 2014), this study adds a micro-angle to the discussion on data science's universal logic and ties it to the discussion of competition/cooperation in the sociology of expert professions. Focusing on data scientists' own ideas regarding their relations to 'domain experts' and the way they talk about these ideas, I, therefore, ask, what is data scientists' aspired jurisdiction, and how do they position their expertise in relation to other experts and expertise? Findings show that data scientists experiment with and alternate between contesting relational narratives of both competition and cooperation. This way, they institutionalize several *options* for their relations with other professional groups and their aspired jurisdiction. The multiplicity of options enables them to experiment with varying levels of authority and adjust their claim of control according to circumstances.

Research context

With more than 6600 high-tech and start-up companies and more than 300 research and development (R&D) branches of international corporations, Israel has a well-developed high-tech industry (Korbet, 2019). Most of these companies are located in the greater Tel Aviv area in the center of the country, forming what is known as the Silicon Wadi, the dense Israeli tech ecosystem. Accordingly, Israel has one of the highest density rates of data scientists (RJ Metrics Survey, 2016), and the Israeli data science professional community is active and vibrant. It holds daily meetups, community hackathons, and boot camps, and it hosts some of the AI and data science summits. Israeli data scientists also participate in the online activity of data science global Internet platforms, such as the data science competition website Kaggle, and online programming and statistics websites, such as Stack Overflow, Stack Exchange, and Cross Validated. They are also active on both local and global data science–designated groups on Facebook, LinkedIn, and other social networks.

Methods

The empirical data ($n=60$) were gathered during a research project that investigated the emergence of data science as a nascent technical profession. Data collection took place in Israel between 2016 and 2017. The 60 interviews were conducted with three categories of interviewees: 50 data scientists, 5 professors who train data scientists, and 5 managers of data scientists, not necessarily data scientists themselves. Sampling focused on individuals who defined themselves as data scientists on the professional social network LinkedIn. At the time of data collection, there were less than 1000 results for the keywords ‘data scientist Israel’, and the growth rate of the community was about one person per day. As of the time of writing (October 2022), there were 23,000 results for the search ‘data science Israel’, which shows that data science is a rapidly growing profession (see more on the international data science profession in Brandt, 2016, and Dorschel, 2021, and self-reports of membership count in Kaggle Survey, 2019). A total of 125 data scientists were contacted on LinkedIn based on their occupational title, and 46 agreed to participate – indicating a relatively high response rate of 37%. Another four data scientists were snowball sampled from interviewees’ recommendations.

The workers’ ages varied from their 20s to 60s, with the largest group being in their 30s (see Figure 1). Five interviewees were women, and 45 were men (see Figure 2). This ratio is in line with the international gender ratio of data scientists (Kaggle Survey, 2019) and of other STEM (science, technology, engineering, and mathematics) professions

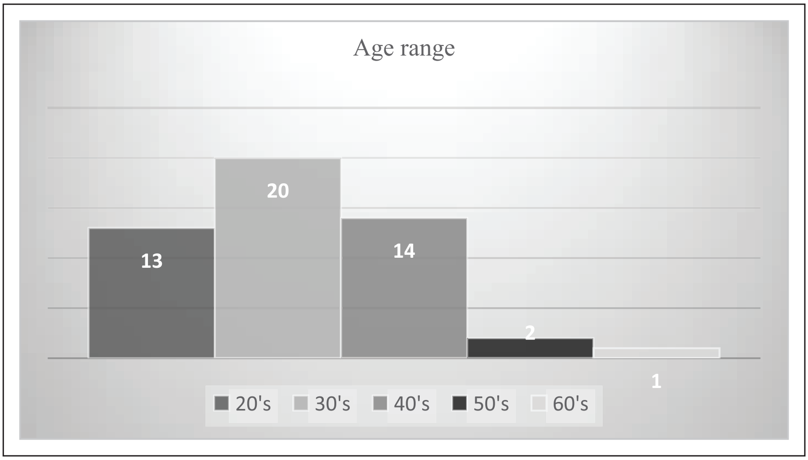

Figure 1. Age range of the sample (excluding professors and managers).

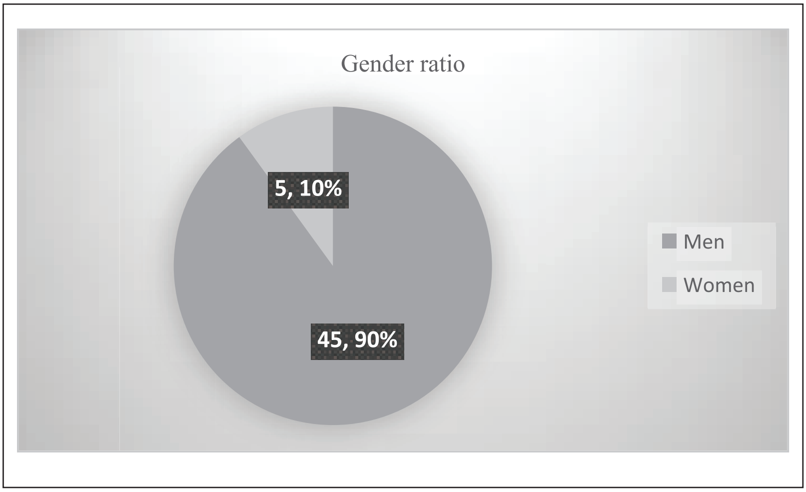

Figure 2. Gender ratio of the sample (excluding professors and managers).

(Wajcman, 2007). The education level of this group varied from having only a bachelor's degree (BSc) to postdoctoral training (see Figure 3).

The second group comprised five professors who train data scientists at four leading Israeli universities and technology institutes. This group was snowball sampled based on interviewee recommendations. It comprised one woman and four men. The professors were from the computer science, software engineering, industrial engineering, statistics, and management departments. Excluding management, all of the departments were located at either engineering or exact science schools. (In Israel, math, physics, chemistry, earth science, and computer science are academically distinct from engineering.) At

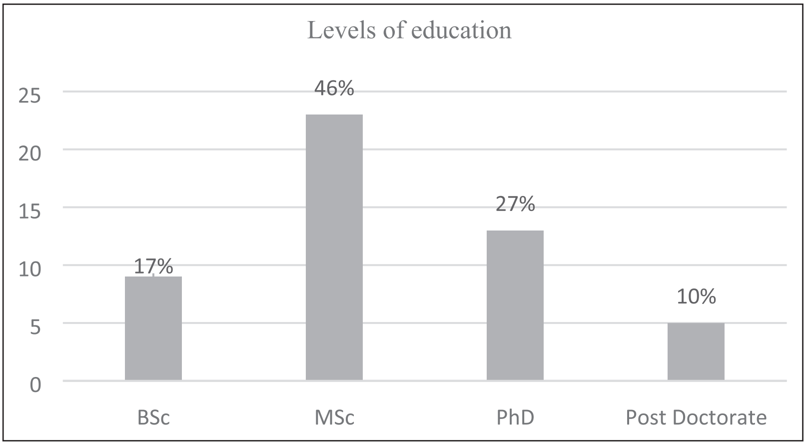

Figure 3. Levels of education (excluding professors and managers).

the time of the research, data science schools, departments, and programs were being established across the Israeli academic system, and all professors in the sample were part of these efforts at their institutions. The third group comprised five managers of data science teams. This group was also snowball sampled based on interviewees' recommendations. It comprised five men: Two of them had their own start-ups, and three were hired managers.

Interviews were open-ended and conducted in Hebrew by the author. They lasted between 1 and 3 hours and were recorded and transcribed. The names of people and companies were anonymized. Interviewees were encouraged to talk about different aspects of their professional life, such as their work process, skills, community, and relations with other experts. I challenged them occasionally regarding their views and asked them to elaborate and provide examples as the conversation evolved (Lamont and Swidler, 2014).

Analysis of the transcribed interviews was conducted according to grounded theory principles (Strauss and Corbin, 1997). Viewing speech as a socially embedded act, I aimed to trace data scientists' continuous sense-making talk on their professional relations (Lamont and Swidler, 2014; Millerand et al., 2013). I first read and reread each interview several times and categorized the relations described therein. For example, 'We have analysts here who work to make themselves redundant' (the workers are labeling for algorithms) was categorized as 'Relations with domain experts – embedding domain knowledge in technological systems', and 'In my opinion, doctors should not diagnose or treat patients, we do this better than them' was categorized as 'Relations with domain experts – outperforming experts'. By the end of this stage of analysis, I had a tree of categories and their matching quotes. Categorization revealed tensions in the relations between data science and other professions, their claim to superiority and authority over domain experts, their desire to serve, and their insecurities and dependencies on domain experts' knowledge. In the second stage of analysis, I assembled categories into themes.

At this stage, I brought into the analysis concepts from the literature on algorithms, expertise, and interprofessional relations, such as power, authority, competition, value, cooperation, generosity, and coproduction (Abbott, 1988; Eyal, 2013; Hansen, 2021; Lange et al., 2019; Ribes, 2019). In the third stage of analysis, I returned to the interviews and studied the frequency, sequence, and switching points between narratives in data scientists' sense-making talk to trace the dynamics of their perspectives. At this stage, it became clear that data scientists usually operate more than one relational narrative and that coherence is achieved by seeing data science as all-encompassing and universal (Avnoon, 2021; Ribes, 2019), one that may be applied to all domains with various levels of authority. In terms of switching, each narrative had its own function in data scientists' sense-making talk.

Findings

The research participants cultivated three narratives for their relations with domain experts, as a discursive form of institutionalization: replace domain experts, absorb, or serve them. Each narrative conveys a different set of power relations with domain experts (Hansen, 2021) and communicates a different aspired jurisdiction. Therefore, I found that each narrative was used differently in data scientists' sense-making talk. See Table 1 for the taxonomy of narratives and their function in data scientists' talk.

Replace

Data scientists talk about replacing experts in two distinct speech events in the interviews: (a) while referring to the far future, as well as to the potentials, hopes, and risks of their expertise and of the technologies they are developing, and (b) while telling war stories, which either happened to them or other data scientists. These stories give a touch of reality to the idealist conviction that algorithms, such as AI, are going to replace expert work (Hansen, 2021).

Potentials and the far future. Replying to a question about the impact of data science on the labor market, Alma, a data scientist with a PhD in finance and two degrees in computer science who works for a big fintech company, said,

> [In terms of] employment, I think there is going to be a problem. I think that machines will start to replace people; we are already seeing it happen. I do not think it is happening fast enough [. . .] I think for a lot of people, all kinds of accountants and lawyers – not the very best, maybe, but those who do the paperwork – slowly the machines will replace [their work].

Alma focused specifically on the learned professions, such as accounting and law, and foretold their replacement by automated systems. Her only reservation was that the very best professionals would survive this change. This type of rhetoric suggests that intelligent machines will dominate the labor market sometime in the future, in wishful anticipation of technological triumphs (Takhteyev, 2012). Some data scientists felt that their core skills and abstract knowledge of machine learning were superior to other methods

Table 1. Narratives and their functions in data scientists' talk.

Perspective	Replace	Absorb	Serve
Function in talk	Potentials and the far future War stories: 1. Myths 2. Personal war stories	Talking about specific domains: 1. Assimilating domain knowledge into systems 2. Demoting experts 3. Creeping training	Real-world interactions with domain experts Limitations: 1. Limitation of the technology 2. Responsibility

of data analysis and the abilities of domain experts. They told war stories to prove this superiority.

War stories. The data scientists in the study told two kinds of stories: mythical tales of the triumphs of AI and their personal experiences. These war stories portray the technology created by data scientists as superior to experts without paying much attention to the details, such as the involvement of experts in the development process or the locality of the technology's success. These stories convey the message that data science is here to replace experts. Typical war stories concern AI's winnings in chess and Go games. Others are about outperforming experts in specific tasks. For example, when presented with a question regarding the impact of data science on the labor market, Hillel, who has a PhD in physics and works as a data scientist for a consultancy firm, recounted a mythical story. This story recounts how algorithms have replaced experts in linguistics tasks:

> Take text to speech. For a long time, it was clear that you take linguists, and they say there are 54 phonemes and you just needed [the algorithm] to practice each phoneme. But then someone says, 'Come on, let us say there are not [54 phonemes]'. 'You tell me [addressing the computer] how many there are?' and then [the algorithm] says, 'I work with 14 [phonemes]'. 'Why?' 'Because'. And by using this, [the algorithm] does a better job.

Hillel discussed the computer's task of converting speech to text where a computer types speech instead of a human. For many years, linguists advised computer scientists on this task and divided speech into 54 phonemes. They used this abstract linguistic division to train computers to convert the speech signal into text. At some point, computer scientists said that the algorithms would work better and convert speech to text more efficiently without the linguistic division into phonemes. The computer scientists allowed the computer (an unsupervised algorithm) to assign sounds to groups, and the algorithm found it easier to divide speech sounds into 14 groups rather than 54. Linguistic knowledge of division into phonemes was rejected while machine learning using unsupervised algorithms delivered the goods. This is an example of how, from the perspective of data scientists, knowledge of algorithms will replace existing knowledge systems.

Idan, a data scientist with a PhD in computer science, works for a large high-tech company. The stories he told in the interview about data science triumphs were not mythical but taken from his workplace experience, for example,

> I will give you a very concrete example. There was a client, and one of the teams tried to predict the size of the workforce the company would need in the next six quarters. There was an exercise the client held in each quarter: He had a group of eight experts sit for two weeks to discuss the next six quarters. They produced a kind of forecast. They produced all kinds of Excel charts and analyzed them. And she [his teammate, who was a data scientist] wrote an algorithm that took exactly two minutes to run. It provided a forecast that was 10 to 15 percent more accurate than that of the experts. Eight people worked for two weeks. Sixteen-man weeks, experts, serious people.

Idan described what he sees as the superiority of algorithms. Telling a David and Goliath tale (the female data scientist vs the group of eight financial experts), he communicated how data scientists' algorithms are superior to experts' discretion. The algorithm that Idan's teammate wrote produced a more accurate forecast in less time than the calculations supplied by the group of analysts. One can construe from Idan's story that the eight analysts' jobs were at risk because their forecasts were less accurate than those produced easily and efficiently by the algorithm.

According to the narrative of *replace*, domain knowledge is inferior to machine learning algorithms, and knowledge of algorithms is competitive and will dominate any other type of knowledge (Hansen, 2021). Here, data scientists' expertise is portrayed as superior, and their relational intention is to monopolize various domains.

Absorb

In contrast to the futuristic anticipation of replacing other professions and the competitive zeal of war stories, while discussing specific domains such as medicine, marketing, or law, data scientists usually use the narrative of *absorb* for their relations with domain experts. In this narrative, expert knowledge is important and is required for algorithms to succeed. This narrative promotes cooperation and coproduction with domain experts to extract domain knowledge and expertise (Ribes, 2019). However, according to this narrative, data scientists predict that domain knowledge can be subsumed by data science in various ways and that the experts themselves will vanish. Data scientists described three ways in which specific domain knowledge can be absorbed into data science: (a) when data science 'extracts' expert knowledge and assimilates it into automatic systems through labeling and consultation until the experts become redundant, (b) when data science reconstructs the content and value of the domain experts' work and relegates experts to a lower position in the division of labor and (c) when the domain experts' training changes beyond recognition to include data science.

Assimilating expert knowledge into automatic systems. In specific domains, expert knowledge becomes assimilated into automated systems – for example, when radiologists' reports are used as labels to create an algorithmic diagnosis or when Internet content is sorted by editors and journalists to automate editing decisions. Data scientists use expert knowledge in three stages of their work process: while labeling examples for an algorithm, while creating features of the algorithm (the variables it uses for analysis), and while testing the model. Some data scientists thought that once expert knowledge is assimilated into labels and features, or while testing a model, the experts' days are numbered – despite having a good working relationship with the experts and valuing their knowledge. Their expertise is temporary – lasting until the algorithm learns from them, outperforms them, and makes them redundant. For example, Alex, a data scientist with a large cyber security company, said the following about the field of radiology:

> Radiologists try to understand a [health] problem based on an image. They will disappear. Today they are the domain experts; they are the only ones who can understand the image. But, for example, there is a technology, deep learning, which knows how to do a very good job of image analysis, and once there is a large enough pool of samples and results [radiology images and radiologists' reports/labels] and this model learns these things, it will make these people redundant because they are just people and the computer will be more accurate, with a lesser chance of error.

Alex expressed the current view among data scientists that radiology's days are numbered, as deep learning algorithms outperform radiologists in diagnosing scans (Hinton, 2016). In this techno-optimistic view, radiologists' role is confined to providing annotations for the algorithm, and they are perceived as redundant once their knowledge has been assimilated into the machine.

Demoting experts to a lower standing. A second approach to absorb the expertise of other professions is to lower domain experts' standing. This view posits that the algorithms have assimilated the experts' knowledge and that the experts have been allocated another role within the division of labor. This role is no longer the authoritative role of experts with theoretical knowledge and cognitive decision-making abilities. When asked about data science's relations to medicine, Professor Gill, who trains data scientists and works in the medical domain, described the role he plans for doctors:

> Regarding doctors, I even wrote to them about it once, in an article on medical ethics [. . .] I argued that, ethically, doctors should advise their patients as little as possible because [the doctors] do not understand enough.
>
> They should not do a probabilistic calculation of what the patient has; I do it better than them, and there is evidence to support this. They should not address how patients should be treated; this, too, we perform better than them, as long as the patient has only one illness.
>
> Their job should be to consider and represent the patient's utility function – because that is something we still cannot do well – by understanding their fears and concerns and representing their preferences. I mean, there could be several solutions for you [as a patient], but one would suit you best.

Professor Gill challenged doctors' authority and jurisdiction. He argued in his academic work that doctors are not as competent as data scientists and algorithms in performing medical diagnosis and treatment. He recommended that the role of doctors should be reconstructed; rather than applying abstract theoretical knowledge to the diagnosis and treatment of patients, a doctor's 'job should be to consider and represent the patient's utility function'. According to this argument, the medical profession would be concerned with providing advice as a care and service profession, and its core task domain would become applying empathic and therapeutic skills. In this scenario, the role of doctors would be to learn the patients' fears and concerns and represent the patient's preferences to algorithms and their designers, whereas data scientists would become the authoritative entity in medical decision-making.

Creeping training. Another way in which data scientists absorb specific domains is by infiltrating the training of these domains. According to the data scientists interviewed for the study, the curricula of many disciplines are growing increasingly computerized because of the emphasis on computer-based data analysis. For example, Chen, a data scientist employed by a marketing start-up and studying for a master's in computer science, described a process where domain knowledge in all branches of academic research is becoming less important than computer science knowledge. Thus, in this narrative, algorithmic tools have become more important than the domain, even in the humanities and social sciences:

> It has started to seep into the humanities [. . .] simulating election systems and all those kinds of things. It is slowly creeping in the entire time. In chemistry, they already have mandatory courses in programming [. . .] and also in human resources. Today, everyone needs to program; it is a required course and not something extra.

Chen very convincingly described how the knowledge and way of thinking of data science and computer science are 'creeping' and 'seeping' into all knowledge domains: humanities, political science, chemistry, psychology, sociology, and management (see Capogna, 2022, for more on this trend, specifically in sociology). In Chen's opinion, mandatory programming courses are not the extra curriculum item that they once were but rather one of the key subjects of all data-based research domains. From the data scientists' perspective, the domain is becoming increasingly marginal, whereas algorithmic modeling and programming are taking center stage.

The data scientists' construction of domain experts and domain knowledge as temporary and the work of experts as either 'labeling' or 'emotional' recall the arguments of neo-Marxist literature on de-skilling (Braverman, 1974). Expertise in this narrative is attributed only to the task of crafting machine learning models. This construction of the relationship between data scientists and domain experts produces an inbuilt antagonism between the groups because the data scientists' success equals the loss of work, or at least the loss of occupational power, for domain experts. A relational authority, or even a relational monopoly, is thus created when data science forms relations of cooperation and coproduction with other occupations and professions and then plans to dominate them.

Serve

In the competitive narrative of replace and the competitive-cooperative narrative of absorb, data scientists challenge the expertise of different domain experts and the stability of their jurisdictions. As Abbott (1988) projected for AI experts in his discussion of the information professions, when exercising their competitive narratives, data scientists rely on their abstract (and practical) knowledge of machine learning to generate status and legitimacy in their relations with other experts (pp. 237–238). In contrast, the narrative of *serve* aims for continuous and respectful cooperation between data scientists and domain experts with no threat to domain experts' jurisdiction and status. Rather, according to this narrative, data scientists view their expertise as providing a tool or a service to domain experts to enhance (or 'augment') the latter's powers. This narrative also conveys the dependency of data scientists on domain experts and their will to cooperate and coproduce with experts (Hansen, 2021; Ribes, 2019). Interviewees used the narrative of *serve* in two ways in their sense-making talk: (a) while recounting daily workplace interactions with domain experts and (b) while describing the limitations of their expertise.

Real-world interactions with domain experts. When asked about data science's relations to domain experts, Zoe, a data scientist with a loan start-up, described the division of labor in the company she worked for:

> Analysts divide into two teams. The one abroad actually approves deals manually. I write a model that [approves loans], and what my model does not approve, the analysts approve or reject. [The goal] is that the model will know which deals are definitely good, and the analyst will decide on the ones that raise a dilemma.

Zoe described a cooperative and mutually beneficial relationship with the analysts in the company she worked for. Her task, she explained, was to write a model that would approve or reject loan applications. The analysts for their part manually reviewed loan applications that were not automatically approved by the model. Zoe's algorithm, therefore, reduced the workload of the analysts and handled routine, easy-to-analyze cases. The task of the algorithm is to help the experts in their routine work and not to replace or absorb their knowledge. According to this view, the analysts were not losing their jobs, the division of labor between them and Zoe was clear, and the team of analysts in the company has only grown.

Limitations of data science. Despite arguments that the work of domain experts will eventually be replaced or absorbed and despite efforts to market machine learning algorithms as the ultimate AI that can replace all labor with automation, many data scientists nevertheless questioned the capabilities of the technology. As the experts who actually produce the technology, data scientists are uniquely positioned to appreciate its internal limitations and the factors that prevent technological systems from ever replacing domain experts' work. One limitation cited by the interviewees relates to the garbage in, garbage out (GIGO) model. Machine learning systems depend on the quality of the input data, and data scientists are aware that the data for their analyses are often inaccurate. For example, Tal, a data scientist with a PhD in computer science who works for a large high-tech company, explained how he views the relations between domain experts and the data science team in his organizations:

> People get very scared, especially when recommendations [made by the algorithm] are accepted that go against a manager's domain-based intuition. Many times, ego and business experience lead them to reject your recommendation. By the way, that is not a mistake. Intuition and gut feelings can be more accurate sometimes. [There can be many reasons for this:] insufficient data; noisy data; inaccurate assumptions of ours; and data that are outdated following extensive changes in a domain.

Tal raised the issue of establishing trust relations with managers as domain experts who may prefer to follow their gut feelings, intuition, business experience, or ego rather than the decision of an algorithm. Apparently critical of the managers, Tal actually justified the domain experts' attitude and authority because he knows the limitations of technology. An analysis may be based on unsatisfactory and noisy data or incorrect assumptions, or there may have been changes in the domain since data collection. In this sense, Tal did not see the technology as preferable to the domain expert's intuition. In line with their education and mathematical understanding, other interviewees stressed that uncertainty invariably coexists with statistical probability analysis and can never be reduced.

When asked about his work with domain experts, Yotam, who has a PhD in computer science and works for a large high-tech company, emphasized that the data scientist's only task was to develop tools to help the expert:

> My goal is not to replace human labor. It is important for human deliberation to be involved. Take doctors, for example, I would not like a doctor to be someone who presses a button or receives a recommendation from someone else. I want him to have the medical knowledge to challenge the machine. If I make medical science wither away, I am shooting myself in the foot. So, I do not believe this is the plan. We are just trying to develop tools that help.

Yotam believed data science is dependent on expert knowledge and that doctors should argue with machines and not lose their authority in the medical jurisdiction. Contrary to Alma's narrative above, which anticipated the replacement of expert labor with AI, and contrary to Alex and Prof. Gill's narrative of absorbing medical knowledge into AI systems, and then pushing doctors out or down the occupational prestige ladder, Yotam does not think AI can ever replace physicians.

Another limitation raised by the data scientists is responsibility. Legal liability and the question of whether data scientists wish to appropriate the domain experts' authority minus the responsibility have become a scholarly concern (Haupt, 2019; Martin, 2019; Metcalf et al., 2019). Some interviewees said that they were fully aware of the issue of responsibility and had concerns about producing systems that were presumed to replace experts because of the legal and moral responsibility attached to the results. For example, when asked about the chances data science will replace experts, Jonathan, who has a PhD in computer science and works for a large high-tech company, expressed the following view:

> Doctors. Never. I do not think anyone would want medical, legal, or financial decisions to be fully [automated]. I think [algorithms] will improve efficiency in these fields. They will act like

> expert opinions, like an additional professional, expert opinion. Not instead. Even if [the algorithm] is very accurate, it [can only be] almost a hundred percent.

> As the person who is supposed to make the decision and tell you if you are sick, healthy, or going bankrupt, or whether to invest your money, to some extent, my name is on [the algorithm's decision].

Jonathan stressed that the algorithms do not work on their own. The data scientist creates and programs them and has their name attached to the algorithms' decision. Therefore, Jonathan considered himself responsible for the model's decisions, but he did not wish to be responsible for decisions regarding health, law, or finance. He was willing to remain in the position of advisor, offering a second professional opinion. His aim was not to replace the expert but to work alongside her. This narrative constructs the data scientists as service providers and advisors and their jurisdiction as advisory to both clients and professionals (Abbott, 1988).

Discussion and conclusions

The data scientists institutionalize three narratives in their relations with domain experts: replace, absorb, and serve. These narratives have different functions in data scientists' sense-making talk, referring to either the far future, war stories, specific domains, daily interactions, or doubts and limitations. This multiplicity of options that can be utilized in relations allows data scientists flexibility in the way they position themselves in their relations to other professions. A different jurisdiction is sought when data scientists seek to replace domain experts than when they seek to absorb or serve. Their dynamic choice of narrative impacts the way they approach domain experts and how they design technology (Hansen, 2021; Lange et al., 2019).

When data scientists see themselves as replacing or absorbing domain experts' knowledge, they construct a dominant and authoritative jurisdiction. They produce an image of the data scientist as 'the only real expert around' and position algorithmic decision-making as superior to expert discretion. This narrative has real-world consequences (Millerand et al., 2013). Scholars have already named this engineering perspective of the superiority of machines and the inferiority of human deliberation as 'technological solutionism' (Metcalf et al., 2019), which refers to the engineering conviction that any task and any problem are best resolved by technology. This perception imbues the idealistic stance of data science with technological determinism (Hansen, 2021) when data scientists promote the belief that machine learning will replace all human labor (boyd and Crawford, 2012). The *replace* and *absorb* narratives appeal to clients' recognition and legitimization by downplaying other forms of expertise and viewing them as redundant or temporary (Capogna, 2022). This rhetoric creates competition with other professions, threatens them, and promotes interprofessional battles over control (Abbott, 1988).

At the same time, data scientists also cultivate the relational narrative of *serve*, which constructs their expertise as cooperative and coproducing, and even dependent on domain experts. This type of rhetoric fosters a pragmatic coproduction of technology and expertise (Eyal, 2013; Hansen, 2021; Huising, 2015; Ribes, 2019) and recognizes the network of interprofessional connections in the production of AI. According to this narrative, new socio-technical configurations are created in the cooperation between data scientists and domain experts producing machine learning algorithms for the service (or 'augmentation') of expert labor, while recognizing experts' needs and the diversities in the role of quantification in various disciplines (Desrosières, 2016). This structuration of a wide network of relations imbues data science with relational vigor. When this narrative is institutionalized by data scientists themselves as a basic engineering logic, it contradicts ideas of big data mythology and the view that machine learning and AI can and will replace human labor. Thus, this collaborative engineering logic is consistent with the argument of social scientists and philosophers: that what is called 'artificial intelligence' is in fact a type of limited technology (boyd and Crawford, 2012; Capogna, 2022; Collins, 2018; Hansen, 2021).

However, this cooperative *serve* narrative leaves data science vulnerable to the authority of other experts (Abbott, 2001; Eyal, 2013; Kahl et al., 2016). Kahl et al. (2016) viewed collaborative language and collaborative technologies as the answer to the problem of maintaining authority in relations. Collaboration reduces competition, and those who are part of a collaboration achieve recognition and legitimation, and even relational authority (Huising, 2015). Eyal (2013) resolved the tension of achieving authority in relations with the concept of 'generosity as power'. He viewed the wide and generous distribution of expertise (as distinct from experts) as a type of power. For Abbott (1988, 2001), stretching a jurisdiction too widely and thus making too many connections necessarily makes a profession vulnerable to encroachment. The current study adds to the discussion by suggesting multiple narratives as another solution to the theoretical problem of maintaining authority in relations, specifically in wide networks of relations.

A multiplicity of options provides the necessary flexibility to adjust according to circumstances and to create new opportunities. Put together, the three relational narratives described here expand the data science network of relations to include many different types of domain experts (Avnoon, 2021; Ribes, 2019), while maintaining varying levels of authority in the relations (Hansen, 2021; Lange et al., 2019). While offering services, or acting as advisors, data scientists continue to hold on to their competitive narratives, namely replace and absorb. Thus, they form a wide network of connections but do not give up on their competitiveness and their claim to authority. The multiplicity of narratives allows data scientists an ambiguous, chameleon-like claim for authority (Cross and Swart, 2021; Dorschel and Brandt, 2021). Coherence is achieved at a higher level when data scientists position their expertise as universal (Avnoon, 2021; Ribes, 2019) and relevant to all domains. Note that data scientists do not have a narrative of *abstain*, viewing their expertise as not applicable to a certain domain.

Moreover, multiple narratives create ambiguity, confusion, engagement, fascination, and tensions in the interactions between data scientists and domain experts when the role and goals of data science work within different domains remain unclear (Capogna, 2022; Hansen, 2021; Lange et al., 2019; Martin, 2019). Therefore, when data scientists as a nascent professional group operate a multiplicity of narratives, they guarantee the continued interest and reaction of domain experts. In the long run, this may help them gain authority and legitimization, as ambiguity does in other aspects of knowledge work (Alvesson, 2001; Freidson, 1973), or it may cause a loss of jurisdiction, a disassembly of data science into subgroups, or an absorption of data science within different domains (Abbott, (1988, 2001)). It may well be that this chameleon-like relational pattern will lead to a loss of trust in data science. Future research may examine whether this ambiguity of narratives persists or, as the profession matures, whether data scientists will settle on one narrative per domain or for all domains. Future research may also investigate whether a temporal pattern of relations exists – for example, whether data science begins with a relational narrative of *serve* and then, slowly, as the technology improves, it absorbs and eventually aims to replace experts.

One limitation of this study is that it focuses on data scientists' talk and not on their actual interactions with domain experts. As structuration scholars have long argued (Giddens, 1984; Jerolmack and Khan, 2014), people tend to behave differently than they talk, and interactions have a life of their own, resulting in structures that are different from those intended by agents. Thus, the actual relations between data scientists and domain experts are not a mirror image of the narratives envisioned by the former. Experts may object to data scientists' narratives, and this may, in reality, create different relations with them, or even convince data scientists to abstain from certain tasks and problems. For example, sociologists of quantification have already begun to criticize the methodology of data science and big data in social science research, specifically the attempt to 'objectify' social reality with theoretically uninformed procedures, the attempt to measure collective performance based on individual performance, the move of social and policy research from academia into industry, and the lack of standardization in computational social research (Capogna, 2022). Future studies should therefore continue to investigate the perspectives of various domain experts on data science professionals and technologies, as well as the actual use of data science expertise in different settings, in the present and over the next few decades.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Netta Avnoon https://orcid.org/0000-0002-3758-2097

Notes

1. 'Domain expert' is data science jargon for an expert in any other field but data science – for example, a doctor, psychologist, manager, engineer, or accountant.
2. The term 'domain agnostic' is data science jargon for their ability to move between domains. See Ribes (2019) and Ribes et al. (2019) for more on the use of the word 'domain' in data science.

References

Abbott A (1988) *The System of Professions: An Essay on the Division of Expert Labor*. Chicago, IL: University of Chicago Press.

Abbott A (2001) *Chaos of Disciplines*. Chicago, IL: University of Chicago Press.

Alvesson M (2001) Knowledge work: Ambiguity, image and identity. *Human Relations* 54(7): 863–886.

Anteby M, Chan CK and DiBenigno J (2016) Three lenses on occupations and professions in organizations: Becoming, doing, and relating. *Academy of Management Annals* 10(1): 183–244.

Avnoon N (2021) Data scientists' identity work: Omnivorous symbolic boundaries in skills acquisition. *Work, Employment and Society* 35(2): 332–349.

boyd d and Crawford K (2012) Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15(5): 662–679.

Brandt PS (2016) *The emergence of the data science profession*. PhD Thesis, Columbia University, New York .

Braverman H (1974) *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century*. New York: Monthly Review Press.

Bruno I, Jany-Catrice F and Touchelay B (eds) (2016) *The Social Sciences of Quantification: From Politics of Large Numbers to Target-Driven Politics*. Cham: Springer.

Capogna S (2020) The work we shall be amidst organisation and community. *International Journal of Innovation and Economic Development* 6(3): 7–19.

Capogna S (2022) Sociology between big data and research frontiers, a challenge for educational policies and skills. *Quality & Quantity* 5: 1–20.

Collins H (2018) *Artifictional Intelligence: Against Humanity's Surrender to Computers*. Hoboken, NJ: John Wiley & Sons.

Cross D and Swart J (2021) Professional fluidity: Reconceptualising the professional status of self-employed neo-professionals. *Organization Studies* 42(11): 1699–1720.

Desrosières A (2016) The quantification of the social sciences: A historical comparison. In: Bruno I, Jany-Catrice F and Touchelay B (eds) *The Social Sciences of Quantification*. Cham: Springer, pp.183–204.

Dorschel R (2021) Discovering needs for digital capitalism: The hybrid profession of data science. *Big Data & Society* 8(2): 1–13.

Dorschel R and Brandt P (2021) Professionalisierung mittels ambiguität: Die diskursive konstruktion von data scientists in wirtschaft und wissenschaft. *Zeitschrift für Soziologie* 50(3–4): 193–210.

Eyal G (2013) For a sociology of expertise: The social origins of the autism epidemic. *American Journal of Sociology* 118(4): 863–907.

Freidson E (1970) *Professional Dominance: The Social Structure of Medical Care*. New York: Atherton Press.

Freidson E (1973) Professions and the occupational principle. In: Freidson E (ed.) *The Professions and Their Prospects*. Beverly Hills, CA: Sage, pp.19–38.

Giddens A (1984) *The Constitution of Society*. Cambridge: Polity Press.

Gillespie T (2014) The relevance of algorithms. In: Gillespie T, Boczkowski PJ and Foot KA (eds) *Media Technologies: Essays on Communication, Materiality and Society*. Cambridge, MA: MIT Press, pp.167–193.

Goffman E (1959) *The Presentation of Self in Everyday Life*. New York: Anchor Books.

Hansen KB (2021) Model talk: Calculative cultures in quantitative finance. *Science, Technology, & Human Values* 46(3): 600–627.

Haupt CE (2019) Artificial professional advice. *Yale Journal of Law & Technology* 21(3): 55–77.

Hinton GE (2016) On radiology. In: *Machine learning and the market of intelligence, creative destruction lab*, Toronto, ON, Canada. Available at: https://www.youtube.com/watch?v=2HMPRXstSvQ

Huising R (2015) To hive or to hold? Producing professional authority through scut work. *Administrative Science Quarterly* 60(2): 263–299.
Jerolmack C and Khan S (2014) Talk is cheap: Ethnography and the attitudinal fallacy. *Sociological Methods & Research* 43(2): 178–209.
Kaggle Survey (2019) State of data science and machine learning. Available at: https://www.kaggle.com/kaggle-survey-2019
Kahl SJ, King BG and 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.
Korbet R (2019) The state of the Israeli ecosystem. Startup Nation Central: Finder Insight Series. Available at: https://lp.startupnationcentral.org/annualreport/ (accessed 30 December 2019).
Lamont M and Swidler A (2014) Methodological pluralism and the possibilities and limits of interviewing. *Qualitative Sociology* 37(2): 153–171.
Lange AC, Lenglet M and Seyfert R (2019) On studying algorithms ethnographically: Making sense of objects of ignorance. *Organization* 26(4): 598–617.
Larson MS (1977) *The Rise of Professionalism: A Sociological Analysis*. Berkeley, CA: University of California Press.
Martin K (2019) Ethical implications and accountability of algorithms. *Journal of Business Ethics* 160(4): 835–850.
Metcalf J, Emanuel M and boyd d (2019) Owning ethics: Corporate logics, Silicon Valley, and the institutionalization of ethics. *Social Research: An International Quarterly* 86(2): 449–476.
Millerand F, Ribes D, Baker KS et al. (2013) Making an issue out of a standard: Storytelling practices in a scientific community. *Science, Technology, & Human Values* 38(1): 7–43.
Muzio D, Kirkpatrick I and Kipping M (2011) Professions, organizations and the state: Applying the sociology of the professions to the case of management consultancy. *Current Sociology* 59(6): 805–824.
Noordegraaf M (2020) Protective or connective professionalism? How connected professionals can (still) act as autonomous and authoritative experts. *Journal of Professions and Organization* 7(2): 205–223.
Ribes D (2019) STS, meet data science, once again. *Science, Technology, & Human Values* 44(3): 514–539.
Ribes D, Hoffman AS, Slota SC et al. (2019) The logic of domains. *Social Studies of Science* 49(3): 281–309.
RJ Metrics Survey (2016) The state of data science. Available at: https://www.stitchdata.com/resources/the-state-of-data-science/
Strauss A and Corbin J (eds) (1997) *Grounded Theory in Practice*. Thousand Oaks, CA: Sage.
Susskind R and Susskind D (2015) *The Future of the Professions: How Technology Will Transform the Work of Human Experts*. Oxford: Oxford University Press.
Takhteyev Y (2012) *Coding Places: Software Practice in a South American City*. Cambridge, MA: MIT Press.
Wajcman J (2007) From women and technology to gendered technoscience. *Information, Community and Society* 10(3): 287–298.

Author biography

Netta Avnoon is a postdoctoral fellow at the Coller School of Management, Tel Aviv University, and a visiting scholar at the Department of Sociology, Columbia University. She studies expertise in diverse fields, focusing on AI and expert labor. She is interested in questions of power, ethics,

legitimacy, gender, and the link between micro-processes in the habitus of the expert and the institutional level of expertise.

Résumé
Comment les scientifiques des données (*data scientists*) définissent-ils leurs relations avec les experts d'autres domaines ? Cette étude s'intéresse plus particulièrement à l'autorité professionnelle à laquelle aspirent les *data scientists* et à leurs multiples récits concernant les relations de la science des données avec d'autres domaines d'expertise. Sur la base de l'analyse de 60 entretiens ouverts approfondis avec des scientifiques des données, des professeurs de sciences des données et des responsables de scientifiques des données en Israël, les résultats montrent que les *data scientists* institutionnalisent trois récits concernant leurs relations avec les experts d'autres domaines: (a) remplacer les experts, (b) absorber les connaissances des experts et (b) fournir un service aux experts. Ces trois récits présentent l'expertise des scientifiques des données comme universelle et omnivore, c'est-à-dire qu'elle est adaptée aux besoins de nombreux domaines et permet aux scientifiques des données d'être flexibles dans leur revendication d'autorité.

Mots-clés
Algorithmes, domaine de compétence professionnelle, expertise, intelligence artificielle, professions, relations interprofessionnelles, science des données

Resumen
¿Cómo enmarcan los científicos de datos sus relaciones con los expertos de otros campos? Este estudio se centra en la jurisdicción profesional a la que aspiran los científicos de datos y sus múltiples narrativas con respecto a las relaciones de la ciencia de datos con otros campos de especialización. Con base en el análisis de 60 entrevistas abiertas en profundidad con científicos de datos, profesores de ciencia de datos y gerentes de ciencia de datos en Israel, los hallazgos muestran que los científicos de datos institucionalizan tres narrativas con respecto a sus relaciones con los expertos de otros campos: (a) reemplazar a los expertos; (b) absorber el conocimiento de los expertos; y (b) prestar un servicio a los expertos. Estas tres narrativas construyen la experiencia de los científicos de datos como universal y omnívora; es decir, son relevantes para muchos campos y permiten que los científicos de datos sean flexibles en su reivindicación de autoridad.

Palabras clave
Algoritmos, ciencia de datos, inteligencia artificial, jurisdicción profesional, pericia, profesiones, relaciones interprofesionales