

Data

In the two previous chapters, we have stretched classic phenomenological accounts of the construction of the social world by attending to the consequences of mediated communications for the space and time of everyday social interaction. But we have not so far encountered anything that fundamentally disrupts the approach to the sociology of knowledge offered by Berger and Luckmann. They understood social knowledge as built up through the accumulation of ‘ordinary’ members of society’s knowledge acquired through everyday ‘thoughts and actions’ (1966, p. 33). However, we have noted that data-based infrastructures of computer-mediated communications now play a key role in social interaction and that this *might* be shifting how we acquire social knowledge. In this chapter, we take that issue head on: we will be concerned with the deep enfoldings in everyday life of automated data-gathering and data-processing which, in their underlying operations, are very far from everyday ‘thought and action’.

What are the implications of ‘data’ – as acquired, processed, configured and re-presented by computer-based systems – for social knowledge? We use the term ‘systems’ here in a descriptive, not theoretical, sense to refer to configurations of computing resource that enable the performance of large-scale information processing, operating to a large extent without direct human intervention, through ‘automating mediation’ via ‘software agents’ (Mansell, 2012, pp. 108–115). We do not intend the word ‘system’ to suggest any allegiance to a *theory* of ‘social systems’: indeed, we reject any such theories.¹ That definitional point aside, this chapter is a turning-point in our argument. This is the point at which a materialist phenomenology starts to diverge substantially from classic phenomenology. As anticipated in Chapter 1, this is also the point where the apparently irreversible breach between the phenomenological tradition (once cast in properly materialist form) and apparently anti-phenomenological (because materialist) accounts of knowledge such as Foucault’s (1970, p. xiv) can be repaired.

The strength of Berger and Luckmann’s work, and the wider tradition

of phenomenology, was to move away from a concern with the social contexts in which ‘ideas’ are generated (a sociological supplement to traditional ‘history of ideas’) towards an interest in ‘everything that passes for “knowledge” in society’, that is, with ‘common-sense knowledge’ (1966, p. 26). But Berger and Luckmann bracketed out the question, previously raised by Schutz, of the ‘social distribution of knowledge’ (1966, p. 28), in order to try to develop a ‘single body of reasoning’ about social knowledge in the style of Talcott Parsons (1966, pp. 28–29). That is why they were comfortable also with bracketing out from their sociology broader questions of epistemology, as if they had no bearing on everyday life. Both decisions now seem problematic. The growth of ‘data’ is part of a major *re*-distribution of knowledge production: any account today of what we know in the social world must confront a *conflict* (or at least a *plurality*) in what passes for social knowledge and in everyday epistemology. We come face to face with that plurality every day. This gives a new prominence to the sociology of knowledge,² but on terms very different from those supposed by Berger and Luckmann.

The classic thesis of the ‘social construction of reality’ assumed that one could build an account of common-sense knowledge simply by bringing together overlapping perspectives on how knowledge arises for human actors in their everyday ‘social contexts’ (Berger and Luckmann, 1966, p. 15). Our thesis in this book – the *mediated* construction of reality – considers media’s implications for such an account of common-sense knowledge, but what if the workings of ‘media’ are not understandable exclusively within our accounts of particular bounded social contexts? Data’s role in media’s operations pushes us further in this direction.

‘Data’ is the symbolic rough material out of which, through processes of accumulation, sorting and interpretation, ‘information’ is generated for use by particular actors with particular purposes (Kallinikos, 2009a). Although, casually, we talk about ‘raw data’, in reality no data is ‘raw’. ‘Raw data is an oxymoron’ (Bowker, 2008, p. 184; Gitelman and Jackson, 2013, p. 13; Kitchin, 2014, p. 20), which means that data always materializes within a particular practice and structure of collection: at its simplest, a *database*. We can leave the details of that to one side for now. Our point, more basically, is that ‘data’ and ‘information’ generated by systems of computers are today a precondition for everyday life; the *selections* from the wider ‘world’ (Boltanski, 2011) made by data processing are consequential for social life. That much was already grasped by Anthony Giddens in the 1980s (1984, p. 309, quoted in Gandy, 1993, p. 13), although Giddens’ particular focus was the state’s role in gathering information, not the wider processes of

state *and* corporate monitoring we see today. Another pioneering account of the social role of data processing by both markets and states was James Beniger's (1986) description of nineteenth-century modernization, but this did not prioritize phenomenological issues.³

The challenge to phenomenology from contemporary data practices stems from three developments subsequent to Giddens' observations. First, the collection of data is now continuous in many processes of social action and interaction, generating volumes of data whose processing is unmanageable without automation. In many rich countries, basic acts, such as booking a train or plane ticket, or keeping in touch with friends, now have as their precondition the unimpeded operation of networked systems of data-gathering and data-processing. Such automated processes are not a special case, or reserved for large institutions such as government departments: they are becoming, for many, the *general background of everyday life*. Second, the largest proportion of data processing now lies in the hands of the 'corporate private sector' (Gandy, 1993, p. 13), that is, organizations whose goals cannot be equated with the general social interest, since they are aimed at private competitive advantage. Those goals are necessarily *external* to the model of social knowledge developed by classic phenomenology, and there is at least an attempt to implement them on a global, not merely national or regional, scale (Mosco 2014). Third, the outcomes of such data processing include the generation, *prima facie*, of *social knowledge itself*, at least in an instrumental sense: information put to use in the management of social interaction. Social actors are sorted in relation to particular action-outcomes on the basis of how data relating to them is categorized and processed. Data processing is, as Oscar Gandy (1993, p. 15), a pioneer of its sociological study, put it, 'a discriminatory technology' that works through 'three integrated functions': 'identification' (the collection of data of administrative relevance), 'classification' (the resulting assignment of individuals to pre-formulated groups), and 'assessment' (the assignment of individuals to particular action-outcomes based on comparing how they are classified).

In the term 'data' we include all the processes and underlying infrastructures for gathering, sorting, collecting, evaluating and acting upon data. Data comprise today a substantial proportion of the 'socially available stock of knowledge' (Berger and Luckmann, 1966, p. 61). Data production is inherently *asymmetrical*, in a way not envisaged in classic phenomenology's model of social knowledge: it is oriented to the purposes of the institutions – private or governmental – that use the data. True, individual social actors are themselves involved in mutual data-gathering, and may

act in ways that adapt to data-gathering processes (we return to this), but this does not change the fact that the primary drivers of data processes as forms of social knowledge are institutions external to the social interactions in question. A large proportion of such data is produced automatically, relying on processes of aggregation and algorithmic calculation that are driven by the needs of those external institutions. True, we could also argue that the growing interdependency of everyday life and media technologies (what we have called deep mediatization) is itself a key driver of data production – at least from the perspective of *today's* version of digital infrastructure: we can surely imagine other versions that are not dependent on continuing data-gathering. That would raise wider issues about the particular types of figurational order that are today becoming dominant, a point to which we will return in Chapter 10. For now, we ask only the basic question: what are the consequences of the deep social embedding of data processing for social phenomenology?

7.1 Data and the Premises of Classic Social Phenomenology

Berger and Luckmann were right to make explicit a key organizing dynamic in social experience, that is, how social actors 'apprehend the reality of everyday life as an *ordered* reality' (1966, p. 35). There is still force to that idea: states of affairs where we *cannot* apprehend our everyday reality as ordered are deeply distressing and disruptive of basic human processes (times of social and civic breakdown, political terror, deep forms of social victimization). But the issues raised by data-processing already confront us with conflicts about *how* reality is ordered, and *what* its order is. Schutz saw the 'how of the individual situation in the lifeworld' (Schutz and Luckmann, 1973, p. 105) as fundamental to social knowledge, but conflicting accounts of that 'how' now circulate: data processes generate many of those accounts. The economy of data collection and processing is now a crucial dimension of the wider market economy, as well as the operations of the state. The collection of data does not operate through the give-and-take – the mutual acknowledgement – of social interaction, but rather through processes of automated extraction exterior to any possible reflexive human action. Social actors may seek sometimes to resist this, but the resistance can only be partial since so many forms of action seem to have, as their precondition, such prior processing and the categorizations on which such processing depends.

As a result, the two premises of Berger and Luckmann's phenomenology

of the social world are challenged: first, that ‘everyday life presents itself as a reality *interpreted by men* [sic] and subjectively meaningful *to them as a coherent world*’; and, second, that the world of everyday life is not only taken for granted as reality by the ordinary members of society, but ‘it is a world that *originates* in their thoughts and acts, and is *maintained as real by these*’ (1966, p. 33, emphases added). We are not arguing that in everyday life social actors no longer try to generate common-sense knowledge in their thoughts and acts, just because *automated processes* of data-processing are deeply embedded in their daily lives. For sure, ‘common-sense knowledge’ of the social world remains ‘the knowledge shared with others, the normal, self-evident routines of everyday life’ (Berger and Luckmann, 1966, p. 37). But there are other forms of knowledge of the social world now in play,⁴ not always self-evident to social actors, over which they have no control and yet which *impact them* deeply, and we need now to integrate this fact into our understanding of what we do and think every day.⁵

Berger and Luckmann were certainly aware of the reliance of social actors on wider patterns of knowledge and institutional knowledge production. They acknowledged the role of language in ‘transcending the “here and now”’ (1966, p. 54), and they had an account of how, more widely, institutions work to underwrite the hierarchy of knowledge that underlies social order, making legitimation not just a normative but a cognitive fact (1966, p. 111). But their account of how this works fits poorly with the role of automated data processing in everyday life today. Their examples of the role of ‘system’ in knowledge production were different in kind from today’s data-processing systems, and much less consequential for the content of social knowledge: the background role of the telephone system in everyday transmission of communication, the bureaucracy that arranges one’s new passport (1966, pp. 56–57). Berger and Luckmann simply could not have anticipated the role of data and information systems in generating *knowledge* for contemporary life, that is, their role in supporting and shaping the *ontology* of everyday interaction. For Berger and Luckmann, ‘institutionalization’, though broadly defined, depends in the end on a fitting together of how *human agents themselves* act in, and make sense of, the world: ‘institutionalization occurs whenever there is reciprocal typification of habitualized actions by types of actors’ (1966, p. 72). However bewildering the scale of such institutional forms of sense-making and knowledge production may seem to the individual actor, the institutional world that appears to the individual *remains* ‘a humanly produced, constructed objectivity’ (1966, p. 78).

How well does this account of the sociology of knowledge fit with

data's role in everyday life today? For sure, our reliance on data in the social world already seems a *social* necessity. The internet evolved as an information space connecting, potentially, every computer and computer-based device on the planet, and every file found there. That huge expansion of scale, if it is within our cognitive reach, requires automated processes. The infinity of images and texts, people and events that we now regard as being 'there for us' online could not be 'there *for us*', as human actors with limited processing capacities, if it were not for the automated processes of search engines ('apps'). For this reason, it is important to think about the wider infrastructure of contemporary communications – including their data aspects – as a crucial dimension of what Berger and Luckmann call objectivation whereby 'the externalized products of human activity attain the character of objectivity' (Berger and Luckmann, 1966, p. 78), and so social reality is constructed. But this 'objectivation' operates by rather different rules from those envisaged by classic phenomenology.⁶

The problem is not merely one of increased complexification and delegation. Berger and Luckmann had already anticipated this, arguing that the resulting pressure against overarching order was overcome by 'establishing a stable symbolic canopy for the whole society' (1966, p. 103) which arranged even realities very distant from social actors' experience within a relational hierarchy (1966, p. 110). Data processes are disruptive not just because they are distant, but because they involve the *unimaginably large-scale and automated* repetition of processes of counting, sorting and configuring data (generating new forms of cognition). More generally, a large proportion of what now passes for social knowledge is held not by persons, but within an impersonal 'reserve' of accumulated text and images (the internet) that is available to us not directly (it is too large for that) but indirectly via *automated search* (Halavais, 2009) and automated processing of other sorts. Under these new conditions, Berger and Luckmann's once unobjectionable statement is jarring:

Knowledge of how the socially available stock of knowledge is distributed, at least in outline, is an important element of that same stock of knowledge. I know, at least roughly, *what I can hide from whom, who I can turn to for information on what I do not know, and generally which types of individuals may be expected to have which types of knowledge.* (1966, p. 61, added emphasis)

And yet Berger and Luckmann remain right when they argue that the pressures for things to hang together are strong, and the need to optimize the convergence of 'relevance' is high (1966, pp. 81–82). The *impression* of

‘meaningful reciprocity’ (1966, p. 82) among social actors is still important, even if there can be no reciprocity (in the sense Berger and Luckmann intended) between humans and the automated processes which accumulate, count and configure data for and about them.

What then if sociology of knowledge acknowledges, in the midst of the social world, *other* forms (indeed forces) of ‘social’ knowledge than those generated by social actors? The result need not be ‘reification’, defined by Berger and Luckmann as the forgetting of the human role in the construction of reality, so that ‘the objectivated world loses its comprehensibility as a human enterprise’ (1966, p. 106): data processes, after all, are themselves the mediated result of any number of social, cultural and political processes, involving humans at some level. But there is something significant in the fact that ‘data’ involves processes that exceed the direct capacities of human agents, whether to perform or to model:⁷ in this sense, data involves a certain kind of *materialization* (via media and their infrastructure) that brings, in turn a particular *institutionalization* of knowledge. The goals, norms and ‘knowledges’ of those processes are necessarily different in kind from those of human actors. The relations of ‘social’ knowledge therefore to questions of legitimacy, value and social order become less straightforward than they appeared to Berger and Luckmann. Let us now attempt to explore this in some specific areas.

7.2 New Institutions for ‘Social’ Knowledge

Like all infrastructures, the infrastructure of ‘data’ *sinks* inside social arrangements (Star and Ruhleder, 1996): if it did not, it would not be doing its work of enabling our lives to run their ordinary course. But since the data infrastructure is a structure for social *knowledge*, we must bring it to explicit reflection, if we are to have a satisfactory account of how the social world is constructed.

The process of institutionalizing knowledge through data has many components: it operates via the materialization of a network of networks so complex that it makes no sense to regard it as animated by one single logic or ‘dominant shaping force’ (in that, it is similar to all technology: MacKenzie and Wajcman, 1999, p. 18). That said, its point is, so far as possible, to act *as* an infrastructure, *as* a practical system. *That*, in broad terms, is what someone expects when s/he trusts our credit card, phone or laptop will work when getting off a plane the other side of the world, or indeed a train the other side of a country. And such expectations, on the part of social actors, are inseparable from the goals of countless businesses

to achieve a seamless *plane of interoperability* between their products and services and those of other businesses. Such corporate ambitions have become unthinkable without data tracking: as Armand Mattelart puts it, ‘the tracking grid now provides meaning on a planetary scale’ (Mattelart, 2010, p. 2). The importance of this infrastructure for wider forms of political and social order is undeniable: ‘technological innovations are similar to legislative acts or political foundings that establish a framework for public order [. . .] in tangible arrangements of steel and concrete, wires and transistors, nuts and bolts’ (MacKenzie and Wajcman, 1999, p. 33) – and, we might now add, code.

Databases and Social Classifications

Crucial to this transformation is the database; ‘the ability to order information about entities into lists using classifications [is] a contemporary key to both state and scientific power’ (Bowker, 2008, p. 108). The database has a distinctive type of power which Bowker defines as ‘jussive’, an ordering power based on an ‘exclusionary principle’ that determines what can and cannot be stored in a particular form (2008, p. 12). The consequences of database operations are in this sense final: ‘what is not classified gets rendered invisible’ (2008, p. 153). The point of database operations is to fix the *starting-point* (the base) from which data operations – counting, aggregating, sorting, evaluation – begin. In that sense, by being placed in a database, ‘data’ becomes ‘unmoored’ from the underlying materials from which it was gathered (Kitchin, 2014, p. 72). It makes no difference to this process that some of the processes developed on the base of the underlying data architecture may themselves go on adapting in response to emerging patterns in data, provided the structure of the underlying database remains the same.⁸ Such adaptations are in any case driven not by some independent ‘will’ present in the processes from which data are gathered but by the emergence of features that, from the perspective of the data process, are defined as significant enough to trigger such adaptations. In all such cases, the ‘knowledge’ that is produced cannot be separated from the purposive selections out of which the database is formed, or subsequently adapted. Insofar as the outcomes are treated as *direct* knowledge about the processes re-presented by the data, they are misleading. As Bowker put it, ‘our memory practices [are] the site where ideology and knowledge fuse’ (2008, p. 228). We have already in Chapter 3 suggested that these developments can be seen as a potential new fourth wave of mediatization, through which our interdependencies progressively *deepen* through infrastructures

for the continuous production and exchange of *data*: the emerging wave of datafication, within the wider wave of digitalization.

Most data involved in the organization of social life are ‘made from the raw material of human experience’ (Gandy, 1993, p. 53) and used for social *classification*. The purpose of data-gathering is not neutral, but precisely discriminatory, that is, ‘to coordinate and control [people’s] access to the goods and services that define life in the modern capitalist economy’ (Gandy, 1993, p. 15). This discrimination requires a massive pooling of computational resources: a single database is not sufficient. It is essential to aggregate separate databases into massively larger ones, enabling the matching of patterns across countless sites of data collection, from which predictions can be made (Gandy, 1993, pp. 71–84). From the development of ‘distributed relational databases’ in the 1990s and 2000s to the work of today’s growing data sector, including global businesses on the scale of Google, Expedia, Acxiom (Kallinikos, 2009b, p. 232; Nissenbaum, 2010, pp. 41–45), the aggregation-for-value of data originally collected from specific locations and in specific contexts now comprises a basic fact of social life.

It is naive to ignore the consequences that the principles of selection underlying such data-gathering may have for particular distributions of power, and for the long-term organization of the social world. As Theodore Porter put it in a broader context, ‘quantification is a technology of distance’, motivated not by a ‘truth to nature’ but by ‘the exclusion of judgement’ (1995, p. ix) that makes particular types of judgement possible and efficient through a reshaping of the social world itself: ‘the quantitative technologies used to investigate social and economic life work best if the world they aim to describe *can be remade in their image*’ (Porter, 1995, p. 43, added emphasis). Indeed, because large-scale institutions necessarily act at a great distance from the realities they seek to influence, the operation of data processes plays a crucial role in *making* a distant, intractable ‘world’ *into* an ordered, calibrated reality that can be interpreted and governed: as anthropologist James Scott put it, ‘legibility’ [. . .] is a central problem in statecraft’, requiring a ‘politics of measurement’ (1998, p. 2, 27). But, as some legal theorists have noted, we must extend this analysis to the whole surveillance apparatus of the contemporary commercial and information technology environment (Cohen, 2012; Pasquale, 2015).

We miss the social form of this complex process, unless we connect up the abstractions inherent in data functions to the *experiential processes* in which those functions have become embedded (Cohen, 2012, p. 20), and so register the potential violence of data-based ‘processes of [. . .] representation and classification’ (2012, p. 24). For this, phenomenological

reference-points are essential.⁹ But a core feature of data infrastructures is deeply at odds with the assumptions of classic phenomenology. That is its *opacity*:¹⁰ ‘the configuration of networked space [. . .] is increasingly opaque to its users’ (Cohen, 2012, p. 202). Data’s continuous processes of selection and comparison have a generalizing force across an infinite domain that guarantees a larger asymmetry in *social knowledge*. As Christine Alaimo and Jannis Kallinikos (2015, pp. 15–16) put it, ‘it is exactly because of the abstract nature of data and the simplicity of the logic of encoding that the social can be represented in all of its (now compatible) highly pliable forms [. . .] *Once the social gets engraved into data, it ceases to be related to established categories and habits*. It is transposed onto and thus enacted according to the very same logic’ by which it was generated. Or, as Jose van Dijck put it more succinctly, ‘it is easier to encode sociality into algorithms than to decode algorithms back into social action’ (2013, p. 172). The translation of social life into ‘data’ therefore casts a large shadow: the domain of descriptors that *fail* to get captured by data processes (Balka, 2011).

The result is a change – a potentially profound change – in our relations to infrastructure. Infrastructures are, at root, tools for human action, operating at the highest level of complexity, a black-boxed substrate of ‘ordinary’ human action. In the digital world, our infrastructural tools (for example, our pages on a social media platform) are increasingly entangled in *powerful* and *distant* processes, which we cannot unpick or challenge. All tools involve mechanisms whose details we forget when we use them, or perhaps never knew: we may guess how a hammer has been put together, even if most know little about how the modern car is constructed, yet both hammer and car are tools of everyday living and our ignorance of their workings (their black-boxedness) is not crucial to the quality of our use. But many of today’s ‘digital tools’, as we use them, are black-boxes of a different sort, black-boxes that *are also in the act of using us*.¹¹ They track our actions algorithmically, not to enable the tool to work better for us, but to generate data for the *toolmaker’s* use: that is, to enable us to be better targeted by advertisers and marketers (Turow, 2011). This is indeed the very rationale of the much-hyped ‘Internet of Things’, yet its transformation of our usual relations to infrastructure seems not to have been noticed.

The result is a social relationship to abstraction very different from that envisaged by Schutz. Schutz saw artefacts as comprising the extreme end of the spectrum of ways in which *humans typify* (abstract from) their world (Schutz, 1967, p. 201). But today’s data-based *artefacts* now themselves operate to *typify humans* mostly for commercial ends and surveillance, to construct a seamless world for commerce and control. We might call

this *tool reversibility*. Tool reversibility is not immediately apparent when we use data-based tools, but it becomes apparent through our practices of use and the obstructions those practices encounter: whenever we use a data-based tool, it is already using us. This is one of the deeper cultural and social implications of the embedding of algorithms in everyday life (Napoli, 2014).

Categorization

Data processes rely, in turn, on categorization. Categories have been important in social theory for more than a century. For Durkheim and Mauss (1969 [1902]), categories (as outputs of a system of classification in so-called ‘primitive societies’) were derivatives of the actual divisions of society itself, and of the very idea of society itself. In most subsequent accounts, the order of causality is reversed with categories contributing to ‘the built information environment of a society’ (Bowker and Star, 1999, p. 5): ‘typification’ playing a similar role in classic phenomenology. But to grasp the degree of abstraction involved in data-based categories, we need to look more specifically at their features.

As David Berry explains, no process of computer-based categorization (and so no sorting, combination or evaluation based on it) can operate unless an ‘object’ has been created: ‘in cutting up the world in this pattern, information about the world necessarily has to be discarded in order to store a representation within the computer [. . .] those subtractive methods of understanding reality [. . .] produce new knowledges and methods for the control of reality’ (2011, pp. 14–15). So, in order to compose the *objects* in a database, such abstraction, first, needs to have occurred. As already noted, there is no raw data, but only ‘data [. . .] produced through measuring, abstracting and generalizing techniques that have been conceived to perform a task’ (Kitchin, 2014, p. 19). Second, where large numbers of objects are to be processed by automated functions or algorithms, processing requires prior organization, the design of a database *structure* ‘to extract the data located in them as rapidly and as effectively as possible’; in this sense, ‘a data structure forms a sort of intermediate level, an abstraction mechanism, in the process of addressing machine memory’ (Fuller and Goffey, 2012, p. 84, 85). Third, the more complex the operations to be completed, the greater the need to *combine* data levels and so enable more complex processing, what Fuller and Goffey call ‘abstraction layers’: ‘the more that different functions of a process [. . .] or software can be integrated by one layer of implementation, the wider it circulates and

coalesces. The more generality an abstraction is capable of, the greater its degree of usefulness; and the greater its tenacity in self-stabilization, the more activity is arrayed around it' (2012, pp. 88–89). In this way, more and more of what was once heterogeneous information can be processed together, but at the price of ever-greater levels of abstraction. Fourth, the processes of *calculation* performed on data must be automated through the use of algorithms. Algorithms are often identified, casually, with the whole process of data-based transformation of everyday experience, but as 'encoded processes for transforming input data into a desired output, based on specified calculations' (Gillespie, 2014, p. 167), algorithms are merely one of many elements in a sequence of progressive abstraction, if an essential one.

Alongside the relations between data processes (involving categories that do social work), we also need to consider what happens when the outputs of those data processes are *played back to social actors*. That they are played back (Isin and Ruppert, 2015, p. 113) is another factor, which a sociology of knowledge that takes data seriously must grasp. Categorization¹² is fundamental to all forms of organization, including social organization. Without it, effective (or at least non-random) interaction with the world would be impossible. Yet in the social realm, categorization has a distinctive feature neglected in phenomenology's account of typification and not even conceivable in Durkheim and Mauss's society-driven model. As Ian Hacking pointed out, classifications of human objects are 'interactive' in a way that, arguably, classifications of non-human objects are not:¹³

Ways of classifying human beings interact with the human beings who are classified. [. . .] classifications do not exist only in the empty space of language but in institutions, practices, material interactions with things and other people [. . .]. people are aware of what is said about them, thought about them, done to them. They think about and conceptualize themselves. Inanimate things are, by definition, not aware of themselves in the same way. (Hacking, 1999, pp. 31–32)

Hacking's insight has particular importance for the digital age when actors and actions are relentlessly categorized in countless ways for various purposes.

Our ways of interacting with categories are not easy to disentangle. They occur not randomly but in a highly structured context linked to the purpose for which data is being gathered in the first place. The simplest example is the social media platform. As Daniel Neyland notes, algorithms do not have a simple or automatic recursive effect on the social

world, but involve a ‘configuration through which users and/or clients are modelled and then encouraged to take up various positions in relation to the algorithm at work’ (Neyland, 2015, p. 122). What is striking about the configurations of social media platforms is that we act on and through them, largely as if there were *no* such configuration: indeed the very idea of ‘platform’ is a constructed space where the interface between everyday interaction and commercial transaction *appears* natural, a seamless data-flow (Gillespie, 2010). Through our underlying desire to maintain our social commitments across the newly configured spaces where they seem to migrate, there emerges a ‘growing *social* commitment to functionality’ (Plantin, Sandvig et al., forthcoming). This has major implications for how the social world is constructed.

Translating Data into Practice

There are at least five fundamental ways in which data abstracted from social experience can translate into frameworks for social practice. These connect with the dimensions in terms of which the other chapters of Parts II and III are organized.

The first relates to the organization of *space*.¹⁴ As Kitchin and Dodge (2011) analyse extensively, many spaces (physical, organizational, informational) are now ‘coded’. Their operations are structured through the software that processes data inputs of various sorts: the highly controlled space of the airport security queue is one clear example (2011, Chapter 7), entry into which is impossible without having met various data-related conditions in a prescribed sequence. This is an aspect of the rise, more generally, of the automated management of social processes. Unlike traditional surveillance, this form of control allows no gaps, since it operates through ‘a grammar of action’, that is, ‘a systematic means of representing aspects of the world [. . .] and an organized language for processing those representations’ (Kitchin and Dodge, 2011, p. 80, drawing on Agre, 1994). Under these conditions, the spaces Kitchin and Dodge call ‘code/spaces’ are *figurations* (in our term) of a particular, highly organized sort, driven by today’s complex forms of interdependency.¹⁵ Social media platforms *feel* like ‘spaces’ where, quite simply, we encounter others, but their existence is shaped by the underlying operation of platform software and its calculative infrastructure. Insofar as they create publics, these are ‘calculative publics’ (Gillespie, 2014, pp. 188–191), not that calculability itself is new: Weber (1978, p. 975) already saw this a century ago as ‘the peculiarity of modern culture’, but it has a *constitutive* role today that is unprecedented.

A related point can be made about the *time* of online media. Online media encourage us to operate in a distinctive time of required reactions related to the 'expected' rhythms of platform interaction: the Facebook timeline, the Twitter hashtag stream (Weltevrede et al., 2014). This time is not natural but the result of configuring *time-sequenced* data in a particular array designed to stimulate ever more interaction. We relate to this array as if it were a natural production by the human parties involved in the exchange, yet without the data-based presentation of the platform, there would be no mutual orientation in space-time, and so no 'interaction'. Many platform devices (such as email reminders) are designed to train people to rejoin the flow of what has been called 'social media time' (Kaun and Stiernstedt, 2014), should they slip out of it. This helps stabilize new data-based figurations that can function as 'social metronomes' (Neverla, 2010, p. 183).

The third translation operates at the level of the *self*. Each of us is familiar with the need to operate as a self under various descriptions: difficulties arise when contradictory descriptions of ourselves converge on a single interaction, and this much is familiar from social phenomenology. Unfamiliar from that classic work is the idea that each of us has not just a self-based identity (*vis-à-vis* the state or corporation with which we must deal), but a constantly updated 'data double' that is the resultant of the vast data stream that each of us generates continuously across various sites of data-tracking (Ruppert, 2011, p. 223, drawing on Lyon, 2003). The 'data double', with its built-in relation to multiple interdependent systems of data capture, depends entirely on the 'standardizing [of] classification systems so that they are comparable and databases can be joined up' (Ruppert, 2011, p. 221). It poses sometimes difficult challenges for individuals whose overall data stream may generate undesired or conflicting data, and, as already noted, the 'data double' is cut off from the shadow body of data that does not – perhaps cannot – feed into the relevant processes of calculation (Balka, 2011; Gillespie, 2014, pp. 173–174).¹⁶

The fourth translation operates at the level of *collectivities*. Data classifies, and so data processes not only work to specify individuals uniquely, but also generate countless groupings to which individuals are treated as belonging (groupings have 'aggregative power'; Ananny, 2015, p. 8). Whether these groupings correspond to anything that might be recognized by social actors as collectivities outside the process of data generation is an open question, but we are familiar with cases where the insistent use of data labels generates a type of action. Think of Facebook 'friends': some of these will have been friends before but many others are

likely to be those one has acquired through the practice of receiving and making friend requests on the platform. As Taina Bucher puts it, ‘friends have become a primary means through which *the production and occlusion of information* can be programmed’ (Bucher, 2013, p. 49). In everyday actions and adjustments, actors become attuned to maximize such data-based groups (another example would be numbers of Twitter followers). Collectivities are sites where, through data processes, new norms of action and reaction emerge. We already know however that data processes are creating new entities for governments and civil society actors to deal with. So when, during Brazil’s Junior Masterchef competition in October 2015, offensive and abusive sexual comments emerged online about a 12-year-old girl contestant, an NGO worker created a hashtag #primeiroassédio, which quickly generated more than 80,000 similar stories across multiple platforms (Gross, 2015). The importance of open hashtags as attractors to form political action has been noted for a number of major global protests, for example the rise of the 15-M and indignados movements in Spain in 2011 (Postill, 2014).

The final translation operates at the level of *organizations and order* and flows directly from the third. A problem for governments and corporations from the vast proliferation of continuous data streams is monitoring what counts as ‘risk’. According to geographer Louise Amoore, governments are increasingly relying not on judgement or deliberation (no longer possible perhaps in the face of such a vast mountain of ‘information’) but on an ‘ontology of association’ that ‘draw[s] into association an amalgamation of disaggregated data, inferring across the gaps [. . .] to derive a new form of data derivative’ (2011, p. 27). According to Amoore, this new ‘politics of possibility’ involves a fundamental abstraction *operating on the flow of time itself*: it is oriented to the predicted future, to ‘a population yet to come’. ‘Data’ must be rendered ‘actionable’ (2011, p. 29), which means selecting and excluding, ‘rul[ing] out, render[ing] invisible, other potential futures’ (2011, p. 38). These practices of exclusion become the basis for governing and ordering whole territories, a point to which we return in Chapter 10.

In all these ways, the domains of the social world, their practices and knowledge are reconfigured, in part, through processes of categorization based on data. The distributed complexity of these processes is a key contemporary example of the figurations of figurations and other large-scale relations between figurations that we argued a materialist phenomenology must understand. The emerging figurational order that results from this transforms the basis on which, in specific social domains such as the family and school, we are bound into figurations of interaction. It also operates,

potentially, to change the very stuff of the social domain that powerful actors such as governments see themselves *as acting upon*. In the next section, we consider some implications of this for how individuals and collectivities act in the social world.

7.3 How We Are in a Social World with Data

Online media present us with appearances which are highly consequential and with which we spend increasing amounts of our time. But these appearances are not ‘social facts’ in Durkheim’s sense, emerging from the flow of interpersonal interactions: rather, they are shaped, at least in part, by the economic and other external imperatives of the platforms through which they appear. When involved in online media, we interact on the basis of habits adapted to these platforms with others whose habits are similarly adapted. This is different from the entanglement people have always had with objects and infrastructures, and for two basic reasons. First, because, being rooted in everyday sociality and knowledge, online media comprise a space governed by *norms*, including expectations of legitimacy (van Dijck, 2013, p. 174). Second, because those norms emerge in relation to actions shaped by particular infrastructures of interaction and exchange, infrastructures already motivated by the corporate goals to *produce* and *stimulate* certain types of effects. This is not, of course, to deny agency, but simply to emphasize that, when we consider how we are in the world ‘with’ data, the ‘facts’ of what we do online, like all datafied ‘facts’, must be weighed carefully by reference to the motivated context in which they occur (boyd and Crawford, 2012): the goal of measuring what goes on in *any* online context as data for evaluation, and the constant stimulation of performance that yields *more* such data for measurement. You do not have to have been active for long on a platform such as Twitter to understand what Burrows and Savage (2014) mean by the ‘metricization of social life’.

Monitoring

Although online platforms themselves direct surveillance, the new monitoring affordances of digital communication technologies involve us all: from the basic question (‘why hasn’t s/he texted yet?’), to Googling others before we meet them, to more persistent forms of mutual monitoring, via multiple forms of ‘social media’ which include not only social networking sites, but also sites for posting ‘user-generated content’, ‘trading and

marketing sites', and 'play and game sites'.¹⁷ For that reason, legal theorist Helen Nissenbaum avoids the term 'surveillance' because of its strong pejorative associations with the state, and proposes the more open term 'monitoring' (2010, p. 22). Indeed the 'digital trail' is a major factor in the life of children, who in a country such as the USA are subject to continuous monitoring by their parents (Schofield Clark, 2013, p. 213). *Self-monitoring* and *self-tracking* (Klauser and Albrechtslund, 2014) is another important part of the picture, sometimes with specific goals (a sick person willingly accepting a measuring device that can warn a local hospital of symptoms of an impending heart attack), but often for more diffuse purposes. Emerging here in everyday practice is what Jose van Dijck calls 'the ideology of dataism [. . .] a widespread belief in the objective quantification and potential tracking of human behaviour and sociality through online media technologies' (2014, p. 2). In some institutionalized fields such as health, practices of continuous monitoring are a new and urgent trend; in others, like education, they build on decades of increased measurement and surveillance in schools (Selwyn, 2015, pp. 74–75). Indeed a consequence of the increasing capacity to combine vast database networks with huge calculative power is that 'the more data there is, the less any of it can be said to be private, since the richness of that data makes pinpointing people "'algorithmically possible'"' (Tucker, 2013, quoting Princeton computer scientist Arvind Narayanan, see Narayanan and Felten, 2014). The result over time may be a certain fatalism. We may come to accept a social world characterized by continuous and enhanced mutual monitoring as our *starting-point* for thinking about the social. If so, this is a new and clear example of what we have called *deep* mediatization.

Berger and Luckmann's principle that 'human "knowledge" is developed, transmitted and maintained in *social situations*' (1966, p. 15) – that is, situations where human beings, by virtue of their mutual dependence on shared resources, must *come together* to act and think – now carries a very different implication. The wave of digitalization (Chapter 3) has created a continuous plane of interaction based in technologies of mediated communication where, in principle, any actor, wherever located, can reach, and be reached by, the communications of any other. The temporality of social situations (Chapter 6) has also been transformed in more subtle ways, giving access to aspects of the flow of daily life that were previously lost, once experienced. The resulting enrichment of experience is inseparable from a new degree of *institutionalization of social form*. As Jose van Dijck (2013, pp. 6–7, added emphasis) puts it: 'through social media, these casual speech acts [of previous everyday life] have turned into *formalized*

inscriptions which, once embedded in the larger economy of wider publics, take on a different value'. Let us think a little more about the implications of this.

Datafication

Social situations, through their increasing involvement in ecologies of measurement and counting, are deeply implicated in data's status as a source of economic value. While many aspects of the metricization of social space are hidden to social actors, this cannot stop data-processing becoming entangled in the emotions of everyday life: 'more than mere tools, algorithms are also stabilizers of trust, practical and symbolic assurances that their evaluations are fair and accurate, and free from subjectivity, error, or attempted influence' (Gillespie 2014, p. 179). Translated into the language of classic phenomenology, algorithms and other aspects of the data infrastructure become a form of 'objectivation', part of 'the process by which the externalized products of human activity attain the character of objectivity' (Berger and Luckmann, 1966, p. 78). That is why an exposé of the *searcher*-sensitive features of the search engine results in which we trust is shocking (Pariser, 2011).

For Berger and Luckmann, institutionalization depended on 'the generality of the relevance structures' achieved in the production of knowledge (1966, p. 97). The interoperable metricized space of social media platforms and online interaction generally are coming to comprise, with little resistance, a new structure for generating social knowledge. According to a recent survey, of the US parents who use social media, 75 per cent go there for advice about how to solve parenting problems (Duggan, et al., 2015). Put another way, in the language of US pragmatism, the 'generalized other' (Mead, 1967 [1934]) that regulates social action is now increasingly sustained by a commercially encouraged flow of online exchange. Gillespie notes that 'algorithms impinge on how people seek information, how they perceive and think about the contours of knowledge, and how they understand themselves in and through public discourse' (2014, p. 183). If the algorithms associated with particular platforms, sites and practices have acquired legitimacy, anyone and any organization that depends for its power on legitimacy must deal with the consequences of 'what appears' *somewhere* in the unbounded, linked space of the internet. The management of this 'new visibility' (Thompson, 2005b; see Brighenti, 2007, and Heinich, 2012) becomes an all-consuming challenge, creating new challenges for the self: 'being spotted "by accident" against one's will

is not an option; missing out on purposeful display becomes a predicament' (Izak, 2014, p. 362).

An interesting example in the private corporate sector is the transformation of the hospitality industry, and its relations with customers and employees, through data-based systems for collecting customer reactions. As Orlikowski and Scott (2014) note, the *force* of platforms such as TripAdvisor (popular at the time of writing) is striking. In an industry whose principal asset is customer anticipation of a good service, a new mechanism of legitimacy in the 'space of appearances' is a profound shift, especially when the results of such recommendations are now distributed ever more efficiently as data within the corporate sector (Hayward, 2015). For political organizations that must sustain legitimacy through a narrative of control of their past but also their ability to manage a whole society's future, the dynamics of legitimacy and information are even more complex (Bimber, 2003). New forms of interdependency are emerging here, based not on digitalization but on datafication, and the link of datafication to categorization. 'People put things into categories and learn from those categories how to behave' (Bowker and Star, 1999, p. 311). Whether or not social actors are aware of the many levels of data processing that work to shape their contexts of action, they interact with data-based contexts such as social media platforms *as if* they were sites for social categorization and normalization.

It might seem, therefore, as if a hurricane is blowing through the domains of the social world and our knowledge of them, threatening to overturn every reference-point and previously bounded context of knowledge production. Two important factors constrain this chaos, although they do not in the process protect the social from datafication, and from the potential normative problems that it generates. First, there is the stabilizing force that, as Berger and Luckmann put it, 'institutions do tend to "hang together"' (1966, p. 81): meanings or (as they put it) 'relevances' overlap between contexts and institutional settings, constraining divergent interpretation, establishing norms of comparison, and (we would add) generating criteria for ensuring things do *not* appear. Arguably, this will work to entrench processes of datafication *into* social reality through a contemporary version of the circular feedback between structuring forces and structured outcomes that Bourdieu called 'habitus'.¹⁸ Second, there is the ongoing and irreducible tension between the grain of social experience and the forms in which it appears online. Social actors are likely to devote increasing efforts to contesting this tension with varying degrees of success.¹⁹ We return to the implications of these struggles for social order in Chapter 10.

Data and its Challenge to Social Knowledge

If we take seriously the possibility that the automated digital tools that measure behaviour and activity online are now a key part of everyday life's background, then phenomenology has been complicated irreversibly. A materialist phenomenology must register how everyday actors are involved in bringing the workings of those tools into their everyday awareness. All are processes of categorization. As Bowker and Star note, categories create a 'social and moral order' (1999, p. 3), but in the case of data-related categories, it is unlikely to go uncontested.

The figurations of figurations that make up the distributed data industries and the domains that rely upon them are transforming the space of social action. In this chapter, we have given key emphasis to the role of data in social media platforms. There is a reason for this: this is precisely where the process of constructing social reality is *remoulded* in detailed forms. As Kallinikos and Constantiou (2015, p. 73) say, 'social media platforms elaborate architectural arrangements through which communal interaction and daily living are transformed into data ready to enter the circuits of calculation and so-called personalization'. But social media platforms are only one area where data processes are becoming deeply embedded in the building-blocks of social action. Others are the growth of data-generating 'wearable' devices in the health sector, and this is just part of the wider 'Internet of Things', whose consequences for the texture of the social world are, at this point, uncertain.

We can put all this in a broader philosophical perspective. The philosopher John McDowell (1994, p. 84) considers how our 'mindedness' (our unfolding conscious relation to the world as human beings) becomes embodied in forms of interaction and resource, as part of what he calls our 'second nature' (the evolving set of social institutions which humans are disposed to develop alongside their first, biological 'nature'). For McDowell, the forms that embody the 'possibility of an orientation to the world' have a history (McDowell, 1994, p. 125),²⁰ and this history is constantly open to revision. This formulation enables us to frame with particular clarity the problem that data pose for the social world as conceived by classic phenomenology.

Berger and Luckmann assume that the forms of our 'mindedness' evolve *only* from the accumulation of sense-making by *human* social actors, but what if today there is an *alternative* 'embodiment of mindedness' (McDowell, 1994, p. 124)? What if 'data', in all their direct and indirect forms, are being installed as an alternative *and exterior* cognitive

infrastructure through which not only do we become minded, but the world becomes *mindful of us*, and everything we do? Since the data processes discussed in this chapter are part of an informational infrastructure that is being spread globally at huge speed, this amounts to a further stage in the deepening of mediatization. The very scale and scope depends upon the *delegation* of knowledge generation and knowledge application to automated processes. Once delegated, those processes become exterior to the process of social knowledge as classical phenomenology conceived it: they become what we might call, building on McDowell, a ‘third nature’, driven by the economic imperatives of the data industries and all the wider goals of capitalist expansion that, in turn, drive those industries. This ‘third nature’, if it is to order social life, requires social actors to adapt to it in a process that, following Agamben (2009, p. 15), we can call ‘subjectification’: the production of entities that can function *as subjects* within this new type of social order. One achievement of materialist phenomenology is to remind us that we *are* those entities – unless, that is, we refuse to be.

In sum, in the wave of datafication, new means for producing social knowledge have emerged with two key features. First, they produce ostensibly social knowledge through automation that is necessarily exterior to everyday processes of human sense-making. Second, they are oriented to goals, driven by wider economic forces, that are different in type from the goals that embodied actors are able to have, unless, that is, they give up on their autonomy entirely. The result is the emergence, unevenly at this stage, of a new kind of sociality – call it ‘computed’ or ‘platformed’ (Kallinikos and Tempini, 2014; van Dijck, 2013, p. 5) – that changes the starting-points for everyday reflexivity and sociological reflection. Social order (which classic social phenomenology set out to explain) is now, through its very conditions of formation, inhabited by a form of already ‘rationalized reason’ (Bernstein, 2002, p. 239, added emphasis),²¹ which cannot be comfortably integrated into the reflection of individual social actors. Positive readings of this world (what Papacharissi calls its ‘algorithmically rendered materiality’: 2015, p. 119) are possible, but they cannot cover over the fissure developing within the production of social knowledge itself. Not surprisingly, in response, some call for the ‘right to disconnect’.²² The only repairs to this fissure must lie in the agency of social life on various levels, to which we turn in Part III.