3

Classification Situations

OUR MACHINES CLASSIFY BECAUSE PEOPLE DO. WE COME TO KNOW and relate to the world by way of categories. To be human is to be able to recognize patterns and distinguish things according to type. Ordinary communication is the most immediate expression of this faculty. We refer to things through sounds and words, and we attach ideas to them that we call concepts. Some of our categories remain tacit; others are explicitly governed by custom, law, politics, or science. The application of category systems for the same things varies by context and in use. The notion of an animal species, for instance, might in one setting best be thought of as described by folklore and myth, in another as a detailed legal construct, and in another as a system of scientific classification.[1]

The way we divide up human populations into groups and classes has this quality too. Independent of whether they "carve nature at the joints," as philosophers like to say, what gives categories their authority—what makes them appear natural to us—is the fact that they are collectively

crafted, sustained, and enforced. This is not a deliberately collaborative process, of course. The most basic insight of sociology is that the joint action of human beings produces a social world that has the character of objective fact. The intersubjective, mutually reinforcing character of our expectations is the basis of the strange facticity, or quasi-objectivity, possessed by socially constructed things. Institutions flicker into existence in the light of mutual expectations and are sustained by them.[2] They are highly scripted, chronically available, repeatedly enacted, and presumptively real. They can also be obligatory and coercive, as in the case of classifications of caste, or gender, or race.

Social classifications are entrenched in people's emotions, in their bodies, and in their everyday practices. This makes them hard to change. But change happens anyway. It has to. Intersubjective expectations must be constantly re-created and revalidated in practice. This process is neither error free nor uncontestable.[3] While the most important categories and classifications are deeply entrenched and have a "systemic" character, it is not the sort of system that is necessarily coherent or perfectly self-replicating. Social life is messy. Since some social categories are advantageous and others detrimental, people struggle over their definition. They press to be fitted into one type rather than another.[4] They come up with new categories. In Ian Hacking's felicitous phrase, people are "made up" and remade all the time that way. New concepts come along to identify with and be identified by.

Certain kinds of classifications, typically those applying to human or social collectives, are "interactive" in that "when known by people or those around them, and put to work in institutions, [they] change the ways in which individuals experience themselves—and may even lead people to evolve their feelings and behavior in part because they are so classified."[5] These "looping effects," in turn, may further transform how institutions intervene in people's lives, perhaps even upending the classification system itself. Human societies are forever being destructured and restructured by the continuous interactions between classifying institutions and the people and groups they sort. This is especially the case when categories are public, visible, or legally recognized. A public or official status is a mark of objectivity that also makes it easier for collective action to find a focal point. More deeply, present-day liberal democracies tend to have a "politics of recognition" that accords moral importance to ideas of self-categorization, personal authenticity, and dignity. This makes it easier for many formerly primordial categories and classes to be contested.[6] It also makes possible the sort of entrenchment and sacralization that is distinctively modern, centered on the individual.

The abundance of rich, multidimensional, digital data and the means to analyze it has profoundly affected how social categories are made and how people sort themselves or are sorted. Relative to their analog predecessors, classifications produced by computer code sifting through digital data are more likely to be anchored in direct measures of behavior. They also tend to be more fine-grained, inductive, and flexible. And they are often more opaque, in the

sense that they may depart from established categories and fail to be readily interpretable in terms of them.

An engine of social differentiation sits on top of this evolving—but increasingly precise and routinized—infrastructure. As people move through the world, reams of data are assembled about them. Analytics tools begin the work of arranging longitudinal and cross-sectional profiles and disaggregating and sorting these "data doubles" into predicted categories.[7] While you exist as a physical person in the world, your data double is the representation of you, your tastes, and your actions that can be reconstructed in whole or in part from the records and traces you leave behind. The categorical systems themselves may vary depending on the purpose at hand—people may be sorted into types of person, market segments, risk brackets, expected value targets, and more. But what unites these systems is that they are actionable. Writing about market research in the early 1990s, Oscar Gandy Jr. called this now ubiquitous process "the panoptic sort":

> The panoptic sort is a system of disciplinary surveillance that is widespread but continues to expand its reach. The panoptic sort is a difference machine that sorts individuals into categories and classes on the basis of routine measurements. It is a discriminatory technology that allocates options and opportunities on the basis of those measures and the administrative models that they inform. The panoptic sort has been institutionalized. It

> is standard operating procedure. It is expected. It has its place. Its operation is even required by law. Where it is not, people call out for its installation. Its work is never done. Each use generates new uses. Each application justifies another. It is efficient, having largely been automated. . . . The panoptic sort is a system of actions that governs other actions. The panoptic sort is a system of power.[8]

As this "cybernetic triage" unfolds,[9] the analysis of tracked and classified behaviors forms the basis of differential treatment, thus affecting social stratification through the allocation of similar sets of opportunities to similarly situated people—what sociologists since Max Weber have called "life chances." Organizations apply labeling and scoring methods to slice people into groups and ranks suited to their particular ends. People find themselves more or less comfortably fitting into these categories. Often these categories are not—or sometimes, as a matter of law, cannot be—constructed from standard demographic classifications such as race and gender. Instead they tend to be behavioral and probabilistic, predicting the likelihood that people will do or like certain kinds of things, or the position they may reach on a particular scale. In that sense these methods construct a "postdemographic" classificatory infrastructure.[10]

We call these outcomes *classification situations.*[11] These are positions in a generated system of categories that are consequential for one's life chances. Classification situations are not merely approximations of commonly identified social groups and identities—though, of course, they may overlap substantially in specific cases. Rather, they are independently generated taxonomies that can come to have distinctive and consequential effects on the "outcomes" people experience in life.

Naming and Ordering

Social order, like computational order, comes in many varieties. To begin, we can distinguish between different ways to classify and measure things in the world. *Nominal* judgments are oriented toward essence. They define what something is. They are judgments of type, labels that describe some intrinsic quality, perhaps in relation to other sorts of things. Think of the efforts by Carl Linnaeus to classify the flora and fauna of the world in his *Systema naturae* early in the eighteenth century. Specimens from all over the world were compared and carefully organized in relation to each other, following rules of resemblance and difference. Later, nonbiological aspects of human life were subject to the same process of collection and classification. By the end of the nineteenth century, the accumulation of human artifacts and knowledge of exotic languages and rituals allowed for the ostensibly scientific definition of cultural differences between human groups.

Nominal judgments require decisions about the criteria for resemblance and then some sort of interpretive assessment of where something belongs. Resemblance legitimates lumping together, but the basis for establishing proper resemblance is always difficult and often contested. Moreover, nominal categorizations tend to work with ideas about ideal or typical members of a class, on the one hand, and departures from some default or standard, on the other. Even though nominal classifications are in principle just names or labels, not numbers or positions, they may still express priority and hierarchy. This can be explicit, as in terms of relations in a tree of categories and subcategories. Or it can be implicit, as when default orderings lead to marked and unmarked categories, with the unmarked case—the one you don't have to point out or specially label—being intuitively prior to the marked.

Nominalization may also be informal or formal. Bottom-up processes driven by homophily and more top-down institutional rules for naming both create categories. In the former case, when left to their own devices people have a tendency to associate with others who are similar to them in various salient ways: birds of a feather flock together.[12] In the latter case, organizations or institutions group like with like, too, in the light of their own goals. While informal sorting might arise from local actions and choices, nominalization finds its formal representation in clustering and classification methods. A recommendation algorithm might find each person's "nearest neighbors" in terms of positions in an abstract space defined by patterns of consumption or purchasing or voting or any such quality. People with similar tastes or other characteristics in common can be lumped together into more or less internally homogeneous groups.

Ordinal classifications, meanwhile, are explicitly organized by measures of position, priority, or value along some countable dimension. Something is ordered, rather than simply named, and distinctions are expressed in terms of scores or ranks on that scale of measurement. In everyday society, ordinal measurements can be found in graded examinations, standardized tests, competitive sports rankings, occupational pay scales, and so on. Almost any repeated activity can be converted into some sort of score, even if only in terms of a count of frequencies.

In both social and technical practice, distinctions between different schemes of measurement are potentially fluid depending on the goals of the measurement. Conversion pathways are common. A continuous measure may be simplified to a numerical rank, or binned into some number of ordered categories, or dichotomized into a binary classification. Nominal clusters or categories will often be created on the basis of calculating continuous distances in some multidimensional space of relative similarities or dissimilarities in conjunction with the application of some criterion for determining when things are similar enough to be placed in the same class.

From the point of view of social order, the most basic kinds of processes are those of *nominalization* and *ordinalization:* the naming of kinds and the designation of ranks. They are not, to repeat, inimical to one another. In practice, ranks can be collapsed into types. Similarly, those in search of social examples of purely nominal classifications will have a hard time finding cases that have not, somewhere, been treated as explicitly or implicitly ordered.

It is tempting to think that it is the act of numerical measurement as such, just the sheer fact of quantification, that makes ordinalization pernicious. This is a mistake. As social phenomena, naming and ranking are much more general than quantification. Catechisms, shibboleths, purity regimes, ritual compliance, degradation ceremonies—in short, symbolically infused distinction making in all its forms—can serve in the place of numerical measures of position. Insofar as they are about distinguishing better from worse, as opposed to simply affirming the uniqueness of every single thing in the world, qualitative modes of classification can be as powerfully disciplining as quantitative ones. What is of real interest is the fusion of socially fundamental processes of naming and ranking with novel tools and data for carrying out those tasks. Large-scale measurement allows for thinking about scores and ranks through the lens of small differences and high dimensionality. What does it mean for computers to intervene in the business of seeing and organizing society?

Testing and Matching

For one thing, machines, like people, are prone to error. A machine learning (ML) classifier is a function or procedure that assigns a class label to a case. Classifiers are not foolproof. They can be tricked into misclassifying objects in photographs, or elements of speech, or kinds of person. These failures can be consequential. Out in the world, false arrests have been made on the basis of false predictions.[13]

A large literature in computer science, law, and social science finds that the social machinery sustaining the deployment of algorithms might contribute to the reproduction of categorical inequalities around gender and race.[14] A key source of trouble is the need, in many methods, for some baseline or "ground truth" of correctly classified items or training data that the model must be built with. It can be hard to find datasets of sufficient size or quality to train models on, and the circumstances of their production are often opaque.[15] Even if they are large and varied, training data is often lacking in ways developers are unaware of or do not think to consider. The result is poor or (to use the most common euphemism) problematic performance in real-world applications. For instance, early commercial facial recognition programs were prone to misgender Black women. Similarly, Amazon had to scrap a much-vaunted ML recruitment tool after the software "learned" from the company's practice of hiring almost exclusively men in technical positions and systematically weeded out the résumés of women applying for these jobs.[16] US law disallows such behavior—though it does not prohibit the use of algorithms to manage hiring and firing.[17]

There are no easy technical fixes to what is at root a sociological problem. Training data, almost necessarily, comes from a social world already structured by deeply entrenched categories and classifications, with various degrees of formality and normativity. Linguistic corpora used to train text classifiers, for instance, are often sourced from a core network of related sources, such as Wikipedia. This can have knock-on effects when the data is subsequently used to prepare models of all kinds.[18] Predictive methods used to determine who will make a good employee, a lousy parent, or a dangerous recidivist are, in the end, built on top of histories of discriminatory practice and asymmetrical surveillance. People who frequently encounter the police, the criminal justice system, social services, or other institutions of social control are likely to be overrepresented in training data drawn from these agencies. These organizations may share data with one another in an effort to enhance the accuracy of their data-hungry methods, which only tends to make things worse. And, of course, the outcome of interest, the thing "the algorithm" is trying to optimize on, matters a great deal too. When Virginia Eubanks studied a method for assessing a child's risk of being abused in Pennsylvania's Allegheny County, she found that cases of child maltreatment-related fatalities and near fatalities, which are of greatest concern to the state, are extremely rare.[19] To produce useful results, the county's predictive model had to optimize on more common outcomes. In the end, two proxies for child harm were used: the likelihood that another call will be made about the child to the abuse and neglect hotline, and the likelihood that the child will be placed in foster care.

The problem is that what motivates both referrals and foster care placements in practice is difficult to disentangle from a general condition of poverty. For instance, malnutrition, lack of decent housing, or lack of health care are considered neglect. As a result, poor and minority families were overrepresented in the outcome variables. The urban poor also interacted frequently with public services, and

so they, too, supplied a disproportionate number of the available data points for the predictive variables. Suburban middle-class families, by contrast, were nowhere to be seen in the data. This is not surprising: they live in places that are out of the reach of social services and deal with problems through private insurance and experts, whose interventions, in the name of privacy, are not recorded in public databases or shared with public institutions. And so the system returned few red flags for them. In sum, oversampling, data redundancy, and an ill-defined objective function created a predictive instrument of great power to profile those who have the least to get by, in effect stigmatizing their social condition as an immoral and dangerous one—and adding yet another layer to the seemingly incomprehensible moral logic of a system that will financially support a child's foster parents, but not their parents, to raise them.[20]

The world is patterned, and the social world is no exception. Nothing, no matter how mundane or tacit or confidential, is a priori irrelevant as a basis for honing a classification scheme. But what sort of structure will be found by a classifier? One danger is that unnoticed but strictly irrelevant features of the data will end up becoming the basis for identification and labeling. In the field of medical image classification, for example, a widely pursued research goal is to develop classifiers able to correctly identify pathologies from images at a level of accuracy better than trained doctors or technicians. In one study, researchers sought to train a deep-learning network to reliably identify cases of pneumonia from chest X-rays. The training data consisted of about 160,000 images sourced from three different locations: a network of hospitals associated with Indiana University, Mount Sinai Hospital in New York City, and the National Institutes of Health Clinical Center in Bethesda, Maryland. While the system performed quite well, it turned out that a key reason for its success was that, based on consistent similarities in the production quality of the images, the model was able to detect which hospital an X-ray originated with and adjust its predictions accordingly. Different hospitals had different rates of pneumonia, and they also had subtle differences in the look of their X-ray images. The model effectively exploited the latter feature to predict the former. It learned, but not in the intended way.[21]

The general problem of spurious correlation plagues data analysts everywhere. Its manifestation in settings like this is distinctive and interesting, however. Across many different kinds of classification and prediction tasks, the likelihood of specious associations producing "shortcut learning" is high.[22] It happens because while these methods are astonishingly good at pattern detection, their mode of establishing similarity and difference is quite different from the way people recognize and classify things. For example, it can be possible to carefully craft an "adversarial" image that breaks the classifier by exploiting the way some internal layer works. The image may look nothing like the items being classified, or the change might be invisible to the human eye, or indistinguishable from noise. But the classifier does not look at things the way we do, and so it breaks in unexpected ways.

A second and more comprehensible mode of failure is the sort that tripped up the pneumonia classifier. In this case, real structure—that is to say, features that users can see in the data and whose predictive relevance they can understand—is used in an illegitimate way. In the pneumonia case, the classifier learned what hospitals looked like when it was supposed to be learning about what pneumonia looked like. Because the character of the hospitals was correlated with the distribution of pneumonia, taking it into account improved the model given the data it learned on.

Even boring, run-of-the-mill methods of the old-fashioned kind are bedeviled by problems of poor model specification, unwarranted inference, and spurious associations. But the new methods make things even more tricky. In the case of the pneumonia classifier, we can straightforwardly say what went wrong in the training of the model and easily see the reason why it would not do to have it implemented that way in practice. But the implications for more purely social data are not so clear. By inciting organizations to treat all data as useful, and by developing tools with a hitherto unmatched ability to find patterns, software engineers have created tools endowed with new and somewhat alien powers. Together the social origins of training data and the unblinking eye of a deep-learning classifier combine first to translate the social world into a model and then, potentially, to recombine and reconfigure categories based on what it sees. The particular features that these systems detect and act on are likely to reflect social realities of *some* sort. But how exactly they do this, which features are selected, and whether the result is fair or just is another matter.

One of the chief motives for using these tools in the first place is their ability to take large volumes of data and see things that a human user or even a traditional statistical analysis cannot. But when the model is opaque, the question of its performance is intrinsically tricky. If we are attempting to accurately identify handwritten numbers or things that are cats, then at least validation is straightforward on the basis of spot checks and comparisons to labeled sets we are confident about. In the case of more challenging classifications—kinds of employee; species of credit risk; varieties of recidivist; terrorist or not—the process is much more murky. Deep-learning models might classify based on weighted combinations and transformations of hundreds or thousands of features, leaving users with little idea which conventionally identifiable features are really important. Worse, deciding whether the result is fair or unfair to any particular person classified is even more difficult.

If we trust the model, we should go along with its output even if the results seem surprising. Again, there is already a tendency to treat the output of conceptually much simpler approaches, such as those used to estimate credit scores, in a somewhat Delphic manner. Because the fine details of credit score estimation are trade secrets, information about how to manage and improve one's score tends to take on the aspect of lore. This or that behavior is supposed to help; perhaps you could try doing it. When

properly "deep" methods are applied in circumstances like this, not just the methods but also the resulting classifications may appear strange and uninterpretable. People find themselves in a system of categories where their own location is determined by difficult-to-understand methods, and the label they end up with might not even be recognizable or known to them. A steady income or a bank balance is replaced by a synthetic risk score. A passport or visa is trumped by membership on some watchlist or other. Generally, placement on such lists depends on people, organizations, objects, events being identified through patterns in the data pertaining to them.

Rule-based "algorithms" and "clusters of attributes" have always served to draw boundaries between kinds of people and things.[23] But modern methods may be especially unnerving—not only because they are difficult to audit or contest, because so much of the data is ad hoc, or because they pry so deep into people's lives, but because they reconfigure the meaning of categories people took for granted and reorganize what can be done with them. For instance, the significance of what John Cheney-Lippold calls the "right of the algorithm" and what Amoore calls the "deep border" of ML, may supplement and even supersede national citizenship and the physical border.[24] The analogically derived certainty of belonging somewhere may be shattered by one's place in a dataset and the inferences made from it. The risk score associated with a recognized face's cluster of attributes may determine the movement of the turnstile at the airport. One's data double crosses the border first.

We have already mentioned some consequences of the partially achieved, partially assigned categories emerging from algorithmic methods. Sorting and slotting procedures of various kinds shape access to goods, services and opportunities across many institutional spheres, from employment, health care, insurance, and education to housing, citizenship, credit, social welfare, and more. They are also busily reformatting the structure of ordinary sociability, from opportunities for friendship and dating to getting around town at the weekend. But we often fail to appreciate the extent to which these technologies now mediate the whole social process itself. They identify new classes of people, reformat identities, help control social action, and produce new criteria for truth telling and ethical judgment.[25] These classification situations are created across different settings—in markets of all sorts, as well as under the eye of the state. In market settings, as we explore in Chapters 4 and 5, their main purpose is to assist in extracting value. In connection with the state, as we shall argue in more detail in Chapter 7, they are used to establish qualification standards for social inclusion and appropriate levels of merit and desert. This tendency to better score and rank users on multiple dimensions, and the subsequent linking and integration of measured profiles across domains, conjures up images of individuals positioned in a vast multidimensional space of personal and behavioral characteristics, each one a vector of characteristics with an associated set of possibilities, prices, and experiences.

Eigenvalues and Eigencapital

Think of the totality of your interactions where behavioral and interactional data is recorded and collected. All of those traces represent a kind of resource. It is accumulated over the long history of your recorded actions and choices, built up from traces left on everything from social media to credit bureaus, shopping websites and fidelity programs, courthouses, social welfare agencies, pharmacies, and the content of emails and chats. It incorporates whatever value is in your social network, along with synthetic measures of your trustworthiness or accountability in the world. It is heterogeneous and multidimensional and, of course, it is not all gathered into a single place or condensed down to a single quantity. But in principle it might be. It might take the form of some vector of information that summarizes your situation and value across many features—something that compactly represents your position in the multidimensional space of classification situations. It would, in short, characterize your social location. In data analysis, this procedure often involves the decomposition of matrices of data into orthogonal eigenvectors, the better to characterize them. (One of the more old-fashioned translations of *eigen* in the terms *eigenvector* or *eigenvalue* is "characteristic.") From an individual's point of view, these quantities and their representations would be a kind of resource. Call it *eigencapital*.[26]

Following the work of Pierre Bourdieu, social theorists and researchers have named many forms of "capital" over the past forty years—absurdly many, perhaps. Do we really need another sort? There is human capital, economic capital, cultural capital, social capital, bodily capital, and more besides. Each type begins with the same initial idea. People may possess some quality or capacity that directly or indirectly benefits their prospects in life, over and above a direct measure of income or wealth, class position, or membership in some demographic category. These qualities or properties might be a little harder to measure, but they are quite real and they can, in some more than metaphorical sense, be cashed out or converted into more conventionally material benefits. An admirable skill, a network of helpful friends in desirable places, the right sense of good taste, or even an appealing face: all can be "capital" of a kind, because each is defined in some context where what is "admirable," "desirable," "right," "good," and "appealing" reflects whatever the entrenched distribution of assets, opportunities, and status happens to be.

On this view, higher-status people tend to think of and present their tastes, abilities, and achievements as the unforced outcome of their natural talents. Even if they are not always acting in consciously strategic terms, this is one of the main means through which they legitimize their social position and, in the process, stay ahead of the competition. A sense of good taste and the right "feel" for what works and what doesn't in particular social settings can be a valuable kind of resource. That is cultural capital. Bourdieu was fascinated by the conveniently natural fit between people's backgrounds and their tastes and ambitions. His theory of practice tried to get a grip on how

people's class position organized or structured their tendency to speak or act in particular ways, to develop some tastes rather than others, and to think they were "cut out" for some kinds of work while not even considering some alternatives as possibilities. Bourdieu's concept for the individual's experience of this process is a notoriously slippery one: the "habitus." This is the "feel for the game" or "sense of the rules" that you carry around with you. It emerges from your experience. It structures your dispositions and your actions. When a situation feels comfortable or a decision feels like the right thing to do, the habitus is the feeling in your gut that grounds that experience.

In *England in the Age of the American Revolution* the historian Lewis Namier remarks that "A man's status in English society has always depended primarily on his own consciousness. . . . Whatever is apt to raise a man's self-consciousness—be it birth, rank, wealth, intellect, daring or achievements—will add to his stature; but it has to be translated into the truest expression of his sub-conscious self-valuation: uncontending ease, the unbought grace of life."[27] It is the process of generating the apparently "unbought grace of life" that fascinated Bourdieu. Direct efforts to acquire, display, and demand deference to one's taste and learning tend to fail. Transparent use of one's wealth to buy status is effective but crass. Better to put it to work in a less blunt fashion and allow it to express itself more indirectly. Best of all would be to be able to put in the time to gradually acquire the accoutrements of good taste and right thinking, and then "forget" they were acquired at all. Like wine left to mature in its cask, what begins as deliberate cultivation eventually manifests itself as the wholly natural expression of authentic inner qualities.

In Bourdieu's picture, this process mostly happens during the long period of formal education. It takes money and, above all, time—two resources that not everyone has in equal measure. One kind of capital, the straightforward monetary kind, is slowly converted into another, the cultural kind. Education gives you public credentials, certainly. In Bourdieu's terms, this is the "institutionalized" form that cultural capital takes. But it also gives you "embodied" cultural capital that you express without needing to show people your college diploma. In the best cases, your habitus lets you comfortably fit into an already-structured social world, one that in the limit case smoothly meshes with your talents and skills in a seemingly natural, spontaneous, effortless manner. On the other side, and of equal interest to Bourdieu, are the times when things do not mesh and you are left feeling out of place or awkward, knowing—and painfully feeling—that you do not really belong.

Bourdieu insisted, and critics have often complained, that cultural and symbolic capital are not easily measurable. In his view they are primarily "known by their effects"—that is, by the extent to which they allow actors to accrue specific material and symbolic profits, such as money, power, or authority. This lends the Bourdieuian approach a flexibility of application that skeptics feel makes a virtue of endogeneity. But there is also something deeply true about the insight that the organization of the outside world, with all its unequally distributed resources and often obscure rules, gets inside people in a way that makes life go more

smoothly for some than for others. This tends to encourage us to see our experiences as manifestations of a natural order and to see our actions as expressions of natural talent or innate goodness rather than as a kind of side benefit of lucking into the right background. In this respect, Bourdieu's analysis of the forms of capital and their role in social reproduction can be seen as an effort, well before its time, to theorize the now ubiquitous concept of "privilege."

Eigencapital is a little different. It has its origins in particulars—in the totality of one's interactions with the digital economy—but it has a generalized, relational character that is not found in the usual list of novel forms of capital. Its specificity is retained in the variable way that it is applied or expressed in specific contexts. The meaning of one's "score" or "stock,"[28] so to speak, depends on the specific setting one is in at the time. More than just an image or metaphor, it is also a contingently realized empirical phenomenon. Estimating and using something like it in practice is not a fantasy but a genuinely huge engineering problem subject to failure or incomplete realization. But in principle an individual's eigencapital is calculable from all of the digital information available about them, encapsulating the totality of their relations as expressed through digital traces, ordered and characterized through numerical methods. In the Bourdieuian manner, it, too, is visible through its effects. Advantages accrue to those who accumulate it. At present it exists mainly *in potentia*. For it to come closer to what we have in mind, present-day tendencies might fuse into a more encompassing system of measurement.

Like cultural capital, we can think of eigencapital as taking *embodied, objectified,* and *institutionalized* forms.[29] In its *embodied* form, eigencapital is expressed in durable dispositions. This is the "habitus," incarnated directly in the body and the mind, and expressed in the overall presentation of self. The well-situated individual naturally inhabits their data double. They feel the benefits of eigencapital directly, automatically. Reputation is no longer confined to a local community of peers. The trust the individual feels confident extending is no longer circumscribed by a concrete social network. Instead, they carry it with them in their bag or on their phone. Moreover, to the extent that the practical expression of such a resource works successfully—and as we have been emphasizing, getting it to work is a huge technical challenge—the process fades into the background. The ideal, in fact, is much like the National Security Agency's defense of its methods of ubiquitous but invisible surveillance. You do not see the bad actors who tried to use your card but were automatically denied. You do not have your integrity questioned by a salesclerk. The camera takes a quick look at you, and you can board a plane or cross a border.

When things go wrong with systems like this, their automated decision-making will seem stupid, rigid, or rule bound. Why doesn't my card know I am simply in a different city, trying to buy a meal? How come my online transaction was flagged just because I am buying something a little unusual as a gift? But when these systems work properly, instead of throwing up a roadblock they smooth the way for us in a pleasant, barely detectable manner. They allow transactions to happen in the blink of an eye; they prevent fraud; they enable good matches; they help us make good choices. The fortunate consumer experiences this as a well-deserved, delightful form of ease. In a way, the infrastructure of eigencapital revives an old kind of privilege. It promises the portable, universally recognized trustworthiness and good reputation of the gentleman abroad, sustained by his word and letter of introduction. It is the "unbought grace of life," but in a newly quantified and nominally egalitarian form. It takes an aspect of life long familiar to the very rich, a specifically personal attention to one's needs, and spreads it just a little further down the social hierarchy, providing a still-small minority with the pleasingly exclusive experience of recognition and authorization.

The varying amount of time different people must spend to access various services offers a good example of the embodied dimension of eigencapital. Sociologists have long studied queues and lines as structures that allow for both control and status. Who can be made to wait, for how long, and for whom? These simple questions are surprisingly robust indices of the structure of interpersonal relations. "The distribution of waiting time," Barry Schwartz remarks, "coincides with the distribution of power."[30] Organizations that serve the poor have a low staff-to-client ratio, so people wait to obtain service. They line up to get food or to climb on the bus. The mark of being rich, on the other hand, is the ability to spend the time of others.[31] In that case it is the staff who wait upon the client, literally and figuratively. It is no surprise that this most basic of social dynamics has been amplified by the availability of data on who exactly is in the queue. With the rise of predictive analytics, the social differentiation of wait times has been automated. For corporate call centers, a first and easy step away from a first-call, first-served basis was by means of some nominal approach. Premium members might get a special number to call. But there is no need to stop there. By modeling various salient features of a customer's account, a financial, telecom, or airline company might easily produce a "customer importance score" expressing the current and likely future value of the person on the line. This can be used to determine response speed and quality of service. It will likely be positively correlated with their wealth, but predictive models may also include variables such as the urgency of the problem, some measure of the person's importance to the company, or their estimated likelihood of dumping it.[32] Whatever the prediction method, it is typically opaque, so the outcome tends to be experienced as fate.

In its more tangible, *objectified* form, eigencapital is realized in transmissible property. Over and above making ordinary experiences run more smoothly, it helps provide access to goods and services, at better prices, and with better social consideration. The well-informed parent carefully manages their child's credit so that the child, too, will appear trustworthy, even if they are not. The hopeful driver installs monitoring software on their phone to lower their car insurance premium. The objectified form is a reminder that eigencapital must be produced in specific ways. For those who have not been trained that way, its accumulation takes effort, discipline, and often money. This form is all about legibility: You can work toward a better position by paying attention to producing "good data" for yourself. But you need to know what good data is in the first place. This recalls modern debates about the auditability of algorithms—the right of people to know the rules by which they are judged, and the right of experts to inspect these rules. Many algorithms embrace this semi-objective character. Credit scoring companies publicize their evaluation criteria, even if the final formula remains a trade secret. Your banking app will helpfully dispense advice on how you can do better. So does your health and fitness app. Eigencapital in the objectified form is prescriptive, and it feels like work. But eventually this labor does pay off, and assiduous effort fades into the background and becomes second nature. This is when the objectified form folds into the embodied form and the benefits of eigencapital—the shortened wait on the phone, or at the airport counter—are experienced as ordinary, well-deserved and effortless.

The simplest eigenstatus of all is an indicator of mere presence or absence, observed or unobserved, on some dimension. If you are not included, you cannot be measured or assessed. And so, at the most elementary level, platforms and other systems strive to include everyone under their purview. Being outside carries an increasingly high cost, as economists have long noted in the context of network effects, and sociologists in the context of the digital divide in access to the internet. But as data collection becomes ubiquitous, so do the expectations of being seen. As the physical world fills with sensors, and people live increasingly hybrid lives, resisting these expectations is hard in practice. Furthermore, it is not necessarily desirable: being invisible to digital infrastructures is suspicious, and organizations code it negatively. The failure to engage and properly care for one's data double is a moral fault at best, a sure sign of illicit behavior at worst.[33] Invisibility is as much a trap as visibility. Simply avoiding debt will not do: you'll just end up with a bad score, which you can only hope to improve by subjecting yourself to more intrusive data inquiries. Likewise, a user's failure to like, share, comment, and message others on a social media platform will prompt a demotion of their contributions relative to others who are more involved. In Chapters 6 and 7 we will look more closely at why this is the case. For now we can say that people who are inactive are of little value to organizations. They are expensive to know and unprofitable to manage. By the systems' standards, they perform poorly. And so they get punished for it. Likewise, public institutions also increasingly operate according to a logic that privileges electronic visibility: the extension of rights depends on digital incorporation and the steady production of data about oneself.

To a first approximation, the acquisition of eigencapital depends on being seen by, and making oneself visible to, digital architectures—with a credit card, an email account, a smartphone, a smart speaker. As personal data of various kinds is pumped throughout the internet, the moral injunction shifts from obligatory incorporation to proper data management. People are taught, often from a young age, that they must "build credit." In the United States, twenty-one states require schools to teach financial literacy. Banks and credit reporting companies helpfully supply educational resources about financial probity. Educators share their preferred pedagogical strategies on blogs and specialized websites, while tech entrepreneurs have designed credit apps specifically for children.[34] These materials generally emphasize not only the benefits of early financial incorporation and visibility but also what it means to work toward a scored position in the world. A generic knowledge of relevant algorithmic categories, constant monitoring of outcomes, and quick intervention in times of crisis are part of the expected posture of the datafied citizen. People's relationship to their personal data exists in a moral universe shaped by both the direction provided by institutions and their own personal conjectures about how to do well in life. Far from being passive, they are emotionally involved in systems of data production and management and sometimes take great pains to develop strategies that "feel right"—or properly balance the need to be visible with the desire, however hopeless, to safeguard their privacy.[35]

Training a population to embrace its own ordinalization can sometimes take the form of a bold exercise of political will. In 2014, the Chinese government declared its intention to deploy a national "social credit" system anchored in general measures of "honesty." The project spearheaded a myriad of experiments by municipal governments harvesting information from dozens of subunits and dedicated local committees on a broad array of punishable and honorable behaviors. Legal judgments against one's person, politically sensitive behaviors, incivility, or moral turpitude can downgrade one's score, while volunteering, government work, or making a donation contribute bonus points.[36] The criteria are generally public, so people know, on the whole, what they are supposed to do (or not do). The specific implementation logics vary a great deal from city to city, but the core principle across various systems and locales is that citizens or organizations whose score does not reach a certain mark face practical hurdles (such as travel restrictions, and exclusion from certain occupations, markets, and services), and, sometimes, public shaming. Those with good scores might experience public praise and faster processing across institutions. Some of these systems rely on an infrastructure of paper pushers, while others use digitally sourced data supplied through partnerships with technology corporations, such as ride operators.[37]

A single summary number is the most minimal form of eigencapital: only one value matters. But when formally connected to a set of rewards and punishments, it can be quite powerful as a governance tool. Indeed, a broader ambition in the Chinese case is to link data and scores across systems in an effort to regulate the behavior of entire populations in the name of collective harmony.[38] This is the theory. The reality is more mundane. Despite the headlines, many of these systems still lack in both capacity and authority.[39] For now, it seems people's ability to live their lives is stymied by more mundane roadblocks that quite ordinary methods of monitoring throw up—or alternatively by darker, and much more opaque, deployments of digital surveillance.[40]

This brings us to the third state of eigencapital. In its *institutionalized* form, eigencapital may exist as a measured quantity that may be widely used and circulated. Here, what matters is the general recognition of the measure across institutions. A single measure condensed from a collection of noisy data sources is obviously a rougher and more approximate token of one's "true" eigencapital. To produce it, differences must be flattened, equivalences must be made between incommensurable qualities.[41] The resulting one-dimensional measure will contain much less information than its multidimensional parent. In the same way that, for Bourdieu, a diploma is a highly reduced and unsatisfying way of apprehending the concept of cultural capital, any particular score or rating will be only a rough approximation of a person's general eigencapital. The benefit is that it makes decisions easy to automate. In the United States and a number of other countries, credit scores have attained this generic status and social significance. They can be bought and sold as such, and combined with other measures to produce superscores. They are, for instance, routinely used as an input for "off label" risk prediction in other markets, such as insurance, housing (tenant screening), or dating.[42] China's experiments in social credit have a similar ambition, though they have yet to attain this kind of generality.

Though the particular conceptualizations vary, the quest for the one score that will bind them all is quite universal.

A search for "trustworthiness" through the patents database maintained by Google shows hundreds of applications related to the algorithmic scoring of individuals and entities going back to the early 2010s. The earliest were filed by the behemoths of e-commerce, such as Amazon and eBay. In the same way that the widespread diffusion of credit scoring—a risk prediction tool—enabled the massive expansion of credit,[43] a platform's success depends on its ability to guarantee the integrity and responsibility of both its sellers and its buyers. The practice is perfectly defensible. Who, after all, wants to deal with an annoying customer, who complains about not receiving packages or returns them broken? Who wants to contract with a seller who cannot complete an agreed-upon deal, or who ships subpar products?

Similar concerns apply across digital ecosystems. After dealing with the Cambridge Analytica scandal in 2018, Facebook announced that it had started giving users a "trust rank" based on their propensity to flag verifiably true news stories as fake. Around the same time, the housing rental company AirBnB, which already uses renters' credit scores to give them an initial rating, was granted a patent for a "trustworthiness and person compatibility" score. According to the patent, "text authored by the person or that provides information about the person" can be used to "indicate that the person has created a false or misleading online profile, provided false or misleading information to the service provider, is involved with drugs or alcohol, is involved with hate websites or organizations, is involved in sex work, has perpetrated a crime, is involved in civil litigation, is a known fraudster or scammer, is involved in pornography, has authored online content with negative language, or has interests that indicate negative personality or behavior traits."[44] Most organizations are interested in weeding out the most undesirable among their users, employees, citizens, claimants.

But even in the most totalizing systems, there is always an outside. Those who try to evade being measured and classified, as well as those who perform poorly by the system's standards, face high costs, unsuitable matches, and, increasingly, outright exclusion. Industrial capitalism has its industrial reserve army and its *lumpenproletariat*. Digital capitalism has its stubborn off-the-grid dwellers, cash economy, and, as it were, its *lumpenscoretariat*.[45] Excluding people who are deemed "too" untruthful, risky, deviant, or demanding—however these traits are defined and evaluated, usually in relation to some specific value outcome—is just a normal part of business. But algorithms allow these efforts to be carried out at scale, and to stick over time. A score that circulates, is replicated, and becomes consolidated into other indexes is far more powerful than a reputation that is confined to a file cabinet, a reference book, or an agent's memory. In Chapters 4 and 5 we turn to the material shape and economic implications of this widespread institutionalization of ordinal reason.