March 2017. Forthcoming, *Historical and Social Research*, special issue on Market Classifications.

Categories All the Way Down

Marion Fourcade
University of California, Berkeley
fourcade@berkeley.edu

Kieran Healy
Duke University
kjhealy@soc.duke.edu

Abstract: Scores and classifications are dual to one another. Cardinal and ordinal measures are repeatedly used to produce nominal classifications of essential worth. Conversely, presumptively natural kinds provide the basis for new measurement and scoring systems. Over time, the iterative application of nominal classifications and quantifying measures produce involuted, nested systems whose structure and origins are hard to disentangle. While careful studies of earlier systems and methods has often uncovered these arbitrary aspects, newer technical tools for classification are at once substantially more opaque than their predecessors and more likely to be employed on very large scales. The classification situations to which they give rise thus have the potential to produce the sort of naturalized facticity characteristic of classical social facts.

The articles in this issue explore scoring and classification tools across a range of economic settings, and from a variety of perspectives. The settings range from the German wine market to the American subprime credit sector, from the sustainability and social investment sector to the British fashion world (Nagel et al, this issue; Schiller-Merkens, this issue). The perspectives taken variously see scoring and classification methods as tools for solving coordination or action problems in markets, as means for establishing and maintaining identities (Pridmore and Hämäläinen, this issue) and as portable judgment devices with the capacity to to be put to use beyond their original context (Chiapello and Godefroy, this issue). Across the contributions is the sense that, as Citron and Pasquale (2014) have suggested, we now live in "scored societies" where increasingly large tracts of social life are subject to these methods, and in an increasingly automated manner. Discussing the credit crisis of 2007-08, MacKenzie (2011: 1830) asks "Should we understand the conduct of those practices and the use of

their results as having been driven by belief in them, or should it be seen as cynical, as driven simply by the pursuit of gain (e.g., by earning fees from ratings)?" As devices or tools for action and judgment, scoring and classification methods seem both in and out of the hands of their users, instrumental but disciplining, indispensable yet opaque. In this short paper, we ask: just how opaque?

THE DUALITY OF SCORES AND CLASSIFICATIONS

Scores and classifications are dual to one another, in two senses. First, tools for scoring and ranking—for measuring and comparing on a cardinal or ordinal scale—are repeatedly used to produce nominal classifications associated with judgements of essential worth (Fourcade 2016). Continuous measures are cut into ranked scales, which in turn become socially acted on as classes or categories or person, organization, or group. Second, over time the nominal classes and categories we interpret as basic to social life provide the starting point for new efforts to measure, score, and rank again. Prior classifications provide the basis for new measurements and scores, and scoring systems give rise to newly classified kinds.

Modern institutions, both public and private, use tools and procedures that track individuals, assess their behavior, and assign them membership in various categories. They use them, variously, in their efforts to monitor conduct, calculate risk, or extract value. Moved by the seemingly infinite possibilities offered by digital technologies, contemporary market organizations relentlessly segment and score large quantities of behavioral data. Seeing and knowing people by way of these tools changes how markets and states work. Their sorting and slotting procedures shape the availability and price of many goods and services, not only in traditional commodities markets but also in health care, insurance, education, legal services, and housing. Beyond these conventionally institutional arenas, we also increasingly find them reformatting the structure of ordinary sociability, from opportunities for friendship and dating to getting around town at the weekend. As we have argued elsewhere, the partially achieved, partially assigned categories that result from this widespread expansion of algorithmic decision-making can be thought of as *classification situations*. They shape the possibilities offered to individuals differentiated by them—in Weberian terms, these systems structure their life chances (Fourcade and Healy 2013).

Across the range of markets and settings they organize, scores and the categories generated by them are market-derived and market-oriented tools. They identify important or valuable individuals, where the criteria for "value" is determined by criteria internal to the particular market in question. To the extent that these criteria for evaluation are shared across market settings, and perhaps more importantly to the extent that data, methods, and tools for evaluation are also shared in this way (Rona-Tas,

this volume), classification situations may cohere in a systematic and increasingly consequential manner.

From the point of view of individuals, meanwhile, classification situations have objective consequences that can be measured in prices lowered or raised, fees incurred or waived, and opportunities proffered or lost. They also have a phenomenological aspect (Fourcade and Healy 2016). Because these new technologies of social classification are personal, pervasive, and moralized, the experience of being "well-situated" by them is often a pleasing one. We feel that the market or service (Amazon, Netflix, etc…) "gets us."

Occasionally, though, this expectation is betrayed. Personal cost, inconvenience, and awkwardness typically accompany a poor match. For the conventionally well-situated, a bad match occurs when the wrong product is pitched, or pitched at the wrong price. In these cases it is increasingly common for people to be consciously annoyed at the choices the algorithm has made for them. (How can Amazon be so stupid as to recommend this to me, given how much they know about my purchasing?) More interesting are cases where the quality of the match is "good" from the market's point of view but potentially "bad" from the point of view of the customer's sense of their own experience or identity. For example, the value of an individual in the subprime credit market may come from them having a "bad" credit score and thus ending up in the "wrong" category. They would prefer to be classified elsewhere, but the potentially stigmatized classification is all that is available. In these cases, the phenomenology of one's classification situation may involve both firms and customers may seeking some destigmatized understanding of the exploitative or predatory arrangements they are about to enter in to. Subprime customers are encouraged to feel (and often do feel) that the expensive credit product is "right for them", or presented by a firm that "understands their needs". Increasingly, the same is true of the experience of those who sign up to for-profit schools and colleges (Cottom 2017), or poor-quality health plans. Deceptive sales pitches for bad products are as old as the market itself, but they find new expression through the machinery of category matching and tailored pricing.

CATEGORIES ALL THE WAY DOWN

The study of classification is nothing new in the social sciences either. Understanding the social foundations of the construction of the categories through which people apprehend the world around them, and struggles over this process (Bourdieu 1984), is the central problem of the sociology of knowledge. Scholars since Durkheim have singled out this question as a necessary precondition to any properly sociological or anthropological inquiry. As Warren Schmaus puts it, "social life as we know it, [Durkheim] thought, would not be possible if people did not share certain conceptions

of time, space, causality, and classification" (Schmaus, 2004, p4). Shared categories, vocabularies and nomenclatures express, enable and sustain social coordination. They also align and mobilize—in other words, they are political in essence. The "economics of convention" approach as it developed in France in the 1980s and 1990s emphasized this point, seeing categories and in particular statistical nomenclatures as devices that *constitute* communities. In a world marked by both uncertainty, different types of conventions organize the pragmatics and formatting of action: people appraise and classify the persons and things around them, and they do so in reference to emergent sets of common expectations, "grammars of worth," and evaluative conventions (Boltanski and Thévenot 2006, Lamont 2012).

It is in the work of Laurent Thévenot and Alain Desrosières that the tight connection between classification and quantification is most explicitly articulated. (See Diaz-Bone, this issue for a summary; Desrosières 1995, Thévenot 2016, Diaz-Bone and Didier 2016.)[1] Quantifying implies sorting, and to sort is to pass through a categorical lens. There is no measurement that does not go through the lens of a classifier. As we, in this issue, ponder over the classifying consequences of market scoring processes, we must remember that these new classification situations, produced by measurement and quantification efforts, are themselves built on top of other classifying practices and the schemes yielded by them. The classifying (a score, a ranking, a rating) is itself a classified product.[2] The composite devices that are our main focus here, such as credit scores, depend in the first instance on choices about the way credit events are defined and measured. A small change in the measurement system, or a reweighting of the precise mix of factors deemed relevant to an assessment, may have dramatic effects on the outcomes.

Quantification not only implies classification, it implies classifications on top of other classifications—indeed a classificatory architecture that pulls in variegated ways of boxing and measuring people and things to some end. The pristine numerical output of a final score may bear a tangled relationship to its underlying strata of classes, groups, and types. In this sense, scores are categories all the way down. Any scoring system is dependent on the categorical work of third parties. To the extent that scores are interdependent in practice, systems will be vulnerable to fads and shifts in data collection, measurement, and organization that happen elsewhere. For instance, a change in a bank's approach to credit limits will automatically reverberate into the credit score of its customers, since the ratio between balance transfer and credit limit is a common component of the latter. A lower limit will worsen the score, while a higher limit will improve it, even in the absence of any active intervention by the users.

[1] Characteristically, the French economics of conventions was born at the French statistical institute.

[2] Both meanings of classified, i.e. categorized and secret, are often pertinent when we discuss scoring methods.

Second, if scores are categories all the way down, then they offer infinite possibilities for combination. No wonder, then, that the design of a new type of score leads immediately to the emergence of a field of competitors, all vying to establish dominance over a particular type of measurement, or at least over a niche market within it. If your magazine wants a piece of the college rankings business, better find some unique way of slicing the data. Indeed the most powerful scores in the economy are those that combine stable market anchoring roles with flexibility in implementation. Credit scores have that quality. They generally anchors the lending business (few would lend without a credit check) but the companies that produce them have also made the device customizable to predict, for instance, the likelihood that an applicant will be tempted by and pay as agreed on a *particular* type of loan. As Sevignani aptly reminds us in his own contribution (this issue), asymmetric relations between owners and users of the means of information, surveillance and communication are a source of exploitation in the classic Marxist sense, where powerful companies are able to appropriate the wealth created by users as they navigate digital systems. Importantly, the systems themselves also facilitate a derivative form of exploitation where the data thus obtained is repurposed and manipulated to facilitate the extraction of profit. In Donald MacKenzie's (2006) phrase, borrowed from Milton Friedman, digital technologies are not simply cameras that provide an objective picture of the customer's creditworthiness or reputation. They have become the engine of the value extraction machine: the wide-ranging knowledge on users enables a fine-tuning of the products on offer to broader aspects of the person, from the ability to detect someone's reservation price to identifying their propensity to be fooled.

Third, if scores are composed of categories, then understanding how the resulting sausage, so to speak, is made, is big business. Wendy Espeland and Michael Sauder's recent study picks apart the structure and effects of the dominant *US News and World Report* ranking of Law Schools in the United States (Espeland and Sauder, 2016). The system is of interest for several reasons. First, the ranking is not "official" in the sense of being sponsored by the state, or even by a professional association of lawyers or legal academics. Nevertheless, it is the chief means by which aspiring law students and Law School Deans alike orient themselves to the public status order of their discipline. Second, the ranking is calculated from a mixture of sources, ranging from the average standardized test scores and undergraduate grade point averages of admitted students, to measures of faculty and student expenditure. It also includes a reputational component extracted from a survey of Law School Deans and placement directors, legal professionals, and judges. Some of these sources are themselves highly refined individual-level instruments being used in the "off-label" manner Rona-Tas describes. Others are organizational features of the schools that are somewhat under the control of the staff. Still others are measures of the existence of the very status order that the ranking will quantify and express. Schools seek to manipulate their place in the

pecking order by focusing their action on these various components of the ranking, that is, on the classifications that are *baked* into the *US News and World Report* performance measure.[3] But this work requires a delicate—and somewhat unstable—balancing act, since some components have inherently contradictory dynamics. For instance, given the existing applicant pool and the institutionalized measures that are available, it may be impossible to simultaneously increase measured diversity and test scores. When faced with dilemmas of this sort, very high-status actors may occasionally move unilaterally to rebalance the regime, ignoring or shifting their criteria while banking that their old-fashioned unquantified public status will carry them through.[4] But most actors in a status order do not have this move available to them. This is a rejoinder to Karoline Krenn's point in her article (this issue) that powerful or wealthy actors have in effect more freedom *vis-à-vis* objective measurement systems than less privileged ones.

Law schools and similar professional rankings are opaque and transparent at the same time. They are internally opaque, in that they incorporate a heterogeneous body of measures and weigh them in a way that, if not entirely arbitrary, is at least open to question. Yet they are transparent in the sense that it remains possible to see the various ingredients. Indeed, one of the central puzzles of the rise of third-party rating and ranking systems in this area is why they have been so successful. The hold they exercise over the minds of applicants and the disciplining effects they have on decision-makers at professional schools seem out of all proportion to both the authority of the entity doing the ranking (a news magazine relatively few people read) and the quality of the methods used to generate the results. Moreover, the feedback built into the measures seems to ensure the reproduction of the existing status order in an obvious way. And yet even so poor a measure of status as this has successfully acquired the mantle of an unavoidable, objective social fact about legal education in the United States. The constraint is deeply felt: in Espeland and Sauder's phrase, the rankings act as "engines of anxiety" for applicants and administrators alike. This anxiety, in turn, fuels a prosperous consulting industry specializing in the management or gaming of ratings. The process is very similar, indeed, to the search engine optimization industry that developed around Google's algorithm, PageRank (Ziewitz 2015). Sometimes the rankers even provide these governance services themselves, as in the case of the *Times*

[3] Think, for instance, of the verbal, quantitative and (now defunct) essay components of the SAT.

[4] See, for example, Harvard Law School's recent decision to accept Graduate Record Examination (GRE) scores from applicants instead of the traditionally required Law School Admissions Test (LSAT) score. The decision was taken on the grounds of diversity, of a general kind. A spokesperson said the school was seeking to "diversify our community in terms of academic background, country of origin, and financial circumstances". Note that, even in a case like this, Harvard's decision is not to abandon its use of a standardized test but to take advantage of a somewhat different test instrument for moderately off-label use. See Elizabeth Olson, "Harvard Law School, Moving to Expand Applicant Pool, Will Accept GRE", *New York Times*, March 9th 2017, p. B5 *https://www.nytimes.com/2017/03/08/business/dealbook/harvard-law-will-accept-gre-scores.html*. Accessed March 9th, 2017.

Higher Education World University Rankings, which is marketing "strategic solutions" for universities to "improve through performance analysis and benchmarking."[5] Indeed the opportunity to sell a suite of associated services may be the prime motivation for investing in the development of a new ranking or scoring method in the first place. The production of classification situations is a valuation practice (Krüger and Reinhart, this issue) that has both evaluative and valorizing, or economic, aspirations (Vatin 2013).

WHITHER THE CATEGORIES?

The relationship between scores and rankings on the one hand, and the categories they rely upon on the other, raises a fundamental problem in the sociology of knowledge. Donald MacKenzie (2011) reminds us that it was in part the financial actors' belief in the facticity of their new composite products, the ABS CDOs, or tranches of tranches of bundles of mortgages, that blinded them to the dangers within. In their efforts to redistribute risk through securitization, people lost sight of both the declining quality of the components (the category 'all the way down', the individual mortgage) and the possibility of even a modest correlation among those ABSs, which the 2008 credit crisis ultimately revealed. As MacKenzie notes, the market participants overlooked these risks partly because they believed them to be good tools, and partly because it was in their financial interest to act as though they were good. There was a lot of money at stake.

As the skills required to understand the internal structure of algorithms become more demanding, ranking and scoring devices become less easily accountable. Furthermore, the inner workings of the vast majority of scores, rankings and algorithms currently in use are *deliberately* shrouded in secrecy. The opacity of instruments in the name of state or trade secrets lies beneath the "black box society" criticized by Frank Pasquale (2015) and Catherine O'Neill (2016). But as Jenna Burrell (2016) has argued, these two modalities of opacity (proprietary codes and technical know-how) have now been superseded by another, more intractable form. Machine learning procedures have been developed in cases where an explicit logic of decision-making remains elusive, or simply where the abundance of data makes such an approach more efficient. In contrast with traditional artificial intelligence, where computers were programmed to follow an algorithm designed by a coder, the machine-learning approach uses statistics to identify patterns directly in the data. The computer "learns" from these patterns, and optimizes its performance of a task accordingly. Computers can also learn to classify data on their own, and thereby predict where new data should fit. In other words they can produce a model, but the difference with human programming is that the rationale for why certain decision rules end up in the model is not always obvious. In the most

[5] *http://timeshighereducationonline.com/clienthub/strategic-solutions.html* Accessed March 8, 2017.

advanced techniques, this rationale is in fact impenetrable for the human mind. Owing to the recent resurgence of 'deep' learning procedures, the model's outputs are now based on multiple, sometimes thousands of processing layers. Each layer produces its own representation of some piece of the data and relays what it has learned to the next layer, all the way to the final layer, which uses all the information passed along the way to generate the classification.

Methods for layered neural networks have been developed since the 1960s, and they began to be seen in more widespread use in applied statistics the 1980s and 1990s. At the time they were seen mostly as "a flexible non-linear extension of multiple logistic regression" (Venables and Ripley 2002, p.342). Their usefulness seemed relatively limited at the time. In comparison to more familiar methods they were both less transparent and more computational trouble. However, continuing research, the rapid expansion of cheap, large-scale computing power, and the concomitant availability of enormous datasets for analysis brought resulted in a step-change in the usefulness of these methods. Their application began to yield rapid progress in notoriously intractable problems such as speech recognition, image classification, and natural language processing. The result has been a huge surge of interest in these approaches, and a new wave of experimentation with them in many different areas.

A characteristic feature of discussion around deep learning is that while its success is results-driven, a satisfactory theory of why these methods work so well is harder to provide. Research and applications continue to surge, but it is striking to see the enthusiasm for these methods intermingled with the frank acknowledgment, even by experts, of how opaque they are in practice. It is common enough for well-understood technical methods to be deployed as packaged tools for use by non-expert (but often still "professional") practitioners. But deep learning techniques have much more of this quality than usual. Due to the high-dimensional character of the data and the model, the way these procedures operate, calculate, and classify is typically impervious to human interpretation. It is often impossible in practice to identify the role of individual inputs, which makes the devices rather intractable to manage when problems arise. That was Google's hard-learned lesson after its image recognition software classified black people as gorillas, and the only workable solution (since the classifier could not be unpicked to fix this error alone) involved preventing *any* photo from being tagged to the word gorilla. Categories all the way down, but what were the categories in the end?

As the tools of deep learning are just beginning to be applied across market settings –for instance in credit scoring–, the issue of opacity is returning to the forefront with a vengeance. The law requires that scoring tools be interpretable or comprehensible to scorer and scored alike, but the new methods are much harder to make sense of than the old, both in a technical way and in a regulatory one (Kroll et al., forthcoming). At the same time, they are also more powerful, and better able to generate the kind of

outcomes that mortgage and credit issuers want (e.g., better predictions of risk). Once again, we see the prospect of enigmatic methods that are at once technically effective, rhetorically useful, and financially rewarding, often combined with a certain kind of blind confidence that nothing will go terribly wrong, as in the credit crisis case.

Traditional mechanisms of social classification are powerful. Legal or political classifications of an arbitrary sort can become imbued with the character of a taken-for-granted facts. Amateurish or barely defensible data collection and ranking schemes turn out to have the capacity to control the status order of professional fields, partly just in virtue of their quantitative character. Perhaps a deeply arbitrary order is better than no order. Perhaps, as Gillespie (2014, p. 192) points out, "we want relief from the duty of being skeptical about information we can never assure for certain." The new classifiers seem to combine and supercharge these features. They are more technically sophisticated than many of the methods that preceded them, and are also set to be applied on a much larger scale. At the same time, they are far more difficult to fathom—perhaps intrinsically so—even for well-informed users. To exaggerate, but only a little, they fuse the rational legitimacy of technical analysis with the enigmatic but undeniable force of a Delphic oracle. The classification situations to which these methods give rise thus have the potential to produce the sort of naturalized facticity characteristic of truly social facts. Both the act of classification and the criteria for it fade into the background, and we are left with what seems simply to be the world itself, delivered to us as a set of natural categories that it is in our best interest to believe in, act upon, or live up to.

REFERENCES

Bourdieu, Pierre. 1984. *Distinction. A Social Critique of the Judgement of Taste*. Harvard University Press.

Burrell, Jenna. 2016. "How the machine 'thinks': Understanding opacity in machine learning algorithms." *Big Data and Society* 3(1): 1-12.

Chiapello, Eve, and Gaetan Godefroy. 2017. "The dual function of judgment devices. Why does the plurality of market classifications matter?" This issue.

Citron, Danielle Keats and Pasquale, Frank A. 2014. "The Scored Society: Due Process for Automated Predictions." *Washington Law Review* 89: 1-33.

Cottom, Tressie McMillan. 2017. *Lower Ed: The Troubling Rise of For-Profit Colleges in the New Economy*. New York: New Press.

Desrosières, Alain. 1995. "Classer et mesurer: les deux faces de l'argument statistique." *Réseaux* 13: 11-29.

Diaz-Bone, Rainer. 2017. "Market Classifications, quantifications and quality conventions in markets—Perspectives of the economics of convention." This issue.

Diaz-Bone, Rainer, and Emmanuel Didier (eds.) 2016. "Conventions and Quantification –Transdisciplinary Perspectives on Statistics and Classifications." *Historical Social Research* 41(2). (Special issue)

Espeland, Wendy, and Michael Sauder. 2016. *Engines of Anxiety. Academic Rankings, Reputation and Accountability*. Russell Sage Foundation.

Fourcade, Marion. 2016. "Ordinalization." *Sociological Theory* 34: 175-195.

Fourcade, Marion, and Kieran Healy. 2013. "Classification Situations: Life Chances in the Neoliberal Economy." *Accounting, Organizations and Society* 38: 559-572.

Fourcade, Marion and Kieran Healy. 2016. "Seeing Like a Market." *Socioeconomic Review.*

Gillespie, Tarleton. 2014. "The Relevance of Algorithms." Pp167-194 In Tarleton Gillespie, Pablo Boczkowski, and Kirsten Foot (eds.) *Media Technologies. Essays on Communication, Materiality, and Society (Inside Technology)*. Cambridge, MA: MIT Press.

Krenn, Karoline. 2017. "Segmented intermediation. Advice concepts in German financial services." This issue.

Kroll, Joshua A., Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson, and Harlan Yu. Forthcoming. "Accountable Algorithms." *University of Pennsylvania Law Review*.

Krüger, Anne and Martin Reinhart. 2017. "Theories of Valuation—Building Blocks for Conceptualizing Valuation Between Practice and Structure." This issue.

Lamont, Michèle. 2017. "Toward a Comparative Sociology of Valuation and Evaluation." *Annual Review of Sociology* 38:201-221.

MacKenzie, Donald. 2006. *An Engine, Not a Camera. How Financial Models Shape Markets*. MIT Press.

MacKenzie, Donald. 2011. "The Credit Crisis as a Problem in the Sociology of Knowledge." *American Journal of Sociology* 116: 1778-1841.

Nagel, Sebastien, Stefanie Hiss, Daniela Woschnack, and Bernd Teufel. 2017. "Between Efficiency and Resilience: The Classification of Companies According to their Sustainability Performance." This issue.

O'Neill, Catherine. 2016. *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy*. Crown.

Pasquale, Frank. 2015. *The Black Box Society*. Harvard University Press.

Pridmore, Jason and Lalu Hämäläinen. 2017. "Segmentation in (in)action: Marketing and 'yet to be installed' role of big and social media data." This issue.

Rona-Tas, Akos. 2017. "The Off-Label Use of Consumer Credit Ratings." This issue.

Schmaus, Warren. 2004. *Rethinking Durkheim and His Tradition*. Cambridge University Press.

Schiller-Merkens, Simone. 2017. "Will green remain the new black? Dynamics in the self-categorization of ethical fashion designers." This issue.

Sevignani, Sebastian. 2017. "Surveillance, classification, and social inequality in informational capitalism: The relevance of exploitation in the context of markets in information." This issue.

Thévenot, Laurent. 2016. "From Social Coding to Economics of Convention: A Thirty-Year Perspective on the Analysis of Qualification and Quantification Investments." *Historical Social Research / Historische Sozialforschung* 41(2): 96-117.

Vatin, François. 2013. "Valuation as Evaluating and Valorizing." *Valuation Studies* 1(1): 31-50.

Venables, W.N. and B.D. Ripley. 2002. *Modern Applied Statistics with S*. Fourth Edition. New York: Springer.

Ziewitz, Malte. 2015. "Spectacular Algorithms." Presentation at the Center for Science, Technology and Society, UC Berkeley, November 5.

Ziewitz, Malte. "Governing algorithms: Myth, mess and methods." *Science, Technology, and Human Values* 41(1): 3-16.