Received: 29 July 2020 | Accepted: 2 February 2021

DOI: 10.1002/aaai.12034

ARTICLE

# Decoding human behavior with big data? Critical, constructive input from the decision sciences

**Konstantinos V. Katsikopoulos**[1] | **Marc C. Canellas**[2]

[1]Department of Decision Analytics and Risk, University, of Southampton Business School, Highfield Campus, Southampton, UK

[2]New York University School of Law, New York, USA

**Correspondence**
Konstantinos V. Katsikopoulos, Department of Decision Analytics and Risk, University of Southampton Business School, Highfield Campus, Building 2, Southampton SO17 1BJ, UK.
Email: k.katsikopoulos@soton.ac.uk

**Abstract**

Big data analytics employs algorithms to uncover people's preferences and values, and support their decision making. A central assumption of big data analytics is that it can explain and predict human behavior. We investigate this assumption, aiming to enhance the knowledge basis for developing algorithmic standards in big data analytics. First, we argue that big data analytics is by design atheoretical and does not provide process-based explanations of human behavior; thus, it is unfit to support deliberation that is transparent and explainable. Second, we review evidence from interdisciplinary decision science, showing that the accuracy of complex algorithms used in big data analytics for predicting human behavior is not consistently higher than that of simple rules of thumb. Rather, it is lower in situations such as predicting election outcomes, criminal profiling, and granting bail. Big data algorithms can be considered as candidate models for explaining, predicting, and supporting human decision making when they match, in transparency and accuracy, simple, process-based, domain-grounded theories of human behavior. Big data analytics can be inspired by behavioral and cognitive theory.

## INTRODUCTION

Who has not heard the motto that data analytics, machine learning, Facebook, or Google…"know us better than we know ourselves"? Even those alarmed by the prospect of silicon superintelligence (Harari 2016; Zuboff 2019) do not doubt its prowess. In the "18 Miles Outside of Roanoke" episode of the TV show *For the People*, a judge bypasses the prediction of a recidivism algorithm but warns the defense attorney that such algorithms are the future of law, drawing an analogy to AI now beating human chess champions (Attie and Verica 2018). This type of argument for the imminent coming of superintelligence based on the success of AI in games is a common one, as pointed out, among others, by Gigerenzer (in press). The problem is that such arguments show a lack of appreciation for the serious challenge that social situations—as when people interact with each other or institutions—that are less stable and well defined than games such as chess and Go pose for making accurate predictions (Wu 2019; Makridakis, Hyndman, and Petropoulos 2020).

We appreciate that AI researchers strive to avoid hyperbole. The phrase *big data analytics* can be a marketing device in parts of academia and industry. It is a blanket term for applications of models from statistics and machine learning to datasets with "volume," "velocity," and "variety" (McAfee et al. 2012). As in the famous quote of Peter Norvig, director of research at Google: "We don't have better algorithms, we just have more data." In this sense, the usual comments for the possible risks and

benefits of AI apply to big data analytics as well. Still, because big data analytics is pronounced to be no less than a revolution in the name of AI, ready to sweep science as well as business (Anderson 2008; McAfee et al. 2012), further investigations of the risks and benefits of big data are necessary.

The risks of big data analytics are very real (Clarke 2016). Biases have been identified in the data and algorithms guiding law enforcement decisions (Richardson, Schultz, and Crawford 2019), medical decisions (Obermeyer et al. 2019; Benjamin 2019), and in language corpora (Caliskan, Bryson, and Narayanan 2017). Big data analytics systems are also vulnerable to adversarial attacks (for example, Brundage et al. 2018; Nguyen, Yosinski, and Clune 2015). Misapplications of big data analytics have resulted in lawsuits and claims of improper government support of those with developmental and intellectual disabilities (Stanley 2017), fraudulent unemployment insurance fraud investigations (Garza 2020), discriminatory car insurance rates (Varner 2020), and incorrect background housing checks (Kirchner and Goldstein 2020).

These well-publicized risks notwithstanding, perhaps one should be less certain of the benefits of big data analytics. For instance, Frederik and Martijn (2019) provide an overview of the capacity of big data analytics for effective personalized online advertising. Consider eBay and a host of other companies, which pay Google to display their ads at the top of a screen for users who are deemed to need just a little nudge to buy. What is the evidence for the effectiveness of such personalized advertising? Surely, it is naïve to even ask because companies as eBay have all the analytics support to make sure they are not burning money ($20 million in some years), right? The profit from these online ads was estimated by eBay to be about $240 million (Frederik and Martijn 2019). But such estimations are fraught with methodological problems. For example, how does one know that those targeted by an online ad would not have purchased the product anyway? Experimental control is required to tease such confounds apart. Blake, Nosko, and Tadelis (2015) ran a series of critical large-scale field experiments. They found that brand keyword ads had no measurable short-term benefits. And, while new and infrequent users could be profitably influenced by ads, this effect was offset for more frequent users, resulting to negative average returns.

Motivated by such findings, this article takes a step back and reviews the purported benefits of big data. We critically examine the capacity of big data analytics to explain and predict individual and group *human behavior*, and thus to support human decision making. Note that we do not question the success of complex algorithms such as deep learning in engineering problems such as processing images, video, speech, and audio (LeCun, Bengio, and Hinton 2015). Our contribution takes a scientific perspective, sampling arguments, and evidence from decision research across multiple disciplines that might be little known to academics and practitioners in AI who do not work on its intersection with fields such as business, politics, and law. The article aims at enhancing the knowledge basis for the development of algorithmic standards in big data analytics. We propose that *simple rules of thumb* that use few pieces of information and combine them in computationally simple ways (Gigerenzer and Todd, 1999; Katsikopoulos et al., 2020) can serve as benchmarks for big data algorithms. We do not provide a comprehensive review but open doors and provide pointers to the literature in the decision sciences broadly construed.

The article is organized as follows. In the next section, we argue that big data analytics is by design atheoretical and does not provide process-based explanations of human behavior; making it unfit to support deliberation that is transparent and explainable to experts and laypeople. In the section after that, we review evidence showing that the accuracy of complex statistical and machine learning algorithms used in big data analytics in predicting human behavior is not consistently higher than that of simple rules of thumb; rather, it has been found to be lower in situations such as predicting election results, criminal profiling, and granting bail. Finally, in a last section, we synthesize these points and conclude that simple, process-based, domain-grounded theories of human decision making should be put forth as benchmarks, which big data algorithms, if they are to be considered as candidate models for explaining, predicting, and supporting human decision making, should match in terms of both transparency and accuracy. Taking a constructive point of view, we join others (Griffiths 2015; Analytis, Barkoczi, and Herzog 2018) in suggesting that big data analytics can be inspired by behavioral and cognitive theory.

## BIG DATA ANALYTICS: LACK OF THEORY, EXPLANATIONS, AND TRANSPARENCY

### Theory and explanations

A foundational piece of big data analytics is Chris Anderson's "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." This title is striking. One might say that it provokes in order to attract attention, and is not to be taken literally. Perhaps. Let us look closer at what Anderson actually says:

> "Petabytes allow us to say: 'Correlation is enough'…We can analyze the data without hypotheses about what it might show. We can

> throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot."

And also

> "This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear. Out with every theory of human behavior, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves" (Anderson 2008).

So, the theory of big data analytics is that human behavior in any domain can be studied *without* a theory of this domain. Content and context do not matter. Rather measurement and statistics suffice to reveal correlations, which in turn supposedly suffice to make accurate predictions.

This idea can go too far. There are ample situations where statistical methods, even sophisticated ones, cannot generalize well if they do not have access to a causal domain theory (Matthews 2020). For example, statistics may identify features correlated with the outcome but which did not cause the outcome, such as when an image recognition algorithm distinguishes between dogs and wolves, not via the features of the animals themselves but by using the snow that is only present in the pictures of wolves because it is part of their natural habitat (Ribeiro, Singh, and Guestrin 2016). After its high accuracy is advertized, users may apply the algorithm confidently to all kinds of wolves on snowless ground and dogs in snow, only to see it fail. Without any domain theory of how wolves differ from dogs, even a rudimentary theory stating which of the two is more likely to live outdoors, users will be at a loss as to why this smart algorithm failed or how to fix it.

But what about practice? Do big data practitioners try to predict what people will do again without a behavioral theory? It seems so. Consider, for example, Nate Silver's work, and specifically how the result of the 2016 US presidential election was predicted (https://projects.fivethirtyeight.com/2016-election-forecast/).

Silver stated that the probability of winning was 71.4% for Hillary Clinton and 28.6% for Donald Trump. He also broke down this prediction by state and gave the probabilities of interesting scenarios, such as a landslide for each candidate. Silver's model took into account the accuracy of election polls all the way back to 1972, plus demographic and economic data. Numbers were adjusted, weighted, and averaged in many ways, including by using linear regressions, some of them regularized ones.

As far as we can tell from Silver's documentation, there is no political, economic, sociological, or psychological theory driving these calculations. By theory we mean a construction that goes beyond characteristics such as including voter income as a predictor in a regression. It is often argued that a statistical model, such as a linear regression, is equivalent to theorizing about the ways in which a voter's politics, economics, and psychology affect whom they vote for. At some formal level, of course, statistics can be seen as descriptions of behavior. But describing a human decision formally is not the same as providing an explanation for how it came about, what goal it is serving, and how it could be changed (Katsikopoulos 2011a; and references therein). This point, a key one in the decision sciences, has been made in legal research as well. In the words of Vincent Chiao:

> "Machine-learning techniques, neural networks in particular, raise a distinct set of concerns. Machine learning is 'atheoretical', in that a machine-learning algorithm 'learns' on its own to draw correlations between outcomes and inputs, including inputs that would not make much sense to a human. In the case of aeroplanes, bridges, and pharmaceuticals, even if lay persons do not understand how they work, still experts do. . . In contrast, in the case of a machine-learning algorithm, it may be the case that no one really understands the basis upon which it is drawing its correlations. Those correlations might be quite reliable, but it might be that no one is in a position to articulate quite why they are reliable and this surely does raise distinctive concerns about intelligibility." (Chiao 2019, p. 136).

Chiao connects the lack of theory in big data analytics to its lack of intelligibility. Big data analytics cannot help us understand why a person behaved a certain way. It cannot provide explanations of the reasons and the *process* by which a decision came about. Since the so-called cognitive revolution (Neisser 1967), describing the cognitive processes underlying observed behaviors is the pronounced goal of behavioral and decision research. A cognitive process specifies the temporal order in which information-processing events occur in the mind, and how those combine to produce a decision or another outcome.

For example, a voter might first consider the point that Trump is a populist (some readers might disagree with this judgment). This might lead the voter to move closer to voting for Clinton, so that that only one more strike against Trump will suffice to have her vote for Clinton. The voter might search her memory for further information on the two candidates. If she recalls that Trump has made racist remarks (again some readers might disagree with this), then she would decide to vote for Clinton. No such process-based explanations are provided by Silver's algorithms for the voting of a person or a group. It is hard to see how there could be such explanations in these algorithms, since there is no underlying theory about the politics, sociology, or psychology of voting.

Postelection, Silver acknowledged the consequences of his a-theoretic approach as he outlined a story of how President Trump won the election (Silver 2017). First, Silver's model ignored key contextual variables of the election. The model did not account for the context of Clinton trying to win a third consecutive term for her party, amidst a mediocre economy, at a time of high partisanship. Alternative models, which did focus on these types of factors, showed that the election was a toss-up or perhaps even slightly favored Trump. Second, Silver's model of voter preferences was not calibrated to the political climate. The model did not account for the instability of voter preferences and the additional uncertainty added by the large number of undecided and third-party voters. The model simply did not capture the psychological reality on the ground.

The workings of big data algorithms like Silver's model are not transparent and do not provide explainable theory. Because of this, and other reasons discussed below, these algorithms are unfit to support deliberation.

## Supporting transparent and explainable deliberation

Legal institutions as well as technical organizations recognize transparency and explainability as fundamental tenets for achieving trust for the tools they use. *Transparency* means making decision-making processes available for scrutiny, and *explainability* means being able to convey those processes to different stakeholders in a way in which they can consume it. These two tenets are necessary for trust because people, who are having decisions made about them, need to understand enough of the decision-making processes so that they feel they were treated fairly. If not, they will reject the decision itself, and by extension the legal institution and the technology.

The European *General Data Protection Regulation* provides rights to "meaningful information about the logic involved" in automated decisions, in other words, a right to explanation (Goodman and Flaxman 2017; Selbst and Powles 2017). In the United States, the Constitution requires procedural due process in government decision making incorporating principles of transparency, accuracy, and political accountability—principles that could be violated by opaque and inexplicable algorithms (Citron 2008). The *Institute for Electrical and Electronic Engineers* requires that the "basis of a particular [algorithmic] decision should always be discoverable" (IEEE 2019). The *Association for Computing Machinery* encourages those using algorithmic decision making "to produce explanations regarding both the procedures followed by the algorithm and the specific decisions that are made" (Garfinkel et al. 2017).

The need for transparent and explainable algorithms has been highlighted in the well-publicized case of *State of Wisconsin v. Loomis*. We discuss it from an angle based on legal theory. Eric Loomis denied involvement in a drive-by shooting but pleaded guilty to a couple of lesser charges, such as attempting to flee a traffic officer and operating a motor vehicle without the owner's consent (Harvard Law Review 2017). The trial court, having accepted the plea, ordered a report to inform sentencing, which included a risk assessment based on the COMPAS algorithm. Because the algorithm is proprietary, the defendant was not given the chance to inspect its logic and challenge it in court. The COMPAS algorithm suggested that the defendant had a high risk of recidivism and the court used this, along with other considerations, to sentence Loomis to 6 years of imprisonment and 5 years of extended supervision. The defendant made an appeal on due-process grounds, which the Wisconsin Supreme Court denied.

We do not discuss the technicalities of this case; for this, see the analysis of Brownsword and Harel (2019). Below we argue that this big data algorithm did not support transparent and explainable deliberation, and that the Court's lack of access to a process-based decision model prevented them from determining the appropriateness of using, or not using, the algorithm.

Even though the Court upheld the use of the COMPAS algorithm, it seemed uneasy about using a secret algorithm to help send a man to prison (Liptak 2017). The Court acknowledged the expert testimony at the trial court, warning about the risks of COMPAS, and cited a report by nonprofit organization *ProPublica* about COMPAS, which concluded that black defendants in Broward County, Florida, were far more likely than white defendants to be incorrectly judged as more likely to reoffend (Angwin et al. 2016). At the same time, the Court noted that *Northpointe*, which had marketed COMPAS, had disputed ProPublica's analysis (Holsinger et al. 2018).

The Court concluded that "if used properly," the COMPAS algorithm does not violate the right of due process at sentencing—yet nowhere did the Court define what "properly" means. A process-based model as a decision support aid would have enabled the Court to understand its own goals and processes, and to understand enough about the COMPAS algorithm to know how it can "properly" inform the Court. Without understanding their own or the COMPAS algorithm's processes, the Court could not meaningfully integrate the concerns about the algorithm into their decision. Ultimately, the Court asserted that the COMPAS algorithm report added valuable information, but that in any case Loomis would have gotten the same sentence based on other factors such as his criminal history and his attempting to flee. Nevertheless, the Court required a written disclaimer to be attached to any future COMPAS report, but again without any explanation of how it should or would affect future judicial decision making (Harvard Law Review 2017).

This interaction of the legal experts with the COMPAS algorithm was far from useful (Washington 2018). The argumentation and counterargumentation seem strange. If Loomis would have gotten the same sentence without the COMPAS report, then the information added by the report is beginning to appear to be less valuable than claimed, or at least less impactful. If so, why was it included? Especially, given that if used "improperly," it could violate Loomis' constitutional rights. Perhaps there is some unique insight that was provided by the algorithm? Doubtful. Nobody in the trial had the chance to benefit from knowing the logic of the algorithm since nobody—except Northpointe—had the chance to interact with the algorithm or have it explained to them. It is difficult enough to interact with nonprocess-based algorithms, but in this case, where the algorithm was also secret, it was impossible to interact with it. Laypeople following the trial must have been just as lost or more.

The lack of a process-based decision aid, paired with an opaque COMPAS algorithm, allowed the Court to adopt the big data algorithm expecting that it worked, without any evidence that it worked. The Court never explained whether or how the COMPAS algorithm actually improved judicial decision making, presumably by reducing recidivism and bias in sentencing. Instead the Court relied on statistical validation studies from other states or Northpointe itself to infer that it would reduce recidivism and bias.

It seems that no one knew then the systematic effect of risk assessment algorithms in the United States because the first empirical studies of judicial use of risk assessments were published later (Stevenson 2018; Stevenson and Doleac 2019). Stevenson and Doleac (2019) found that Virginia "judges' decisions are influenced by the risk score, leading to longer sentences for defendants with higher scores and shorter sentences for those with lower scores. However, [they found] no robust evidence that this reshuffling led to a decline in recidivism, and, over time, judges appeared to use the risk scores less." Given similar results in Kentucky, Megan Stevenson explained that these results challenge the belief that "actuarial tools outperform human intuition in predicting crime… While there are reasons to believe that the risk assessment tools provide new and useful information, the margin of gain is unclear" (Stevenson 2018).

In sum, it is difficult to see how algorithms such as COMPAS can support transparent and explainable deliberation. But perhaps such complex algorithms can reliably lead to accurate decision making?

## COMPLEX BIG DATA ALGORITHMS AND SIMPLE RULES: ACCURACY

### Overview

How accurate are big data algorithms in predicting human decision making? One way of answering is to compare their accuracy with simple rules. Simple rules represent a diametrically opposing philosophy to big data analytics.

The simple rules presented here, also called simple/fast-and-frugal/psychological *heuristics*, process a few pieces of information and do so in computationally simple ways (Katsikopoulos 2011b; Gigerenzer and Gaissmaier 2011; Şimşek 2013). Note that such heuristics are precise algorithms, not verbal descriptions as elsewhere in psychology. But in contrast to big data algorithms, in simple rules the information used is available through knowledge that people—laypeople or experts—often already possess or can easily access, and the computations performed are again within human reach, such as simply adding numbers or comparing them. For instance, recall the Clinton voter who decided based on just summing two binary pieces of information widely available, Trump's alleged populism and racism. Of course, these rules, as all models, cannot be perfectly accurate—for example, rules might have trouble with exceptions—even if they are derived from the behavior of experts.

At a first glance, such simple rules might appear to be dangerously close to "expert systems" (Jackson 1998), which were pursued and abandoned in applications of AI in previous decades. The two approaches differ, however, in three crucial aspects: First, instead of the emphasis of the expert-systems approach on building a large base of information, the simple-rules approach focuses on identifying a few key pieces of information and studies how those can be used to achieve superior performance. Second,

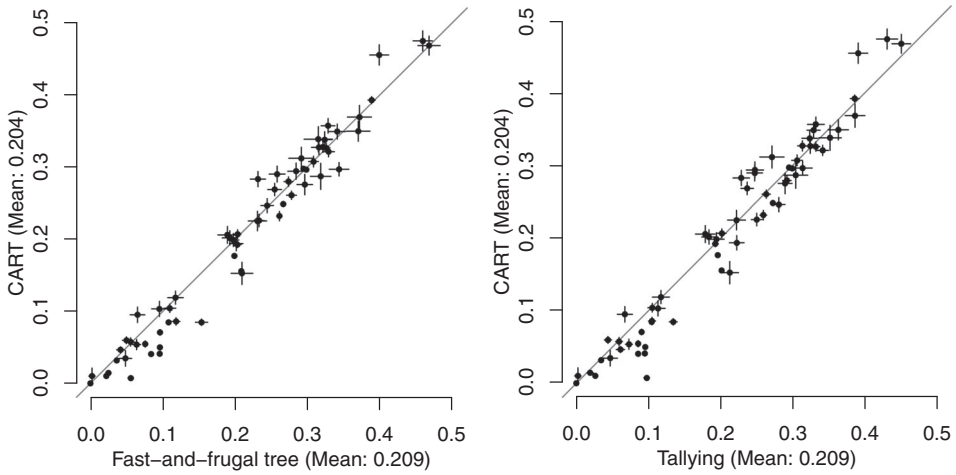


**FIGURE 1** Simple rules, such as fast-and-frugal trees and tallying, overall match the performance of more complex algorithms (CART) and sometimes outperform them (reprinted from Katsikopoulos et al. 2020, with permission of the MIT Press)

the proposed simple rules are intentionally kept simple and transparent, in order to increase user buy-in. Third, the performance of the simple rules (and any competing models) is evaluated based on machine-learning methodologies such as out-of-sample and out-of-population tests.

Simple rules have been discussed disciplines concerned with decision making, such as psychology (Gigerenzer and Todd 1999), economics (Rubinstein 1998; Katsikopoulos and Gigerenzer 2008), business (Sull and Eisenhardt 2015), management science (Hogarth and Karelaia 2005; Katsikopoulos, Durbach, and Stewart 2018), and, in fact, also machine learning (Holte 1993; Şimşek 2013). In law, Epstein (2009) has advocated simple rules, and the concept of the "reasonable man" (Gardner 2015) can be seen as a driver to simplicity in common-law methodology.

The comparative accuracy of simple and more complex algorithms has been the subject of multiple studies, each involving dozens of datasets across domains such as business, economics, law, medicine, politics, psychology, sociology, and transportation (Holte 1993; Czerlinski, Gigerenzer, and Goldstein 1999; Martignon, Katsikopoulos, and Woike 2008; Şimşek 2013; Lichtenberg and Şimşek 2017; Buckmann and Şimşek 2017). While these datasets are not necessarily "big," the statistical and machine learning algorithms used were those that are employed in big data analytics as well.

The results of such studies have been synthesized, and there is consensus on two points (Martignon and Hoffrage 2002; Gigerenzer and Gaissmaier 2011; Katsikopoulos 2011b; Katsikopoulos, Durbach, and Stewart 2018): (i) on the average, the predictive accuracy of simple rules and more complex algorithms is similar, and (ii) each kind of algorithms has regions of superior performance.

Figure 1 (from Katsikopoulos et al. 2020) provides an illustration of these two points. In this figure, the performance of Breiman et al.'s (Breiman et al. 1984) classification and regression trees (CART), a classic family of algorithms in statistics and machine learning that aim to be transparent, is compared to that of two families of simple rules, *fast-and-frugal trees* (left panel) and *tallying* (right panel). Fast-and-frugal trees (Martignon, Katsikopoulos, and Woike 2008) are a special case of decision trees that use a small number of questions and allow a classification after each question is asked; for an example see the granting-bail case below. Tallying (Hogarth and Karelaia 2005) is a special case of linear regression where the weights equal one; see the election-prediction case.

Katsikopoulos et al. (2020) compared the classification error of these three algorithms in 64 classification tasks, containing 95 to 32,561 instances (median 904) and three to 1418 cues (median 19). Each point in the figure shows the mean error of two algorithms, CART and a simple rule, in one task. The vertical line in the cross around a point shows two standard errors (on each side of the mean) of CART and the horizontal line shows the same for the simple rule. For a point on the diagonal, the error is the same for CART and the simple rule; if the point is above the diagonal, the simple rule made more accurate predictions; if the point is below the diagonal, CART made better predictions. Across the 64 tasks, each simple rule predicted nearly as well as CART, falling behind by only half a percentage point. There is an advantage for CART in problems where the error is small, that is, in easy tasks, and an advantage for simple rules when the error is larger, that is, in more difficult tasks.

In the remainder of this section, we will focus on particular cases where simple rules were compared with more complex algorithms. We do so in order to give detail on how simple rules work and show their transparency and explainability. Beyond these goals, our focus is on accuracy.

## Predicting election outcomes

Nate Silver's big data algorithms predicted a 71.4% chance of Clinton winning the 2016 election. Other polls and prediction markets made the same prediction about who had the clearly better chance of winning.

Historian Allan Lichtman, on the other hand, predicted that Trump would win. Lichtman (2016) developed a simple rule he derived based on his domain knowledge, blending theories of politics, economics, sociology, and psychology. Lichtman's *13 keys to the White House* rule does not deliver incredibly precise probabilities of winning but just a prediction of who will win. It is based on a historical analysis of voting behavior in US presidential elections from 1860 to 1980. The keys were fixed once and for all before the 1984 election. Each key is an issue that matters to US voters. Find below Lichtman's 13 keys, each stated so that it is either true or false in a particular election (note that many of them are the factors that Silver acknowledged were missing from his big data algorithm; Silver 2017).

*Key 1: Incumbent-party mandate*. Incumbent party holds more seats in the House of Representatives after this midterm election than the previous one.

*Key 2: Nomination contest*. No serious contest for incumbent-party nomination.

*Key 3: Incumbency*. Incumbent-party candidate is the sitting president.

*Key 4: Third party*. No significant third-party or independent campaign.

*Key 5: Short-term economy*. Economy not in recession during campaign.

*Key 6: Long-term economy*. Real annual per capita economic growth during the term equals or exceeds mean growth during two previous terms.

*Key 7: Policy change*. Incumbent administration effects major changes in national policy.

*Key 8: Social unrest*. No sustained social unrest during the term.

*Key 9: Scandal*. Incumbent administration untainted by major scandal.

*Key 10: Foreign or military failure*. Incumbent administration suffers no major failure in foreign or military affairs.

*Key 11: Foreign or military success*. Incumbent administration achieves a major success in foreign or military affairs.

*Key 12: Incumbent charisma*. Incumbent-party candidate is charismatic or national hero.

*Key 13: Challenger charisma*. The challenging-party candidate is not charismatic or national hero.

How to combine these keys to reach a decision? Lichtman proposed the following simple rule:

> *If six or more keys are false, the challenger will win.*

For example, consider the 2012 election, where Mitt Romney challenged Barack Obama. Lichtman counted all keys as true except 1, 6, and 12, and correctly predicted that Obama would win. Some of the keys, such as whether the candidate is the sitting president, require no judgment, while others, such as charisma, do.

In late September of 2016, Lichtman considered the keys to be settled and counted. Keys 1, 3, 4, 7, 11, and 12 turned against Clinton, the incumbent-party candidate. Thus, the prediction was that Trump will win. Now, there is one important caveat. According to Lichtman, the keys predict the majority vote, which Trump did not get. Thus, the 13-keys rule got the president right, but not the majority vote. No prediction rule is perfect, however, and the rule was closer to the outcome than big data algorithms. Additionally, its predictions have been accurate for all elections since 1984 when it was fixed.

The 13-key rule can be easily understood. The rule also reveals an intriguing logic that contradicts campaign wisdom: The keys all refer to the party holding the White House and their candidate, not to the challenger (with the exception of the challenger charisma key). The keys deal with the economy, foreign policy successes, social unrest, scandals, and policy innovation. If people fared well during the previous term, the incumbent candidate will win, otherwise lose. The 13-key rule delivers a simple theory, a process-based explanation for behavior, and creates a platform for discussion.

## Criminal profiling

The claim that crime-related predictions are more accurate when made by a computer algorithm than the human mind is not unique to the big data era. "Actuarial techniques" have been proposed for decades (Dawes, Faust, and Meehl 1989) and made available as software packages similar to COMPAS. In *geographical profiling*, given the geographical locations of a number of crimes and assuming that those were performed by the same offender, the goal is to identify the offender's residence. *CrimeStat* is a package by Levine and Associates (2000), which outputs the probability that any location in a prespecified 2D-grid is the residence of the serial offender. The description of CrimeStat by Snook, Taylor, and Bennell (2004) is fairly detailed and the underlying algorithm seems transparent

assuming familiarity with basic mathematics. Real-world data was used to calibrate the algorithm.

Brent Snook, Taylor, and Bennell (2004) tested the claim that CrimeStat would be more accurate than people. They recruited 215 prospective undergraduate university students and their guardians, and trained them to a simple-rules approach to geographical profiling, such as the *circle heuristic*, which states:

> *The majority of offenders' homes can be located within a circle with its diameter defined by the distance between the offender's two furthermost crimes.*

Note that, unlike the 13-keys rule, the circle heuristic does not lead to a unique answer, but is more of a guide to locating the offender's residence. The participants in the Snook et al. study had to solve 10 geographical profiling problems (from real serial murder cases), represented on 2D-maps produced by CrimeStat based on the location of three murders. One group of participants had to work on their own, another group was provided with the circle heuristic, and a third group was provided with another heuristic. Whereas the unaided group performed worse than the other two groups and CrimeStat, there was no statistically significant difference between the two groups supported with simple rules and CrimeStat's actuarial technique. In fact, the laypeople provided with the circle heuristic performed slightly better than the actuarial technique as measured by the mean map distance between the predicted offender's residence and her actual residence.

In the field, the benefits of actuarial techniques and their big data algorithm implementations for law enforcement are similarly unclear. In 2010, the Los Angeles Police Department (LAPD) developed *PredPol*, a predictive policing software that predicts where future unlawful activities will occur. After initially being adopted by departments across the country, numerous departments have stopped using the software because "it did not help reduce crime and essentially provided information already being gathered by officers patrolling the streets" (Puente 2019). LAPD's own internal audit concluded that there was "insufficient data to determine if the PredPol software helped reduce crime." Surprisingly, the CEO of PredPol rejected any claims that their big data algorithm should address the very issue it was supposedly being used for: "It's virtually impossible to pinpoint a decline or rise in crime to one thing. I'd be more surprised and suspicious if the inspector general found PredPol reduced crime" (Puentes 2019).

The criminal profiling case prescribes how decisions should be made, while the next case describes how courts of judges actually do make their decisions. Social, political, and legal theory and practice need insight and accuracy in both prescribing and describing human behavior.

## Granting bail

Assume now that a suspect serial offender has been identified. While awaiting trial in jail, she may apply to be granted bail (unconditional release). In the UK, such decisions are made by magistrates. How should they decide whether to grant bail or oppose it? The Bail Act of 1976 and its subsequent revisions say that magistrates should consider the nature and seriousness of the offense, the character, community ties, and bail record of the defendant, as well as the strength of the prosecution case, the likely sentence if convicted, and any other factor that appears to be relevant. The legal ideal of due process is based on a thorough analysis of the available information. However, the law is mute on how exactly magistrates should combine the various pieces of information. What do magistrates do?

Even though its status is weakening (Kahneman, Slovic, and Tversky 1982; Gigerenzer and Todd 1999), the ideal of "fully rational" standard economic rationality is still a dominant description of human behavior. It has not been claimed that the human brain implements the most sophisticated of big data algorithms (for example, support vector machines or random forests), but some other algorithms, such as linear regression, are routinely proposed. Dhami (2003) tested empirically if magistrate bail-or-jail decision making is better described by a linear model or a simple rule.

Dhami observed several hundred hearings in two London courts. The information available to the magistrates included the defendants' age, race, gender, strength of community ties, seriousness of offense, kind of offense, number of offenses, relation to the victim, plea (guilty, not guilty, no plea), previous convictions, bail record, the strength of the prosecution case, maximum penalty if convicted, circumstances of adjournment, length of adjournment, number of previous adjournments, prosecution request, defense request, previous court bail decisions, and police bail decision. The magistrates also saw whether the defendant was present at the bail hearing, whether or not they were legally represented, and by whom.

Dhami evaluated algorithm accuracy by first calibrating each algorithm on half of the whole dataset, and then by testing it in the other half, repeating the process multiple times to average out random variation. Across the two courts, the family of algorithms which weighted and added 25 features achieved 79% predictive accuracy, whereas the family of simple rules that only used three features

predicted 89% of the unseen data points. An example of a particular simple rule is the following:

> *Always oppose bail unless (1) prosecution granted bail and (2) neither police nor previous courts imposed conditions on bail.*

Note that this simple rule explicates a fast-and-frugal decision tree wherein three questions are asked in sequence (has prosecution granted bail? has police imposed bail conditions? have previous courts imposed bail conditions?) and if the answer to any of those questions is "no," then immediately bail is opposed without moving to the next question.

This rule suggests a gap between descriptions of how magistrates are deciding and the prescribed due process, which magistrates claim to be following (Dhami 2003). To understand this gap, one needs to think about the magistrates' situation. Their task is to do justice to each defendant and the public, by balancing the likelihood of the two possible errors: a *miss* occurs when a suspect is released on bail and subsequently commits another crime, threatens a witness, or does not come to court. A *false alarm* occurs when a suspect is imprisoned who would not have committed any of these offenses. Yet magistrates do not have the information to balance the two errors. The law does not give them any instructions. Even if there were such instructions, the English legal institutions do not collect statistics about the error rates of magistrates' decisions. And even if statistics were kept about how often misses occur, it would be impossible to do so for false alarms; no method can determine whether jailed individuals would have committed a crime had they been bailed.

In this situation, magistrates apparently focus on a task they are more capable of solving than making the correct decision: to protect themselves. Gigerenzer (2007) calls this *defensive decision making*. Magistrates can be proven wrong only if a released suspect fails to appear in court or commits a crime while on bail. To protect themselves against potential accusations by the media or the victims, magistrates follow the defensive logic embodied in the simple rule above. The rule is transparent enough to allow the magistrates to understand it and adjust it if necessary.

## SIMPLE RULES AS BENCHMARKS AND INSPIRATION FOR BIG DATA ANALYTICS

### Benchmarking big data analytics

This article aims to provide arguments and evidence that challenge sweeping assertions that big data analytics are *obviously* superior for explaining and predicting human behavior, and thus should be leading the development of algorithmic standards in areas such as business or law. Accepting such assertions uncritically is tantamount to committing a *big data hubris* (Lazer et al. 2014), where "small" data and simple rules are considered inherently inferior to big data processed by complex models. In contrast, Lazer et al. (2014) found that using a few variables publicly available on the website of the Centers for Disease Control (CDC) in simple linear models led to more accurate predictions of the prevalence of flu-related doctor-visits than the big data Google Flu Trends algorithm. Katsikopoulos et al. (in press) showed that an even simpler model, one that uses only the most recent observation on the CDC website, was more accurate than Google Flu Trends. One might note that the data on the CDC website is also big. But there is no big data algorithm used at CDC since the process is essentially just counting. Additionally, Google Flu Trends was notoriously opaque as the variables it used and the way it combined them were not revealed publicly. In general, big data analytics is by design atheoretical, which limits its support for processes we should hold dear such as providing transparent explanations of human behavior that we can understand and deliberate about.

Interestingly, decision research across multiple disciplines shows that theory-driven, simple rules of thumb can, under some conditions, be *both* more transparent and more accurate than complex algorithms. As Rudin and Radin (2019) said, it is not clear why we use black-box algorithms when we do not need to in order to be accurate. This is the case in the prediction of election outcomes, criminal profiling, granting bail, and in a host of other high-stake situations such as identifying threats in a security checkpoint while also trying to minimize civilian casualties, or in monitoring and regulating investment banks (Katsikopoulos et al. 2020). While a complete theory of the situations in which transparency does not need to be traded off with accuracy is still elusive, some conditions have emerged. For example, the *stable world* principle (Katsikopoulos et al. 2020; Gigerenzer in press) holds that tradeoffs are *only* necessary when the decision environment is stable, that is, does not change or changes in predictable ways that can be captured by well worked out formalisms such as probability. Social interactions, however, are not stable situations (Makridakis, Hyndman, and Petropoulos 2020). Such principles and conditions have also been formalized (Hogarth and Karelaia 2005; Katsikopoulos, 2011b; Lichtenberg and Şimşek 2019; Castle in press).

We should pause and ask: Is there sufficient evidence that the currently available theories of human behavior, which are understandable and reasonably accurate, have been now matched by "smart" big data algorithms in both transparency and accuracy? There might be evidence for

this in some cases, and there could soon be more. What we should not do is rush to judge which algorithms can form suitable models for supporting human decision making without checking the scientific evidence. *That* would not be smart.

## Inspiring big data analytics

Some AI and machine learning researchers are pursuing the development of accurate and at the same time transparent models. These approaches seem to employ less domain-grounded theory and more statistics. We find this work intriguing and promising. We conclude this article by commenting on these approaches and offering some constructive suggestions about how they could be married with the approach of simple, process-based, domain-grounded theory.

Bourgin et al. (2019; for a similar idea, see Trafton et al., 2020) start from a neural network and try to make it more accurate at predicting people's choices under risk through the following procedure: (i) identify a model from psychology that predicts the human data better than machine learning models (Erev et al. 2017); (ii) generate synthetic data from the psychological model and use it to train the neural network; (iii) fine-tune the neural network with natural human data, possibly also employing ensemble (that is, combinations of) models. Indeed, the authors were able to demonstrate that this procedure resulted in a model with best accuracy among the models tested. Bourgin et al. (2019) also suggest that the same procedure could be used to boost the neural network's understandability and explainability, although they do not show exactly how.

The simple-rules approach can enter in stages (i) and (ii) of Bourgin et al.'s approach, by using a simple cognitive rule for choice under risk such as the priority heuristic that predicts more accurately than behavioral models including prospect theory (Katsikopoulos and Gigerenzer 2008), and then testing the accuracy of the thus trained neural network. We are unsure, however, about how the approach of Bourgin et al. (2019) can improve transparency. To begin with, the Erev et al. (2017) model used does not specify the underlying cognitive processes, but only the resulting choices. Even if a process model such as the priority heuristic were used in Bourgin et al.'s procedure, it is not clear how this would boost the transparency of the proposed neural network. The use of ensembles might be an additional complication as Lessmann et al. (2015, p. 134) caution: "Using a large number of models, a significant minority of which give contradictory answers, is counterintuitive to many business leaders." The idea of Peysakhovich and Näcker (Peysakhovich and Näcker 2017), to examine the match of the predictions between domain and machine learning models in order to hypothesize mechanisms that the latter might be implicitly expressing, could help here.

There is also a more direct approach to making machine learning models more explainable. The idea is to construct a new model that approximates the predictions of the original one, and is easier to explain to stakeholders (Lundberg and Lee 2017; Molnar 2020). An article in the *AI Magazine* uses this approach in the context of life insurance—first predicting mortality risk using random survival forests and then explaining the predictions to the customer using a simpler model (Maier et al. 2020).

An issue with this approach is that it decouples explanation and prediction because each one of these functions is performed by a different model. This seems to be causing all sorts of issues. Data scientists often have trouble understanding the exact relationship between the two models (Kaur et al. 2020) or explaining this relationship to stakeholders (Passi and Jackson 2018), which is ultimately undermining the value of both models and the whole enterprise (Kumar et al. 2020). The regularization method proposed by Lichtenberg and Şimşek (2019) in order to derive accurate models that approximate tallying might help if it could also be applied to other simple rules.

So, we finish with a couple of rhetorical questions. Could one try simple, process, grounded theory in order to build a model that is at once explanatory and predictive? This approach might not always work of course (for some problems, building good theory could be too expensive). But could this be a place to look at *first*?

## ACKNOWLEDGMENTS

We are grateful to Pantelis Pipergias Analytis, Rachel Haga, Jan Malte Lichtenberg, and Özgür Şimşek for their comments on this work. A preliminary version of this work was presented at the workshop *Data-Driven Personalization, Markets, and Contract Law*, organized at the University of Southampton Law School by Uta Kohl, Jacob Eisler, and James A. Davey, where it benefitted from audience comments, and was published in U. Kohl and J. Eisler (Eds.) (2021). *Data-Driven Personalization and the Law*, Cambridge University Press.

## CONFLICT OF INTEREST

None declared.

## REFERENCES

Analytis, P. P., Barkoczi, D., and Herzog, S. M. 2018. "Social learning strategies for matters of taste." *Nature Human Behaviour* 2(6): 415–24.

Anderson, C. 2008. "The end of theory: The data deluge makes the scientific method obsolete." *Wired Magazine* 16(7): 16–7.

Angwin, J., Larson, J., Mattu, S., and Kirchner, L. 2016. "Machine bias." *ProPublica*.

Attie, E., and Verica, T. 2018. "18 Miles Outside of Roanoke." In *For the People*, edited by P. W Davies. New York, NY: ABC Studios.

Benjamin, R. 2019. "Assessing risk, automating racism." *Science* 366(6464): 421–2.

Bourgin, D. D., Peterson, J. C., Reichman, D., Russell, S. J., and Griffiths, T. L. 2019. "Cognitive model priors for predicting human decisions." In *International Conference on Machine Learning*, 5133–41.

Blake, T., Nosko, C., and Tadelis, S. 2015. "Consumer heterogeneity and paid search effectiveness: A large-scale field experiment." *Econometrica* 83(1): 155–74.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. 1984. *Classification and Regression Trees*. Boca Raton, FL: CRC Press.

Brownsword, R., and Harel, A. 2019. "Law, liberty and technology: Criminal justice in the context of smart machines." *International Journal of Law in Context* 15: 107–25.

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., and Filar, B. 2018. "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation." *ArXiv Preprint:1802.07228*.

Buckmann, M., and Şimşek, Ö. 2017. "Decision heuristics for comparison: How good are they?." *NIPS Workshop on Imperfect Decision Makers* 58: 1–11.

Caliskan, A., Bryson, J. J., and Narayanan, A. 2017. "Semantics derived automatically from language corpora contain human-like biases." *Science* 356(6334): 183–6.

Chiao, V. 2019. "Fairness, accountability and transparency: Notes on algorithmic decision-making in criminal justice." *International Journal of Law in Context* 15: 126–39.

Citron, D. K. 2008. "Technological due process." *Washington University Law Review* 85: 1249–313.

Clarke, R. 2016. "Big data, big risks." *Information Systems Journal* 26: 77–90.

Castle, J. L. (in press). "Comment on "Transparent modeling of influenza incidence: Big data or a single data point from psychological theory"?." *International Journal of Forecasting*.

Czerlinski, J., Gigerenzer, G., and Goldstein, D. G. 1999. "How good are simple heuristics? "In *Simple Heuristics that Make us Smart*, edited by G. Gigerenzer, P. M. Todd, and the ABC Research Group, 97–118. New York, NY: Oxford University Press.

Dawes, R. M., Faust, D., and Meehl, P. E. 1989. "Clinical versus actuarial judgment." *Science* 243(4899): 1668–74.

Dhami, M. K. 2003. "Psychological models of professional decision making." *Psychological Science* 14(2): 175–80.

Epstein, R. A. 2009. *Simple Rules for a Complex World*. Cambridge, MA: Harvard University Press.

Erev, I., Ert, E., Plonsky, O., Cohen, D., and Cohen, O. 2017. "From anomalies to forecasts: Toward a descriptive model of decisions under risk, under ambiguity, and from experience." *Psychological Review* 124(4): 369.

Frederik, J., and Martijn, M. 2019. "The new dot com bubble is here: It's called online advertising." *The Correspondent*, November 6, 2019. https://thecorrespondent.com/100/the-new-dot-com-bubble-is-here-its-called-online-advertising/13228924500-22d5fd24).

Gardner, J. 2015. "The many faces of the reasonable person." *Law Quarterly Review* 131(1): 563–84.

Garfinkel, S., Matthews, J., Shapiro, S. S., and Smith, J. M. 2017. "Toward algorithmic transparency and accountability." *Communications of the ACM* 60(9): 5.

Garza, A. D. L. 2020. "States' Automated Systems Are Trapping Citizens in Bureaucratic Nightmares with Their Lives on the Line." *Time*. https://time.com/5840609/algorithm-unemployment/. Accessed February 17, 2022.

Gigerenzer, G. 2007. *Gut Feelings: The Intelligence of the Unconscious*. London, UK: Penguin.

Gigerenzer, G. (in press). *How to Stay Smart in a Smart World*. Penguin.

Gigerenzer, G., and Gaissmaier, W. 2011. "Heuristic decision making." *Annual Review of Psychology* 62: 451–82.

Gigerenzer, G., and Todd, P. M., and the ABC research group 1999. *Simple Heuristics that Make Us Smart*. New York, NY: Oxford University Press.

Goodman, B., and Flaxman, S. 2017. "European Union regulations on algorithmic decision-making and a "right to explanation."" *AI Magazine* 38(3): 50–7.

Griffiths, T. L. 2015. "Manifesto for a new (computational) cognitive resolution." *Cognition* 135: 21–3.

Harari, Y. N. 2016. *Homo Deus: A Brief History of Tomorrow*. Random House.

Hogarth, R. M., and Karelaia, N. 2005. "Simple models for multiattribute choice with many alternatives: When it does and does not pay to face tradeoffs with binary attributes?." *Management Science* 51: 1860–72.

Holsinger, A. M., Lowenkamp, C. T., Latessa, E., Serin, R., Cohen, T. H., Robinson, C. R., and VanBenschoten, S. W. 2018. "A rejoinder to Dressel and Farid: New study finds computer algorithm is more accurate than humans at predicting arrest and as good as a group of 20 lay experts." *Federal Probation* 82: 50–5.

Holte, R. C. 1993. "Very simple classification rules perform well on most commonly used datasets." *Machine Learning* 11: 63–90.

IEEE. 2019. *Ethically Aligned Design a Vision for Prioritizing Wellbeing with Artificial Intelligence and Autonomous Systems*, Piscataway, NJ.

Jackson, P. 1998. *Introduction to Expert Systems*. Addison Wesley.

Kahneman, D., Slovic, P., and Tversky, A. (Eds.) 1982. *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press.

Katsikopoulos, K. V. 2011a. "How to model it? Review of "Cognitive Modeling" (J. R. Busemeyer and A. Diederich)." *Journal of Mathematical Psychology* 55(2): 198–201.

Katsikopoulos, K. V. 2011b. "Psychological heuristics for making inferences: Definition, performance, and the emerging theory and practice." *Decision Analysis* 8(1): 10–29.

Katsikopoulos, K. V., Durbach, I. N., and Stewart, T. J. 2018. "When should we use simple decision models? A synthesis of various research strands." *Omega – The International Journal of Management Science* 81: 17–25.

Katsikopoulos, K. V., and Gigerenzer, G. 2008. "One-reason decision-making: Modeling violations of expected utility theory." *Journal of Risk and Uncertainty* 37(1): 35–56.

Katsikopoulos, K. V., Şimşek, Ö., Buckmann, M., and Gigerenzer, G. 2020. *Classification in the Wild: The Science and Art of Transparent Decision Making*. MIT Press.

Katsikopoulos, K. V., Şimşek, Ö., Buckmann, M., and Gigerenzer, G. (in press). "Transparent modeling of influenza incidence: Big data or a single data point from psychological theory?." *International Journal of Forecasting* (with commentaries).

Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H., and Wortman Vaughan, J. 2020. "Interpreting Interpretability: Understanding Data Scientist's Use of Interpretability Tools for Machine

Learning." In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14.

Kirchner, L., and Goldstein, M. 2020. "How Automated Background Checks Freeze Out Renters." *The New York Times*.

Kumar, I. E., Venkatasubramanian, S., Scheidegger, C., and Friedler, S. 2020. "Problems with Shapley-value-based explanations as feature importance measures." *arXiv preprint: 2002.11097*.

Lazer, D., Kennedy, R., King, G., and Vespignani, A. 2014. "The parable of Google Flu: Traps in big data analysis." *Science* 343(6176): 1203–5.

LeCun, Y., Bengio, Y., and Hinton, G. 2015. "Deep learning." *Nature* 521(7553): 436–44.

Lessmann, S., Baesens, B., Seow, H. V., and Thomas, L. C. 2015. "Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research." *European Journal of Operational Research* 247(1): 124–36.

Lichtenberg, J., and Şimşek, Ö. 2019. "Regularization in Directable Environments with Application to Tetris." In *Proceedings of the 36th International Conference on Machine Learning Research*, Long Beach, CA.

Lichtenberg, J., and Şimşek, Ö. 2017. "Simple regression models." *Proceedings of the NIPS 2016 Workshop on Imperfect Decision Makers*. PMLR, 58: 13–25.

Lichtman, A. J. 2016. *Predicting the Next President. The Keys to the White House 2016*. Rowman and Littlefield.

Liptak, A. 2017. "Sent to prison by a software program's secret algorithms." *The New York Times*, May 1, 2017.

Lundberg, S. M., and Lee, S. I. 2017. "A unified approach to interpreting model predictions." In *Advances in Neural Information Processing Systems*: 4765–74.

Maier, M., Carlotto, H., Saperstein, S., Sanchez, F., Balogun, S., and Merritt, S. 2020. "Improving the accuracy and transparency of underwriting with AI to transform the life insurance industry." *AI Magazine* 41(3): 78–93.

Makridakis, S., Hyndman, R. J., and Petropoulos, F. 2020. "Forecasting in social settings: The state of the art." *International Journal of Forecasting* 36(1): 15–28.

Martignon, L., and Hoffrage, U. 2002. "Fast, frugal and fit: Simple heuristics for paired comparison." *Theory and Decision* 52(1): 29–71.

Martignon, L., Katsikopoulos, K. V., and Woike, J. K. 2008. "Categorization with limited resources: A family of simple heuristics." *Journal of Mathematical Psychology* 52: 352–61.

Matthews, J. 2020. "Patterns and anti-patterns, principles, and pitfalls: Accountability and transparency in AI." *AI Magazine* 41(1): 82–9.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., and Barton, D. 2012. "Big data: The management revolution." *Harvard Business Review* 90(10): 60–8.

Molnar, C. 2020. *Interpretable* Machine Learning. LuLu.

Neisser, U. 1967. *Cognitive Psychology*. Psychology Press.

Nguyen, A., Yosinski, J., and Clune, J. 2015. "Deep neural networks Are easily fooled: High Confidence Predictions for Unrecognizable Images." In *Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition*, 427–36.

Obermeyer, Z., Powers, B., Vogeli, C., and Mullainathan, S. 2019. "Dissecting racial bias in an algorithm used to manage the health of populations." *Science* 366(6464): 447–53.

Passi, S., and Jackson, S. J. 2018. "Trust in data science: Collaboration, translation, and accountability in corporate data science projects." *Proceedings of the ACM on Human-Computer Interaction* 2(CSCW): 1–28.

Peysakhovich, A., and Näcker, J. 2017. "Using methods from machine learning to evaluate behavioral models of choice under risk and ambiguity." *Journal of Economic Behavior & Organization* 133: 373–84.

Puente, M. 3 July, 2019. "LAPD Pioneered Predicting Crime with Data. Many Police Don't Think It Works." Los Angeles Times.

Ribeiro, M. T., Singh, S., and Guestrin, C. 2016. ""Why should I trust you?": Explaining the predictions of any classifier." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–44.

Richardson, R., Schultz, J., and Crawford, K. 2019. "Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice." *New York University Law Review* 94: 15–55.

Rubinstein, A. S. 1998. *Modeling Bounded Rationality*. MIT Press.

Rudin, C., and Radin, J. 2019. "Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition." *Harvard Data Science Review* 1(2).

Selbst, A. D., and Powles, J. 2017. "Meaningful information and the right to explanation." *International Data Privacy Law* 7(4): 233–42.

Silver, N. 2017. "The Real Story of 2016." *FiveThirtyEight*. https://fivethirtyeight.com/features/the-real-story-of-2016/. Accessed February 17, 2022.

Şimşek, Ö. 2013. "Linear decision rule as aspiration for simple decision heuristics." *Advances in Neural Information Processing Systems* 26: 2904–12.

Snook, B., Taylor, P. J., and Bennell, C. 2004. "Geographic profiling: The fast, frugal, and accurate way." *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* 18(1): 105–21.

Stanley, J. 2017. "Pitfalls of Artificial Intelligence Decisionmaking Highlighted In Idaho ACLU Case." *ACLU*. https://www.aclu.org/blog/privacy-technology/pitfalls-artificial-intelligence-decisionmaking-highlighted-idaho-aclu-case. Accessed February 17, 2022.

Harvard Law Review. 2017. State v. Loomis: Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing, Cambridge, MA.

Stevenson, M. 2018. "Assessing risk assessment in action." *Minnesota Law Review* 103: 303–84.

Stevenson, M., and Doleac, J. 2019. "Algorithmic Risk Assessment in the Hands of Humans." In *EZA Discussion Papers, 12853*.

Sull, D., and Eisenhardt, K. M. 2015. *Simple Rules: How to Thrive in a Complex World*. Houghton Mifflin Harcourt.

Trafton, J. G., Hiatt, L. M., Brumback, B., and McCurry, J. M. 2020. "Using Cognitive Models to Train Big Data Models with Small Data." In *Proceedings of the 19th International Conference on Autonomous Agents and Multi Agent Systems*, 1413–21.

Varner, M. 2020. "Texas Drivers Sue Allstate over Secret "Suckers List."." *The Markup*. https://themarkup.org/allstates-algorithm/2020/05/05/texas-drivers-sue-allstate-over-secret-suckers-list. Accessed February 17, 2022.

Washington, A. L. 2018. "How to argue with an algorithm: Lessons from the COMPAS-ProPublica debate." *Colorado Technology Law Journal* 17: 131–60.

Wu, T. 2019. "Will artificial intelligence eat the law? The rise of hybrid social-ordering systems." *Columbia Law Review* 119: 2001–28.

Zuboff, S. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile Books.

## AUTHOR BIOGRAPHIES

**Konstantinos V. Katsikopoulos** is Professor of Behavioral Science and Head of Research at the University of Southampton Business School, and Chair of the Behavioral OR Group (in the OR Society). He holds a PhD in human factors engineering from UMass Amherst, and has been a visiting assistant professor at MIT and deputy director at the Max Planck Institute for Human Development. His research focuses on integrating standard decision models with people's simple rules of thumb, to build theory fit for practice, as in the monograph "Classification in the Wild: The Science and Art of Transparent Decision Making" (MIT Press).

**Marc C. Canellas** is an Assistant Public Defender for Arlington County, Virginia. He is a member and past Chair of the IEEE-USA Artificial Intelligence Policy Committee. He earned his J.D. from New York University School of Law and holds a Ph.D. in aerospace and cognitive engineering from the Georgia Institute of Technology. His litigation and research focus on affirmative data science, using data science to affirm people's human and civil rights, and on the governance of human–machine systems, particularly with respect to carceral technology in the criminal, housing, and family regulation systems.

**How to cite this article:** Katsikopoulos, K. V., and Canellas, M. C. 2022. "Decoding human behavior with big data? Critical, constructive input from the decision sciences." *AI Magazine* 43: 126–138. https://doi.org/10.1002/aaai.12034