



Transparent modeling of influenza incidence: Big data or a single data point from psychological theory?

Konstantinos V. Katsikopoulos^{a,*}, Özgür Şimşek^b, Marcus Buckmann^c, Gerd Gigerenzer^d

^a University of Southampton, United Kingdom

^b University of Bath, United Kingdom

^c Technical University of Berlin, Germany

^d Max Planck Institute for Human Development, Germany

ARTICLE INFO

Keywords:

Google Flu Trends
Big data
Naïve forecasting
Recency
Simple heuristics

ABSTRACT

Simple, transparent rules are often frowned upon while complex, black-box models are seen as holding greater promise. Yet in quickly changing situations, simple rules can protect against overfitting and adapt quickly. We show that the surprisingly simple recency heuristic forecasts more accurately than Google Flu Trends (GFT) which used big data analytics and a black-box algorithm. This heuristic predicts that “this week’s proportion of flu-related doctor visits equals the proportion from the most recent week.” It is based on psychological theory of how people deal with rapidly changing situations. Other theory-inspired heuristics have outperformed big data models in predicting outcomes, such as U.S. presidential elections, or other uncertain events, such as consumer purchases, patient hospitalizations, and terrorist attacks. Heuristics are transparent, clearly communicating the underlying rationale for their predictions. We advocate taking into account psychological principles that have evolved over millennia and using these as a benchmark when testing big data models.

© 2020 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

Simple forecasting rules can be surprisingly accurate in predicting events in sport, business and crime (Goldstein & Gigerenzer, 2009; Makridakis & Hibon, 2000). This appears to be happening for events that are difficult to predict because of a dynamic time course fraught with uncertainty. While simple rules do not fit past data as well as complex models, under uncertainty they can sometimes predict the future better by reducing error due to overfitting the past. At the same time, simple rules are transparent and can easily be applied, understood, and taught. Yet since these less-is-more effects were shown (Dawes, 1979; Dawes & Corrigan, 1974), they have met

with reactions of disbelief by those who prefer “optimal” models with many free parameters and big data. For instance, when Makridakis and Hibon (1979) showed in the *Journal of the Royal Statistical Society* that simple models predicted better than complex models in 111 time series, the comments published along reveal disbelief and disinterest rather than scientific curiosity (Hogarth, 2012). Makridakis reacted with more competitions, where critical experts could submit their own forecasting methods to predict over thousands of time series, and obtained similar results (Makridakis & Hibon, 2000; Makridakis, Spiliotis, & Assimakopoulos, 2020). He concluded that despite the empirical evidence, those developing complex statistical models “pay little attention or ignore such studies” (cited in Hogarth, 2012, p. 68).

Medicine is also a field where uncertainties, in predicting the course of treatment or infection, abound. While simple rules such as bed-side rules and fast-and-frugal

* Corresponding author.

E-mail address: k.katsikopoulos@soton.ac.uk (K.V. Katsikopoulos).

decision trees are widely practiced by doctors in diagnosis and prediction (Gigerenzer, 2014; Wegwarth, Gaissmaier, & Gigerenzer, 2009), forecasting models used by academics tend to be complex and black-box, independent of whether the situation is fraught with uncertainty or not. In this article, we use a classic application of big data analytics to test how well, in a situation of uncertainty, an extremely simple rule can predict the time-course of health-related behavior.

Big data, such as people's web searches, have been used to capture important outcomes, including unemployment figures (Ettredge, Gerdes, & Karuga, 2005), sales of consumer goods such as films and video games (Goel, Hofman, Lahaie, Pennock, & Watts, 2010) and, prominently, the incidence of influenza (flu; Ginsberg et al., 2009). To build *Google Flu Trends* (GFT), Ginsberg et al. (2009) selected 45 variables from 50 million queries submitted to the Google search engine and combined them linearly to predict the proportion of flu-related doctor visits across the U.S. The model was trained on data from 2003 to 2007 and tested on data from 2007 to 2008. It achieved a 0.97 mean correlation¹ with the weekly estimates from the Centers for Disease Control and Prevention (CDC; Ginsberg et al., 2009). But in 2009 GFT failed to predict the outbreak of the swine flu (Cook, Conrad, Fowlkes, & Mohebbi, 2011; Olson, Konty, Paladini, Viboud, & Simonsen, 2013). As is often the case, if a model fails, it is made more complex, and Google updated the model by increasing the number of variables to approximately 160 (Cook et al., 2011). The updated model overestimated the proportion of flu-related doctor visits in 100 out of 108 weeks from August 2011 to September 2013, in some cases, overshooting by more than 50% (Butler, 2013; Olson et al., 2013). In response, the Google engineers asked “is our model too simple?” and updated it once again in 2013 (Copeland et al., 2013). After a third update in 2014, GFT was shut down in 2015.

Beyond shortcomings in accuracy, GFT has been also strongly criticized for its lack of transparency (Lazer, Kennedy, King, & Vespignani, 2014). The search queries and variables used in the models were not revealed publicly, which prevents replication. For the three model updates, the exact number of variables was not reported. A further, important critique of GFT is that there was no testing of alternative models from approaches different than big data analytics. Such testing was done elsewhere (Goel et al., 2010; Lazer et al., 2014). In one approach, the predictions of GFT were linearly combined with CDC data. In a second approach, time-lagged CDC reports were inputted into a linear regression. One key finding was that GFT was less accurate than regressions on time-lagged CDC reports, as measured by the mean absolute error (MAE; Lazer et al., 2014).

Although regression and other linear models have been labelled as “simple,” studies have shown that economists cannot correctly interpret regression weights (Soyer & Hogarth, 2012) and physicians reject the use of such models in their practice because they do not understand them

(Green & Mehr, 1997). The solution would be to develop models that are understandable and accurate at the same time (Rudin & Radin, 2019). According to a common perception, however, this is not possible because of a purported tradeoff between making a model understandable and making it accurate.

2. The recency heuristic

We provide evidence that forecasting models can be both transparent and accurate. On the basis of theory from psychology and other behavioral sciences, one can build simple heuristics that are accurate in the appropriate decision environments (Gigerenzer, Todd, & the ABC Research Group, 1999; Hertwig, Pleskac, Pachur, & the Center for Adaptive Rationality, 2019; Hogarth & Karelaia, 2007; Katsikopoulos, Şimşek, Buckmann, & Gigerenzer, 2020; Todd, Gigerenzer, & the ABC Research Group, 2012). Psychological research indicates that in cases of change and disruption people rely on recency. Brown's (1838) *law of recency* states that recent experiences come to mind more easily than those from the distant past, and are often the sole information guiding human decisions. Furthermore, people use recency adaptively, depending on the structure of the environment (Anderson & Schooler, 1991). Here, we propose the following *recency heuristic* for influenza incidence:

“Predict that this week's proportion of flu-related doctor visits equals the proportion from the most recent week.”

More precisely, let p_t be the prediction of the heuristic for week t and o_t be the CDC report for week t . The prediction of the recency heuristic for week t equals the CDC report for week $t - 1$; that is, $p_t = o_{t-1}$. This calculation assumes that the prediction is made on a Friday, the day on which CDC releases the report for the previous week according to the official website of the CDC; see https://www.cdc.gov/flu/weekly/overview.htm#anchor_1539281266932.

The recency heuristic relies on a single variable and uses zero free parameters. The heuristic is an instance of *naïve forecasting*, typically frowned upon despite evidence that simple forecasting methods often outperform more complex ones (Dosi, Napoletano, Roventini, Stiglitz, & Treibich, 2020; Green & Armstrong, 2015; Hogarth, 2012; Katsikopoulos, Durbach, & Stewart, 2018; Sherden, 1998). For example, Green and Armstrong (2015) identified 97 quantitative comparisons between simple and more complex forecasting models in the literature. The authors found that, across the different error measures reported, complex models increased error on average by 27% compared with simple models.

3. Method and results

We compared the predictions of the recency heuristic with the predictions of GFT, as well as with those of a two-parameter linear regression on the same variable, o_{t-1} , employed by the heuristic (every week the two parameters of the linear regression were updated, as done by Lazer et al. (2014)). GFT nowcasts flu-related doctor

¹ Correlation is a poor measure of accuracy because it ignores mean bias. For instance, predicted values of 2, 3, 4, and 5 are perfectly correlated with observed values of 12, 13, 14, and 15.

visits and is not subject to the reporting lag of the CDC data.² We used all weeks for which GFT predictions are available, from March 18, 2007 to August 9, 2015. To the best of our knowledge, this is the widest time window on which GFT and alternative models have been compared.

We computed three performance metrics for each model: MAE, as was also done by Lazer et al. (2014), mean absolute percentage error (MAPE), and median relative absolute error (MedRAE). More precisely, let o_t , p_t , and e_t denote, respectively, the observed value (CDC report), the predicted value, and the error (of a model) in week t , with $e_t = p_t - o_t$. Then, over n weeks, $MAE = \frac{1}{n} \sum_{t=1}^n |e_t|$ and $MAPE = 100 \times \frac{1}{n} \sum_{t=1}^n |e_t/o_t|$. When computing MedRAE, relative absolute error was computed by dividing the absolute error of a given model by the absolute error of the recency heuristic, $|e_t/(o_{t-1} - o_t)|$; because the denominator of this ratio was often close to zero and, thus, the ratio had a skewed distribution with many extreme values, we used the median (Hyndman & Koehler, 2006). Recall that the observed values are the proportions of doctor visits that are flu-related, which can range from 0% to 100%. We report prediction errors in percentage points. For example, if observed and predicted values are 5% and 2%, respectively, then the error is 3 percentage points.

Fig. 1 shows the performance of the recency heuristic and GFT for each week. We do not show the performance of linear regression because it is basically indistinguishable from that of the heuristic. The upper panel shows the observed and predicted time series o_t and p_t , and the lower panel shows the percentage error time series $PE_t = 100 \times e_t/o_t$.

Table 1 shows the values of the summary statistics MAE, MAPE, and MedRAE for the recency heuristic, the linear regression, and GFT, as well as the benchmark model of predicting a 0% of flu-related doctor visits for each week.

The MAE in percentage points is 0.20 for the recency heuristic and 0.38 for Google Flu Trends. Similar results hold across all ten surveillance regions of the U.S., with the heuristic having a smaller absolute error in the majority of weeks in all regions. The heuristic also predicts as accurately as the two-parameter linear regression, and in eight of the ten regions the heuristic has a smaller absolute error in the majority of weeks. As can be seen in Table 1, the comparative performance of the models is essentially the same for all three metrics, MAE, MAPE, and MedRAE.³

² Influenza surveillance data collection by CDC is based on a reporting week that starts on Sunday and ends on the following Saturday. The recency heuristic and GFT both predict the percentage of flu-related doctor visits in the current Saturday–Sunday week. The prediction of the recency heuristic becomes available on the Friday of the week, when CDC releases the flu-related doctor visits of the previous week. The prediction from GFT is taken from its archival website <https://www.google.org/flutrends/about/>, where the following is noted: “Each week begins on the Sunday (Pacific Time) indicated for the row. Data for the current week will be updated each day until Saturday (Pacific Time)”. What was archived on the GFT website is the final prediction for the week, which was presumably made on the Saturday of the week.

³ Kandula and Shaman (2019) point out that CDC reports are finalized after the conclusion of a surveillance week. The analyses

Table 1

Summary statistics for the performance of the recency heuristic, linear regression, GFT, and the benchmark model of predicting a 0% of flu-related doctor visits for each week, for all weeks from March 18, 2007 to August 9, 2015, the horizon for which GFT predictions are available. The numbers for MAE and MedRAE are in percentage points and for MAPE in percentages. In all cases, smaller numbers indicate smaller error.

	MAE	MAPE (%)	MedRAE
Recency heuristic	0.20	9%	1.00
Linear regression	0.20	10%	0.99
GFT	0.38	20%	1.87
Benchmark model: Predict zero	1.80	100%	13.62

If we exclude the period before its first update in 2009 in the wake of the swine flu, the MAE of GFT remains at 0.38. This result contradicts any expectation that this update actually improved the performance of GFT. In this period (i.e., after the first update of Google Flu in the wake of the swine flu), the MAE of the heuristic and the regression also remain essentially the same; for both models, MAE = 0.19.

A possible adjustment for the seasonality of the data would be to use the corresponding observation from the previous season. The difficulty is that it is not clear what a season is. If a season is taken to be a year, this seasonal version of the heuristic can be expected to not perform well because flu years are known to be often very different from each other. Indeed, we found that this model missed the swine flu and had a MAE of 0.73.

4. Why did the recency heuristic outperform GFT?

In general, the higher accuracy of the recency heuristic compared with GFT appears to result from the uncertainty of the behavior of viruses as well as the behavior of users who submit search queries. There are at least three specific reasons.

First, GFT was trained with years of data where influenza incidence was high in the winter and low in the summer and thus failed to predict any flu that does not follow this pattern, such as the swine flu (underprediction in the summer of 2009; Fig. 1). Any model which makes substantial use of past data would have the same problem. In contrast, the recency heuristic, which ignores historical trends except for the most recent data point, is not susceptible to this problem and can quickly adapt to the outbreak of a new virus (Fig. 1).

Second, as the lower panel of Fig. 1 shows, the recency heuristic has a fairly constant error over years because it is not influenced by any data except for the most recent

presented here, and in previous work including Lazer et al. (2014), used the final CDC reports. We obtained the initial CDC reports from Kandula and Shaman. Initial CDC reports are not available for weeks 21–39 in 2007 and 2008, weeks 21–37 in 2010 and week 39 in 2013. For the weeks where initial CDC reports are available, we rerun our analyses. The MAE of the recency heuristic, linear regression, GFT, and benchmark model was 0.27, 0.24, 0.42, and 1.94, respectively, all in percentage points. The MAPE for the four models was 13%, 12%, 20%, and 100%, respectively, and the MedRAE 1.00, 1.20, 1.62, and 13.60 (all in percentage points), respectively.

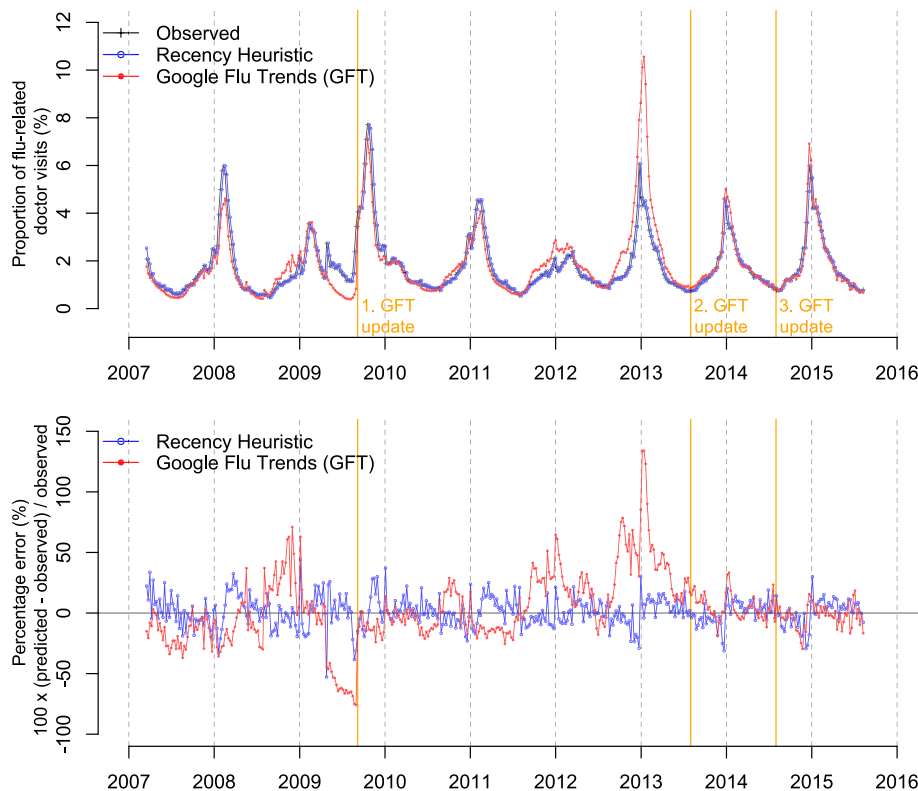


Fig. 1. Performance of the recency heuristic and GFT for all weeks from March 18, 2007 to August 9, 2015, the horizon for which GFT predictions are available; see <https://www.google.org/flutrends/about/>. The dashed vertical lines indicate the first week of each year. The solid vertical lines show the times of the three GFT updates. The predictions of GFT for each week are from the GFT model active at the time (i.e., the most recent GFT update). The upper panel depicts proportion of flu-related doctor visits, observed, and predicted. The lower panel shows the percentage error time series, where values above zero denote over-prediction and values below zero denote under-prediction.

observation. That is not so for GFT, which has a more volatile error. The first two versions of GFT perform worse as time increases from when they were introduced.

Third, the accuracy of GFT might be compromised by people entering search terms out of curiosity—motivated by media reports—rather than medical symptoms (Copeland et al., 2013). In contrast, the predictions of the recency heuristic are not susceptible to this kind of influence of human behavior.

5. Simple heuristics or complex models?

5.1. Recency and other simple heuristics

Since the 1910s, marketing practitioners have been using recency to forecast consumer behavior (Artinger, Kozodoi, vonWangenheim, & Gigerenzer, 2018). Wübben and von Wangenheim (2008) studied empirically the performance of a *hiatus heuristic* for predicting the future purchase behavior of past customers: “Predict that a past customer will continue buying if and only if she or he has made at least one purchase during the last t months”.

The value of the parameter t of the hiatus heuristic was reported to be 9 months by managers in an airline and an apparel firm. The accuracy of this fixed hiatus heuristic was assessed against the purchasing behavior of at least 2330 customers for at least 1.5 years per firm,

and compared with that of a Pareto/negative binomial distribution model (Ehrenberg, 1988; this model's parameters were estimated from the first half of the data). The heuristic achieved lower error than the Pareto model in both comparisons: 23% vs. 27% for the airline and 17% vs. 25% for the apparel firm. In a test with an online-CD retailer, the methods had equal error. Furthermore, in a study of 60 datasets where critical events may or may not be repeated, such as consumer purchases, patient hospitalizations, and terrorist attacks, the hiatus heuristic predicted more accurately than random forests and logistic regression (Artinger et al., 2018).

The recency and hiatus heuristics are examples of heuristics relying on a *single cue* (Hogarth & Karelaia, 2005; Simsek & Buckmann, 2015). Other single-cue heuristics include the *recognition heuristic* (Goldstein & Gigerenzer, 2009) which has been used to predict, for example, financial returns of stocks or sport results such as the winners of the 2005 Wimbledon men's singles' matches, performing as well (70%) as the seedings of the Association of Tennis Professionals (Scheibehenne & Bröder, 2007).

When more than one cue is used, heuristics combine the cues in simple ways, as by *tallying* the cues in a unit-weights linear model (Dawes, 1974) or ordering them in a *fast-and-frugal tree* (Martignon, Katsikopoulos, & Woike, 2008). According to psychological theory, adding

and ordering are among people's core cognitive capacities (Katsikopoulos et al., 2020). A tallying heuristic, derived from a deep study of the U.S. presidential elections from 1860 to 1980 and fixed in 1984, correctly predicted that Donald Trump would win the 2016 U.S. presidential election (Lichtman, 2016), when big data analytics and polls had predicted a high probability of winning for Hilary Clinton (<https://projects.fivethirtyeight.com/2016-election-forecast/>). Lichtman's heuristic has also correctly predicted all U.S. presidential elections since 1984. A fast-and-frugal tree for predicting bank failure, derived after consultation with a team of economists from the Bank of England, performed competitively with the usual tool of financial economics, logistic regression, in out-of-sample prediction of the banks that failed during the 2008 financial crisis (Aikman et al., in press).

5.2. The ecological rationality of simple heuristics

Simple heuristics tend to work well under uncertainty where the future might differ from the past in unpredictable ways, whereas, more complex algorithms work well in more stable situations; this idea is called the *unstable-world principle* (Katsikopoulos et al., 2020). The big successes of the computationally intensive models of AI have been in well-defined games, such as chess and Go, and in relatively stable situations, such as face recognition, while simple heuristics have been found to be more successful when the situation can change rapidly, as in policing, sports, and business (Gigerenzer & Brighton, 2007; Goldstein & Gigerenzer, 2009). Similarly, Sherden (1998, p. 64–65) concluded that economists' forecasts outperformed simple heuristics for predicting highly stable variables, such as government spending, while heuristics outperformed economists' forecasts for highly volatile variables, such as interest rates, and the two were about as accurate in the "middle ground," such as when predicting real GNP growth.

The theory of *ecological rationality* (Todd et al., 2012) suggests that the performance of a strategy, be it simple or complex, should be evaluated in reference to the structure of the environment, as opposed to assuming that complex methods always predict better than simple ones. A more specific way to understand when simple heuristics are likely to predict as or more accurately than complex methods is the *bias-variance decomposition* of the prediction error into the sum of a bias and a variance term (Geman, Bienenstock, & Doursat, 1992): Consider repeatedly sampling training datasets from a population. Bias is the difference between the mean prediction of a model and the true mean value in the population, and it tends to decrease by adding free parameters to a model. Variance is the variability of the individual predictions of the model for each sample, around their mean. Variance tends to increase by adding free parameters to a model.

It is the trade-off between bias and variance, not the bias per se, that determines a model's predictive accuracy. A simple heuristic might compensate for relatively high bias by having low or zero variance. Nikolopoulos and Petropoulos (2018) caution that complex forecasting

methods might be focusing too much on minimizing bias and end up overfitting the training data. Furthermore, there exist precise conditions on the statistical structure of the environment, such as non-compensatoriness, simple, and cumulative dominance, under which the bias of a heuristic is *equal* to the bias of an "optimal" linear model (Baucells, Carrasco, & Hogarth, 2008; Hogarth & Karelaia, 2005; Katsikopoulos, 2011; Katsikopoulos & Martignon, 2006; Martignon & Hoffrage, 2002; Simsek, 2013).

Thus, the superior performance of the recency heuristic over GFT can be understood in a couple of ways. First, consider the unstable-world principle. In situations where the future is not like the past but changes unexpectedly, the recency heuristic can adapt to the change much faster than big data analytics. Second, if the bias of the recency heuristic is not much larger than that of GFT, then the simple heuristic is likely to have smaller total error due to its zero variance (because it has no free parameters). How such factors interact is difficult to determine in the present case, one reason being that the GFT algorithm and its various updates have been kept secret.

6. Discussion

Over-reliance on big data, which are processed by complex models based on mathematical convenience rather than domain theory, has led to failed forecasts in critical situations, such as epidemics, including coronavirus disease 2019 (COVID-19; Ioannidis, Cripps, & Tanner, 2020). For example, SIR (susceptible–infected–recovered) models require data that are not available, such as the number of infected persons in a population, which differs from the available number of persons with a positive test. These models also make unreasonable assumptions such as that every person has the same probability of interacting with everyone else in the population. Furthermore, even if the data were perfect, it is not possible to uniquely identify the parameters of SIR models (Fokas, Cuevas-Maraver, & Kevrekidis, 2020). On the other hand, simple heuristics, such as the recency heuristic, do not rely on big data from the past that may no longer be relevant in the future, avoid overfitting the past in parameter estimation, and are based on psychological theory. Finally, there is a virtue when simplicity and accuracy go together: those who use the algorithm can understand how it works.

The investigation of simple heuristics remains the exception in big data analytics and machine learning. We recommend as a general rule the testing of complex models against simple heuristics that are based on behavioral and psychological theory (Lawrence, Goodwin, O'Connor, & Önköl, 2006). Testing alternative models from different approaches is key to designing meaningful forecasting studies, as for example in the CDC's FluSight competition (Lutz et al., 2019). Such systematic investigations can provide solutions to the alleged accuracy-understandability tradeoff and indicate the situations in which we can expect algorithms that are both accurate and understandable.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We thank the associate editor and two anonymous reviewers for their constructive comments. The analysis of the relative performance of GFT and the recency heuristic reported in this article is an extension of earlier work summarized in (Katsikopoulos et al., 2020).

References

- Aikman, D., Galesic, M., Gigerenzer, G., Kapadia, S., Katsikopoulos, K. V., Kothiyal, A., ..., Neumann, T. (in press) Taking uncertainty seriously: simplicity versus complexity in financial regulation. *Industrial and Corporate Change*.
- Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, 2, 396–408.
- Artinger, F., Kozodoi, N., vonWangenheim, F., & Gigerenzer, G. (2018). Recency: Prediction with smart data. In *AMA winter conference* (p. L2).
- Baucells, M., Carrasco, J. A., & Hogarth, R. M. (2008). Cumulative dominance and heuristic performance in binary multiattribute choice. *Operations Research*, 56(5), 1289–1304.
- Brown, T. (1838). *Lectures on the philosophy of the human mind*. London, UK: William Tait.
- Butler, D. (2013). When google got flu wrong: US outbreak foxes a leading web-based method for tracking seasonal flu. *Nature*, 494, 155–157.
- Cook, S., Conrad, C., Fowlkes, A. L., & Mohebbi, M. H. (2011). Assessing google flu trends performance in the United States during the 2009 H1N1 pandemic. *PLoS One*, 6, 8.
- Copeland, P., Romano, R., Zhang, T., Hecht, G., Zigmond, D., & Stefansen, C. (2013). Google disease trends: an update. In *International society of neglected tropical diseases 2013, international* (p. 3). Society of Neglected Tropical Diseases, Available at: <https://storage.googleapis.com/pub-tools-public-publication-data/pdf/41763.pdf>.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34, 571–582.
- Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, 81, 95–106.
- Dosi, G., Napoletano, M., Rovenenti, A., Stiglitz, J. E., & Treibich, T. (2020). Rational heuristics? expectations and behaviors in evolving economies with heterogeneous interacting agents. *Economic Inquiry*, 58(3), 1487–1516.
- Ehrenberg, A. S. C. (1988). How good is best? *Journal of the Royal Statistical Society, Series A*, 145(3), 364–366.
- Ettredge, M., Gerdes, J., & Karuga, G. (2005). Using web-search data to predict macroeconomic statistics. *Communications of the ACM*, 48, 87–92.
- Fokas, A. S., Cuevas-Maraver, J., & Kevrekidis, P. G. (2020). A quantitative framework for exploring exit strategies from the COVID-19 lockdown. *Chaos, Solitons & Fractals*, 140, Article 110244.
- Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, 4, 1–58.
- Gigerenzer, G. (2014). *Risk savvy. How to make good decisions*. New York: Viking.
- Gigerenzer, G., & Brighton, H. (2007). Can hunches be rational?. *Journal of Law, Economics and Policy*, 4, 155–176.
- Gigerenzer, G., Todd, P. M., & the ABC Research Group (1999). *Simple heuristics that make us smart*. New York, NY: Oxford University Press.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457, 1012–1014.
- Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with Web search. *Proceedings of the National Academy of Sciences*, 107, 17486–17490.
- Goldstein, D. G., & Gigerenzer, G. (2009). Fast and frugal forecasting. *International Journal of Forecasting*, 25(4), 760–772.
- Green, K. C., & Armstrong, J. S. (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research*, 68, 1678–1685.
- Green, L., & Mehr, D. R. (1997). What alters physicians' decisions to admit to the coronary care unit?. *Journal of Family Practice*, 45, 219–226.
- Hertwig, R., Pleskac, T. J., Pachur, T., & the Center for Adaptive Rationality (2019). *Taming Uncertainty*. Cambridge, MA: MIT Press.
- Hogarth, R. M. (2012). When simple is hard to accept. In P. M. Todd, G. Gigerenzer, & the ABC Research Group (Eds.), *Ecological rationality: Intelligence in the world* (pp. 61–79). New York, NY: Oxford University Press.
- Hogarth, R. M., & Karelaia, N. (2005). Ignoring information in binary choice with continuous variables: When is less more? *Journal of Mathematical Psychology*, 49, 115–124.
- Hogarth, R. M., & Karelaia, N. (2007). Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review*, 114(3), 733–758.
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688.
- Ioannidis, J. P., Cripps, S., & Tanner, M. A. (2020). Forecasting for COVID-19 has failed. *International Journal of Forecasting*. Available at: https://www.sciencedirect.com/science/article/pii/S0169207020301199?casa_token=4utons9fc5oAAAAA:7KrD0dUwuZXRcJFojhVmWutxBbaeXFqzJAP1Ula7ghT3ZpBkkwG1tkXovNBDaLrTnw8GUyw.
- Kandula, S., & Shaman, J. (2019). Reappraising the utility of google flu trends. *PLoS Computational Biology*, 15(8), Article e1007258.
- Katsikopoulos, K. V. (2011). Psychological heuristics for making inferences: Definition, performance and the emerging theory and practice. *Decision Analysis*, 8(1), 10–29.
- Katsikopoulos, K. V., Şimşek, Ö., Buckmann, M., & Gigerenzer, G. (2020). *Classification in the wild: The science and art of transparent decision making*. Cambridge, MA: MIT Press.
- Katsikopoulos, K. V., Durbach, I. N., & 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., & Martignon, L. (2006). Naïve heuristics for paired comparisons: Some results on their relative accuracy. *Journal of Mathematical Psychology*, 50(3), 488–494.
- Lawrence, M., Goodwin, P., O'Connor, M., & Önköl, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, 22(3), 493–518.
- Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of google flu: Traps in big data analysis. *Science*, 343, 1203–1205.
- Lichtman, A. J. (2016). *Predicting the next president: The keys to the White House*. Lanham, MD: Rowman and Littlefield.
- Lutz, C. S., Huynh, M. P., Schroeder, M., Anyatonwu, S., Dahlgren, F. S., Danyluk, G., ..., Mgbere, O. (2019). Applying infectious disease forecasting to public health: a path forward using influenza forecasting examples. *BMC Public Health*, 19(1), 1659.
- Makridakis, S., & Hibon, M. (1979). Accuracy of forecasting: An empirical investigation (with discussion). *Journal of the Royal Statistical Society, Series A*, 142, 97–145.
- Makridakis, S., & Hibon, M. (2000). The M3-competition: results, conclusions and implications. *International Journal of Forecasting*, 16, 451–476.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1), 54–74.
- Martignon, L., & Hoffrage, U. (2002). Fast, frugal, and fit: Heuristics for paired comparison. *Theory and Decision*, 52(1), 29–71.
- Martignon, L., Katsikopoulos, K. V., & Woike, J. (2008). Categorization with limited resources: A family of simple heuristics. *Journal of Mathematical Psychology*, 52(6), 352–361.
- Nikolopoulos, K., & Petropoulos, F. (2018). Forecasting for big data: Does suboptimality matter? *Computers and Operations Research*, 98, 322–329.

- Olson, D. R., Konty, K. J., Paladini, M., Viboud, C., & Simonsen, L. (2013). Reassessing google flu trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. *PLoS Computational Biology*, 9, 10.
- Rudin, C., & 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.
- Scheibehenne, B., & Bröder, A. (2007). Predicting wimbledon 2005 tennis results by mere player name recognition. *International Journal of Forecasting*, 23(3), 415–426.
- Sherden, W. A. (1998). *The Fortune Sellers: The Big Business of Buying and Selling Predictions*. New York: John Wiley and Sons.
- Şimşek, Ö. (2013). Linear decision rule as aspiration for simple decision heuristics. In *Advances in neural information processing systems*, vol. 26 (pp. 2904–2912).
- Şimşek, Ö., & Buckmann, M. (2015). Learning from small samples: An analysis of simple decision heuristics. In *Advances in neural information processing systems*, vol. 28 (pp. 3159–3167).
- Soyer, E., & Hogarth, R. M. (2012). The illusion of predictability: How regression statistics mislead experts. *International Journal of Forecasting*, 28, 695–711.
- Todd, P. M., Gigerenzer, G., & the ABC Research Group (2012). *Ecological rationality: Intelligence in the world*. New York, NY: Oxford University Press.
- Wegwarth, O., Gaissmaier, W., & Gigerenzer, G. (2009). Smart strategies for doctors and doctors-in-training: heuristics in medicine. *Medical Education*, 43, 721–728.
- Wübben, M., & von Wangenheim, F. (2008). Instant customer base analysis: Managerial heuristics often get it right. *Journal of Marketing*, 72(3), 82–93.