



*Citation for published version:*

Katsikopoulos, KV, Şimşek, Ö, Buckmann, M & Gigerenzer, G 2022, 'Reply to commentaries on "Transparent modelling of influenza incidence": Recency heuristics and psychological AI', *International Journal of Forecasting*, vol. 38, no. 2, pp. 630-634. <https://doi.org/10.1016/j.ijforecast.2021.10.011>

*DOI:*

[10.1016/j.ijforecast.2021.10.011](https://doi.org/10.1016/j.ijforecast.2021.10.011)

*Publication date:*

2022

*Document Version*

Peer reviewed version

[Link to publication](#)

*Publisher Rights*

CC BY-NC-ND

**University of Bath**

**Alternative formats**

If you require this document in an alternative format, please contact:  
[openaccess@bath.ac.uk](mailto:openaccess@bath.ac.uk)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

## Reply to commentaries on “Transparent modelling of influenza incidence”:

### Recency heuristics and psychological AI

Konstantinos Katsikopoulos, Özgür Şimşek, Marcus Buckmann, and Gerd Gigerenzer

On June 11, 2018, before the World Cup in football began, *Goldman Sachs* announced that after hours of crunching numbers, one million simulations, and 200,000 probability trees induced, it had produced forecasts for the matches (Stehn et al., 2018). In the semi-finals, Germany would beat Portugal, while Brazil would advance over France; and in the final, Brazil would prevail over Germany. The Goldman Sachs team wrote: “We are drawn to machine learning models because they can sift through a large number of possible explanatory variables to produce more accurate forecasts than conventional alternatives [such as Poisson regression]” (Stehn et al., 2018, p. 3; footnote 1).

Goldman Sachs failed to predict the reigning world champion—in the final, France beat Croatia. But this is not our point. This had happened in the 2014 World Cup too, when a less quantitatively sophisticated Goldman Sachs model had also predicted that Brazil would win the Cup (the prediction was that Brazil would win 2-1 over Germany in the semi-finals but a 1-7 loss ensued instead and eventually Germany won the Cup). Our point is that the Goldman Sachs economists reacted to failed predictions in a similar way as the Google engineers did a few years before when they revised Google Flu Trends three times (Katsikopoulos et al., in press). The economists increased the number of predictors by adding individual player characteristics to team

statistics and making the model more complex. The Goldman Sachs team employed the heuristic rule: “When a model fails, make the data bigger and the model more complex”.

Like every heuristic, this rule is rational in some situations but not in all.

Specifically, increasing the number of data points and model complexity might lead to higher accuracy in stable worlds but not necessarily in unstable or rapidly changing situations such as the flu and sports (Goodwin, in press; Katsikopoulos et al., in press). Here, simple rules can predict equally well, or even better, while being more transparent than complex models. In our view, the question is not whether simple models or complex ones are more accurate. Rather, the fundamental research task is to address questions such as: For which situations do big data and complex algorithms make statistical forecasting more accurate? For which situations are simple heuristics more accurate? Can one build a formal theory that explains the results for the two classes of situations? What a-priori theoretical guidance can one provide for selecting forecasting models?

To answer these questions, one needs to experiment and uncover the structure of the ecology of forecasting, and analyze the match between this structure and forecasting models, complex as well as simple. Because “forecasting performance is context dependent” (Fildes and Petropoulos, 2015, p. 1742), the forecasting model should be chosen to match the situation. This mapping of models to situations comprises the theory of *ecological rationality* (Gigerenzer and Gaissmaier, 2011; Gigerenzer, Todd, and the ABC Research Group, 1999; Hogarth and Karelaia, 2007; Todd, Gigerenzer, and the ABC Research Group, 2012). In principle, both big data and

complex models as well as small data and simple heuristics have regions of superior performance, and the goal is to draw these regions in a map that practitioners can use (Hogarth and Karelaia, 2007).

Castle (in press) introduces the term *adaptability* and proposes that a main question of a theory of which model to use when—i.e., a theory of ecological rationality—should be which models are adaptable and which are not. Adaptability refers to a model’s ability to capture changes in the underlying data-generating process. Castle cites examples from the literature wherein complex models were found to be more adaptable than simpler ones. At the same time, she shows that, for a type of change in a linear data-generating process—for more details and a discussion of her analysis, see next section—the recency heuristic is more adaptable than the forecasting model that would be optimal assuming that there is no change. The precise relationship of adaptability to simplicity/complexity needs to be investigated systematically, and such investigations should be embedded in research programs in ecological rationality. A difference between Castle’s notion of adaptability and the concept of ecological rationality is that adaptability refers to a single model whereas ecological rationality refers to the meta-level of choosing the right model for a given situation.

Ecological rationality is one focus of the approach of *psychological AI*, Herbert Simon’s classic idea of using insights from how people make decisions to build smart computer models (Simon, 1969; Gigerenzer et al., 1999; Katsikopoulos et al., 2020). The recency heuristic in Katsikopoulos et al. (in press) is one example of this approach. A second focus of psychological AI is the *transparency* of models to their intended

users, which might include researchers, policy makers, practitioners, and laypeople. A model is transparent to a group of users if it meets the following four requirements: people can understand, memorize, teach, and apply the model (Katsikopoulos et al, 2020). The recency heuristic (Katsikopoulos et al., in press) satisfies these four requirements. Note that, in our definition, transparency refers only to the formulation of heuristics or other models, not to the processes by which heuristics or models have evolved, been discovered, or derived.

In our view, transparency is a protected value: In a democracy, practitioners who employ models as well as citizens who are affected by the output of models, especially in sensitive domains such as health and justice, have the right to understand why, as Moss (in press) writes, “the model said so”, and how a model output led to a decision. Furthermore, not only is transparency a theoretical principle, but a lack of transparency can have serious negative effects in practice. Such effects include loss of public trust and hence decreased adoption (Moss, in press), or increased tinkering of practitioners with the model output, for example, in order to ascertain ownership of the process, while inadvertently leading to double counting of the same predictor (Goodwin, Moritz, and Siemsen, 2018).

The commentators to our paper (Katsikopoulos et al., in press) provide useful suggestions for studying issues of transparency and ecological rationality in forecasting, with which we engage in the next two sections of this reply. We discuss how ecological rationality focuses on the comparative empirical study and theoretical analysis of

different types of models, and how the notion of transparency in psychological AI differs from the approach of explainable AI.

### **Ecological Rationality**

Ecological rationality is an analytical discipline that studies the conditions under which heuristics and other models succeed. An important component of ecological rationality is the systematic empirical comparison of complex models and simple heuristics, as all four commentators have endorsed. How should such testing be performed?

First, as Goodwin (in press) emphasizes, the accuracy of a forecasting model should be tested by using data different from those used to train the model, whereas business, economics, and management often rely on data fitting without assessing predictive power (Aikman et al., 2021). There are two different ways of following this guideline: *out-of-sample* and *out-of-population* testing. In out-of-sample testing, a standard in machine learning, both the training and test set are randomly drawn from the same population. The evaluation of the recency heuristic and Google Flu Trends went beyond out-of-sample testing and employed out-of-population testing. This means that the models were tested in a population different from the population from which the training set had been drawn. For example, models can be tested in predicting the yet-unknown future or in predicting events in other geographic areas. In out-of-population prediction, entirely unforeseen outcomes can occur, such as the swine flu. Predicting out of population is much more difficult than out-of-sample prediction.

Castle (in press), as well as Ben Taieb and Taylor (in press), discuss how to reduce overfitting in multiparameter models. We note that the recency heuristic has no free parameters—similar to many other simple heuristics (Gigerenzer, Hertwig, and Pachur, 2011)—so the heuristic cannot overfit by definition. Ben Taieb and Taylor (in press), as well as Moss (in press), suggest the use of weighted averages of models for forecasting Covid-19 incidence so that over-estimation and under-estimation from individual models can cancel out. The weights can be learned from the past performance of the models although, as the commentators acknowledge, the available historical data might be limited, as in cases such as the evolving Covid-19 pandemic. Such ensembles of models can reduce to a single member of the ensemble under conditions specified by Schurz (2019). Castle (in press) claims that overfitting can be controlled by using only those predictors that are “extremely statistically significant”. No evidence is provided for the empirical efficacy of this approach, however. In Google Flu Trends, reducing 50 million search terms to the best performing 45 ones did not prevent the model from failing. Arguably, neither extremely small significance levels nor ensemble techniques are able to deal with undetected and unexpected system shocks. These shocks can happen in out-of-population but not in out-of-sample prediction. We agree with Castle that system shocks present a major challenge to accurate forecasting. Simple heuristics such as the recency heuristic can deal efficiently with a shock in a rapidly changing environment. Ben Taieb and Taylor (in press) suggest that “global” machine-learning models, which are trained across different datasets at once, might be able to outperform simple heuristics in unstable or rapidly changing environments. We

think that this is an excellent question for a research program of systematic empirical comparisons.

In general, complex models with many free parameters will suffer from error due to variance in out-of-sample prediction and from bias in out-of-population prediction (e.g., shocks). In contrast, the recency heuristic, which has no free parameters, will suffer from bias but not from variance in out-of-sample prediction and can be more robust in out-of-population prediction, as in Castle's (in press) shock model. Similar results hold for other simple heuristics, and conditions are known for which a simple heuristic has exactly the same bias as a more complex model; see next section. On the other hand, simple heuristics might not predict as well as more complex models in stable worlds.

### **Theoretical analysis of heuristics**

The issues of out-of-population prediction and overfitting are relevant for evaluating performance but the focus of the study of ecological rationality is on deriving conditions under which a given model, for example the recency heuristic, performs well relative to other models. Castle (in press), in Section 3 of her commentary, performs an analysis of the ecological rationality of the recency heuristic. She assumes that the data generating process is linear, that is, the upcoming observation is a linear function of the most recent observation plus a white noise term and another perturbation term. The model is dynamic, meaning that at some point the mean of the distribution of observations shifts. Castle (in press) realistically assumes that this shift is undetected by the forecaster. She



shows mathematically that, in this situation, the recency heuristic is more *robust* than the forecasting model that would be optimal assuming no mean shift: In the next period after the mean shift, the recency heuristic has an expected bias of zero whereas the “optimal” forecasting model has a positive bias.

Castle’s analysis, and other comparisons of so-called naïve forecasting with more complex models (Goodwin, Petropoulos, and Hyndman, 2017), are in the spirit of existing mathematical analyses of the ecological rationality of judgment and decision models (for a review, see Katsikopoulos, 2011). For instance, for a linear data-generating process—called environment or ecology in the judgment and decision making literature—several conditions are known under which a heuristic achieves a zero or small (relative to a complex model) bias, such as non-compensatoriness in the space of attributes (Hogarth and Karelaia, 2005; Katsikopoulos, 2013; Katsikopoulos and Martignon, 2006; Martignon and Hoffrage, 2002; Şimşek, 2013), and cumulative or simple dominance in the space of decision options (Baucells, Hogarth, and Carrasco, 2008). Other analyses present conditions under which the heuristic of using only one variable outperforms more complex benchmarks such as multiple linear regression (Davis-Sober, Dana, and Budescu, 2010a; 2010b). Dosi et al. (2020) propose a recency heuristic for forming financial expectations in macro-economic environments, build mathematical models of such environments, and propose conditions under which the heuristics can outperform models which are deemed “rational” in macroeconomics.

As Castle notes, the relationship between simplicity and robustness is worthy of further investigation. Ben Taieb and Taylor (in press) write that machine learning

methods “often struggle to beat simple benchmarks especially on short and highly noisy series.” Finding the precise conditions under which such phenomena occur is another task of ecological rationality.

As a reviewer of this rejoinder pointed out, questions of ecological rationality should not necessarily be framed in an “either-or” fashion. *Hybrid* models, combining ideas from the behavioral sciences and techniques from big-data analytics or standard forecasting, can be built and analyzed as well. In the context of predicting influenza incidence, a possibility is to combine the forecast modeling technique of *trend damping*, which has also been observed in human forecasting (Reimers and Harvey, 2011; Harvey and Reimers, 2013), with the recency heuristic. Green (2021) did so, and independently we also combined trends with heuristics, as shown below.

### **Empirical analysis of heuristics and hybrid models**

Let  $o_t$  be the observation for time period (in the case of influenza, week)  $t$  and  $p_t$  be a prediction for time period  $t$ . Recall that in the recency heuristic  $p_t = o_{t-1}$ . Green (2021) proposed the hybrid model that predicts  $p_t = o_{t-1} + a(o_{t-1} - o_{t-2})$ , where  $a$  is a parameter controlling the extent to which the latest trend ( $o_{t-1} - o_{t-2}$ ) is damped. Note that setting  $a = 0$  means ignoring the trend and using the recency heuristic. Damping trends can improve forecasting accuracy, possibly remedying concerns about the long-term accuracy of relying exclusively on recency (Green, 2021; Moss, in press). We estimated  $a$  as 0.44 by minimizing the MAE of forecasting CDC data from March 17,

2004 to March 17, 2007

([https://www.cdc.gov/flu/weekly/overview.htm#anchor\\_1539281266932](https://www.cdc.gov/flu/weekly/overview.htm#anchor_1539281266932)).

An alternative approach to fixing the value of  $a$  is to consider how people might reason heuristically about trends. Recall the recency heuristic: “Predict that this period’s proportion of flu-related doctor visits equals the proportion from the most recent period” (Katsikopoulos et al., in press). Applying recency in the same way to trends (rather than to observations) leads to a new psychological heuristic, the *trend-recency heuristic*:

“Predict that this period’s trend in the proportion of flu-related doctor visits equals the trend in the proportion from the most recent period”.

That is, the trend-recency heuristic predicts that  $p_t - o_{t-1} = o_{t-1} - o_{t-2}$ , and hence  $p_t = o_{t-1} + (o_{t-1} - o_{t-2})$ . The heuristic corresponds to  $a = 1$  in Green’s model, which means no trend damping. Thus, both recency heuristics are nested within the family of trend-damping models. We now present an analysis of various trend damping models and heuristics.

Table 1 provides the mean absolute error  $MAE = \frac{1}{n} \sum_{t=1}^n |p_t - o_t|$  and the mean absolute percentage error  $MAPE = 100 \times \frac{1}{n} \sum_{t=1}^n |(p_t - o_t)/o_t|$  on the test set used in Katsikopoulos et al. (in press), from March 18, 2007 to August 9, 2015, for the trend-damping model with  $a = 0.44$  and for the trend-recency heuristic. For completeness, the table also includes the MAE and MAPE of the recency heuristic, as well as linear regression on the most recent observation (as done by Lazer et al., 2014), Google Flu Trends, and the benchmark model of always predicting zero.

	<b>MAE</b>	<b>MAPE (%)</b>
<b>Trend Damping Model</b> <b>(<math>\alpha = 0.44</math>)</b>	0.17	8.7%
<b>Trend-Recency Heuristic</b> <b>(<math>\alpha = 1</math>)</b>	0.19	10.4%
<b>Recency Heuristic</b> <b>(<math>\alpha = 0</math>)</b>	0.20	9.4%
<b>Linear Regression</b>	0.20	9.8%
<b>Google Flu Trends</b>	0.38	19.8%
<b>Benchmark: Predict Zero</b>	1.80	100%

Table 1. Summary statistics for the performance of an estimated trend-damping model, the two recency heuristics, linear regression, Google Flu Trends, and a benchmark, for predicting the proportion of flu-related doctor visits for each week from March 18, 2007 to August 9, 2015 ([https://www.cdc.gov/flu/weekly/overview.htm#anchor\\_1539281266932](https://www.cdc.gov/flu/weekly/overview.htm#anchor_1539281266932)), which is the test set used in Katsikopoulos et al. (in press). The numbers for MAE are in percentage points and for MAPE in percentages.

From Table 1 we see that incorporating or damping trends only slightly improves forecasting accuracy, as measured by MAE, beyond the recency heuristic (which ignores trends). When accuracy is measured by MAPE, the recency heuristic actually outperforms the trend-recency heuristic and is within 1% of the estimated trend-damping model. Using a recency-based model, with or without trends, appears to be a

robust choice<sup>1</sup>. In Austria, using a simple trend model led to more accurate Covid-19 forecasts than those produced by a team of experts advising the government who relied on a combination of extended SIR models, agent-based SIR models, and state-space epidemiological clockwork models (Schweiger, 2021).

As suggested by Ben Taieb and Taylor (in press), providing an indication of uncertainty around a point estimate, as the ones produced by recency heuristics, could support scenario planning and risk management, as well as public discourse, for example during a pandemic (Taleb, Bar-Yam, and Cirillo, 2020). A challenge for future research is generating such indications of uncertainty for recency heuristics, or any other model, accurately, based on evidence rather than merely mathematically convenient assumptions (e.g., using the actual empirical distributions of financial investment returns rather than just assuming normal distributions).

## Transparency

Moss (in press) pointedly stresses that at a time where models are increasingly influencing public policy, the justification for using a model should not be “because the model said so...” but rather “the model said so, because...” Indeed, transparency of the rationale underlying a model’s predictions in sensitive legal, financial, or medical contexts, to its users such as citizens, policy makers, and other decision makers, should

---

<sup>1</sup> There might be a “flat-maximum” effect (Lovie and Lovie, 1986) with regard to setting the value of the damping parameter  $a$ , as for example for  $a = 0.20$  MAE is 0.183 and MAPE is 8.8%, and for  $a = 0.70$  MAE is 0.176 and MAPE is 9.1%. We checked all  $a$  from 0 to 1 in increments of 0.01, and found that MAE ranged from 0.172 to 0.198 and MAPE ranged from 8.7% to 10.4%. In forecasting problems where there is such an effect, research resources may be diverted from parameter estimation to other important activities, such as gathering higher quality data.

be a right in a democracy, as advocated by the General Data Protection Regulation of the European Union.

It is not clear that epidemiological models of the SIR variety meet standards of transparency. Ben Taieb and Taylor (in press) assert that “simple SIR-type models are also transparent” but they do not address *to whom* are these models transparent. As Goodwin (in press), highlights, users “may not be mathematicians or statisticians”. Researchers might find SIR models easy to understand, but this is far from obvious for doctors, politicians, and laypeople. Moss (in press) is right in saying that the atomic rules that these models are based on are simple on their own, but the interaction of these rules and the emergent behavior are not necessarily simple; actually, the fact that simplicity somehow can give rise to complexity is a standard selling point of agent-based simulation models (Conte and Paolucci, 2014).

As Moss (in press) says, the use of ensemble models may reduce transparency. In a large comparative analysis of machine learning algorithms for credit scoring, Lessmann et al. (2015, p. 134) note: “The difficulties of introducing advanced scoring methods including ensemble models are more psychological than business related. Using a large number of models, a significant minority of which give contradictory answers, is counterintuitive to many business leaders”.

The AI community is increasingly acknowledging the importance of explainable models. However, based on the (misleading) assumption that complex models are always more accurate than simple ones, current efforts are focused on the development of mathematical techniques to explain the predictions of complex black-box models,

rather than developing simple, transparent, and accurate models in the first place (Rudin and Radin, 2019). For example, DARPA (Defense Advanced Research Projects Agency) presumes that algorithms which are more understandable to people must also be less accurate; see <https://www.darpa.mil/attachments/DARPA-BAA-16-53.pdf> (right-hand panel of Figure 5, p. 14). But there are plenty of examples where more transparent heuristics are also more accurate than complex black-box models (Katsikopoulos et al., 2020). Gigerenzer (in press) and Katsikopoulos and Canellas (in press) provide examples from areas such as jury decision making and predictive policing where the bias towards the use of black box algorithms led to waste of resources with no discernible gains in accuracy while leading to increases in racial and other kinds of discrimination.

Two of our commentators invoke George Box's motto that *all models are wrong but some are useful*. We wholeheartedly agree. For us, a useful model is both accurate and transparent. In situations of uncertainty (as opposed to risk; Aikman et al., 2021), one route to achieving these objectives is psychological AI.

### References

- Aikman, D., Galesic, M., Gigerenzer, G., Kapadia, S., Katsikopoulos, K., Kothiyal, A., ... and Neumann, T. (2021). Taking uncertainty seriously: simplicity versus complexity in financial regulation. *Industrial and Corporate Change*, 30(2), 317-345.
- Baucells, M., Carrasco, J. A., and Hogarth, R. M. (2008). Cumulative dominance and heuristic performance in binary multiattribute choice. *Operations Research*, 56(5), 1289-1304.

- Ben Taieb, S., and Taylor, K. S. (in press). On big data models for infectious disease forecasting: Commentary on “Transparent modeling of influenza incidence: Big data or single data point from psychological theory?” *International Journal of Forecasting*.
- Castle, J. (in press). Commentary on “Transparent modeling of influenza incidence: Big data or single data point from psychological theory?” *International Journal of Forecasting*.
- Conte, R., and Paolucci, M. (2014). On agent-based modeling and computational social science. *Frontiers in Psychology, 5*, 668.
- Davis-Stober, C. P., Dana, J., and Budescu, D. V. (2010a). A constrained linear estimator for multiple regression. *Psychometrika, 75*(3), 521-541.
- Davis-Stober, C. P., Dana, J., and Budescu, D. V. (2010b). Why recognition is rational: Optimality results on single-variable decision rules. *Judgment and Decision Making, 5*(4), 216-229.
- Dosi, G., Napoletano, M., Roventini, A., Stiglitz, J. E., and Treibich, T. (2020). Rational heuristics? expectations and behaviors in evolving economies with heterogeneous interacting agents. *Economic Inquiry, 58*(3), 1487-1516.
- Fildes, R., and Petropoulos, F. (2015). Is there a golden rule? *Journal of Business Research, 68*(8), 1741-745.
- Gigerenzer, G. (in press). *How to Stay Smart in a Smart World*. London, UK: Penguin.
- Gigerenzer, G., and Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62*, 451-482.
- Gigerenzer, G., Hertwig, R., and Pachur, T. (2011). *Heuristics: The Foundations of Adaptive Behavior*. New York, NY: Oxford University Press.
- Goodwin, P. (in press). Commentary on “Transparent modeling of influenza incidence:



- Big data or single data point from psychological theory?" *International Journal of Forecasting*.
- Goodwin, P. (2011). High on complexity, low on evidence: Are advanced forecasting methods always as good as they seem? *Foresight: The International Journal of Applied Forecasting*, 23, 10-12.
- Goodwin, P., Moritz, B., and Siemsen, E. (2018). Forecast decisions. In K. Donohue, E. Katok, and S. Leider (Eds). *The Handbook of Behavioral Operations* (pp. 433-458). New York, NY: Wiley.
- Goodwin, P., Petropoulos, F., and Hyndman, R. J. (2017). A note on upper bounds for forecast value-added relative to naïve forecasts. *Journal of the Operational Research Society*, 68, 1082-1084.
- Green, K. C. (2021). Comparison of forecasts of weekly weighted average US percentage of doctor visits associated with flu symptoms, *Unpublished article*. Adelaide: University of South Australia.
- Harvey, N., and Reimers, S. (2013). Trend damping: Under-adjustment, experimental artifact, or adaptation to features of the natural environment? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(2), 589.
- Hogarth, R. M., and Karelaia, N. (2005). Simple models for multiattribute choice with many alternatives: When it does and does not pay to face trade-offs with binary attributes. *Management Science*, 51(12), 1860-1872.
- Hogarth, R. M., and Karelaia, N. (2007). Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review*, 114(3), 733-758.
- 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. (2013). Why do simple heuristics perform well in choices with binary attributes? *Decision Analysis*, 10(4), 327-340.
- Katsikopoulos, K. V., and Canellas, M. C. (in press). Decoding human behavior with big data? Critical, constructive input from the decision sciences. *AI Magazine*.
- Katsikopoulos, K. V., and Martignon, L. (2006). Naïve heuristics for paired comparisons: Some results on their relative accuracy. *Journal of Mathematical Psychology*, 50(3), 488-494.
- Katsikopoulos, K. V., Şimşek, Ö., Buckmann, M., and Gigerenzer, G. (2020). *Classification in the Wild: The Science and Art of Transparent Decision Making*. Cambridge, MA: 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*.
- Lazer, D., Kennedy, R., King, G., and Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. *Science*, 343, 1203-1205.
- 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), 124-136.
- Lovie, A. D., and Lovie, P. (1986). The flat maximum effect and linear scoring models for prediction. *Journal of Forecasting*, 5(3), 159-168.
- Martignon, L., and Hoffrage, U. (2002). Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory and Decision*, 52(1), 29-71.
- Moss, R. (in press). Because the model so: Commentary on “Transparent modeling of Influenza incidence: Big data or single data point from psychological theory?” *International Journal of Forecasting*.

- Nikolopoulos, K., and Petropoulos, F. (2018). Forecasting for big data: Does suboptimality matter? *Computers and Operations Research*, 98, 322-329.
- Reimers, S., and Harvey, N. (2011). Sensitivity to autocorrelation in judgmental time series forecasting. *International Journal of Forecasting*, 27(4), 1196-1214.
- 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.
- Schurz, G. (2019). *Hume's Problem Solved: The Optimality of Meta-Induction*. Cambridge, MA: MIT Press.
- Schweiger, R. (2021). Corona-prognose: Weit daneben ist auch vorbei (in German), *Profil* (May 20, 2021); <https://www.profil.at/oesterreich/corona-prognose-weit-daneben-ist-auch-vorbei/401387751>.
- Simon, H. A. (1969). *The Sciences of the Artificial*. Cambridge, MA: MIT Press.
- Şimşek, O. (2013). Linear decision rule as aspiration for simple decision heuristics. In *Advances in Neural Information Processing Systems*, 26, 2904-2912.
- Stehn, S. J., Trivedi, K., Fawcett, N., Chaudhary, M., Incoronato, L, and the Global Macro Research Team (2018). *The World Cup and Economics 2018*. New York, NY: Goldman Sachs Global Investment Research.
- Taleb, N. N., Bar-Yam, Y., and Cirillo, P. (2020). On single point forecasts for fat-tailed variables. *International Journal of Forecasting*.
- Todd, P. M., Gigerenzer, G., and the ABC Research Group (2012). *Ecological Rationality: Intelligence in the World*. New York, NY: Oxford University Press.