This postprint was originally published by Elsevier as:
Katsikopoulos, K. V., 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, 38*(2), 630–634.
https://doi.org/10.1016/j.ijforecast.2021.10.011

Reply to commentaries on "Transparent modelling of influenza incidence": Recency heuristics and psychological AI

Konstantinos V. Katsikopoulos[a,*], Özgür Şimşek[b], Marcus Buckmann[c], Gerd Gigerenzer[d]

[a] University of Southampton Business School, United Kingdom
[b] University of Bath, United Kingdom
[c] Technical University of Berlin, Germany
[d] Max Planck Institute for Human Development, Germany
[*] Corresponding author, E-mail address: k.katsikopoulos@soton.ac.uk

Abstract

We structure this response to the commentaries to our article ‘‘Transparent modeling of influenza incidence: Big data or a single data point from psychological theory?’’ around the concept of *psychological AI*, Herbert Simon’s classic idea of using insights from how people make decisions to make computers smart. The recency heuristic in Katsikopoulos, Şimşek, Buckmann, and Gigerenzer (2021) is one example of psychological AI. Here we develop another: the *trend-recency* heuristic. While the recency heuristic predicts that the next observation will equal the most recent observation, the trend-recency heuristic predicts that the next trend will equal the most recent trend. We compare the performance of these two recency heuristics with forecasting models that use trend damping for predicting flu incidence. Psychological AI prioritizes *ecological rationality* and *transparency*, and we provide a roadmap of how to study such issues. We also discuss how this transparency differs from explainable AI and how ecological rationality focuses on the comparative empirical study and theoretical analysis of different types of models.

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, Trivedi, Fawcett, Chaudhary, & Incoronato, 2018). In the semi-finals, Germany would beat Portugal, while Brazil would advance over France; 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 [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., 2021). 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, Şimşek, Buckmann, & Gigerenzer, 2020; Katsikopoulos et al., 2021). 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 do they not? Can one build a formal theory that explains the results for the two classes of situations? What theoretical guidance can one provide for selecting forecasting models?

To answer these questions, one needs to experiment and reveal 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 & 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 & Gaissmaier, 2011; Gigerenzer, Todd, & the ABC Research Group, 1999; Hogarth & Karelaia, 2007; Todd, Gigerenzer, & 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 (Hogarth & Karelaia, 2007).

Castle (in press) introduces the term *adaptability* and proposes that the 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 the recency heuristic is more adaptable than the ‘‘optimal’’ forecasting model for a type of change in a linear data-generating process (for more details and a discussion of her analysis, see the next section). 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.

A second focus of psychological AI is the *transparency* of different models to their intended users, which might include researchers, policy makers, practitioners, and laypeople (Katsikopoulos et al., 2020). A model is transparent to a group of users if they can understand, memorize, teach, and apply it (Katsikopoulos et al., 2020). 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 also 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, to ascertain ownership of the process, while inadvertently leading to double counting of the same predictor (Goodwin, Moritz, & Siemsen, 2018).

The commentators to our paper (Katsikopoulos et al., 2021) provide useful suggestions for studying issues of transparency and ecological rationality in forecasting, with which we engage in the next two sections of this response.

1. 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, a guideline not always followed in business, economics, and management which continue to rely only on data fitting. 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 sets 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. Out-of-population prediction 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, & Pachur, 2011)—therefore, the heuristic cannot overfit by definition. Ben Taieb and Taylor (in press) and 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 negate each other. 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, however, is provided for the empirical efficacy of this approach. 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 prediction but not in outof-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 quickly 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 agree that this is an excellent direction for a program of systematic empirical comparisons.

1.1 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 theoretical ‘‘optimal’’ forecasting model: 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 the so-called naïve forecasting with ‘‘optimal’’ models (Goodwin, Petropoulos, & Hyndman, 2017) are in the spirit of existing mathematical analyses of the ecological rationality of decision models (for a review, see Katsikopoulos, 2011). For instance, for a linear data-generating process—called environment or ecology in the decision literature— several conditions are known under which a heuristic achieves a zero or small bias, such as non-compensatory weights (Hogarth & Karelaia, 2005; Katsikopoulos, 2013; Martignon & Hoffrage, 2002) or non-compensatory validities (Katsikopoulos & Martignon, 2006) of attributes, and cumulative or simple dominance in the space of decision options (Baucells, Carrasco, & Hogarth, 2008; Şimşek, 2013). Other analyses present conditions such as minimax (Davis-Stober, Dana, & Budescu, 2010a, 2010b), under which the heuristic of using only one variable outperforms more complex benchmarks such as multiple linear regression. Dosi, Napoletano, Roventini, Stiglitz, and Treibich (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 ‘‘optimal’’ models.

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’’. The precise conditions under which such phenomena occur is another question 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 (Harvey & Reimers, 2013; Reimers & Harvey, 2011), with the recency heuristic. Green (2021) did so, and independently, we also combined trends with heuristics, as shown below.

1.2. Empirical analysis of heuristics and hybrid models

Let o_t be the observation for time period (in the case of influenza, week) t. Recall that the prediction of the recency heuristic for t is $p_t = o_{t-1}$. 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).

An alternative approach to fix 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., 2021). 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.

Table 1 provides the mean absolute error $\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n} | p_t - o_t |$ and the mean absolute percentage error $\mathrm{MAPE} = 100 \times \frac{1}{n}\sum_{t=1}^{n} | e_t / o_t |$ on the test set used in Katsikopoulos et al. (2021), 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, Kennedy, King, & Vespignani, 2014, Google Flu Trends, and the benchmark model of always predicting zero.

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), which is the test set used in Katsikopoulos et al. (2021). The numbers for MAE are in percentage points and for MAPE in percentages.

	MAE	MAPE (%)
Trend damping model (a = 0.44)	0.17	8.7%
Trend-recency heuristic (a = 1)	0.19	10.4%
Recency heuristic (a = 0)	0.20	9.4%
Linear regression	0.20	9.8%
Google Flu trends	0.38	19.8%
Benchmark: Predict zero	1.80	100%

From Table 1, we see that incorporating or damping trends can only slightly improve 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[1]. In Austria, using the maximum trend observed over any two days in the most recent week 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, especially during a pandemic (Taleb, Bar-Yam, & 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 mathematically convenient assumptions.

2. 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 be a right in a democracy.

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 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 simple; actually, the fact that simplicity somehow gives rise to complexity is a standard selling point of these models.

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, Baesens, Seow, and Thomas (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 assumption that complex models are generally 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 & Radin, 2019). For example, DARPA (Defense Advanced Research Projects Agency) presumes that algorithms that are more understandable to people must also be less accurate; see https://www.darpa.mil/attachments/DARPA-BAA-16-53.pdf (Figure 5, p. 14). But there are many examples where more transparent heuristics are also more accurate than black-box complex models (Katsikopoulos et al., 2020). Gigerenzer (in press) and Katsikopoulos and Canellas (2021) provide examples from areas such as jury decision-making and predictive policing where the bias toward the use of black-box algorithms led to waste of resources with no discernible gains in accuracy while leading to increase in racial and other kinds of discrimination.

[1] There might be a ''flat-maximum'' effect (Lovie & 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%.

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. One possible route to achieve these objectives is psychological AI.

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.

References

Baucells, M., Carrasco, J. A., & Hogarth, R. M. (2008). Cumulative dominance and heuristic performance in binary multiattribute choice. *Operations Research*, *56*(5), 1289–1304.

Ben Taieb, S., & Taylor, K. S. (2021). 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*, in press.

Castle, J. (2021). Commentary on transparent modeling of influenza incidence: Big data or single data point from psychological theory?. *International Journal of Forecasting*, in press.

Davis-Stober, C. P., Dana, J., & Budescu, D. V. (2010a). A constrained linear estimator for multiple regression. *Psychometrika*, *75*(3), 521–541.

Davis-Stober, C. P., Dana, J., & 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., & Treibich, T. (2020). Rational heuristics? expectations and behaviors in evolving economies with heterogeneous interacting agents. *Economic Inquiry*, *58*(3), 1487–1516.

Fildes, R., & Petropoulos, F. (2015). Is there a golden rule? *Journal of Business Research*, *68*(8), 1741–1745.

Gigerenzer, G. (2021). *Staying smart in a smart world. London*. UK: Penguin, in press.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482.

Gigerenzer, G., Hertwig, R., & Pachur, T. (2011). *Heuristics: The foundations of adaptive behavior*. New York, NY: Oxford University Press.

Gigerenzer, G., Todd, P. M., & the ABC Research Group (1999). *Simple Heuristics that Make Us Smart*. New York, NY: Oxford University Press.

Goodwin, P. (2021). Commentary on transparent modeling of influenza incidence: Big data or single data point from psychological theory? *International Journal of Forecasting*, in press.

Goodwin, P., Moritz, B., & Siemsen, E. (2018). Forecast decisions. In K. Donohue, E. Katok, & S. Leider (Eds.), *The handbook of behavioral operations* (pp. 433–458). New York, NY: Wiley.

Goodwin, P., Petropoulos, F., & 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*. Adelaide:University of South Australia, Unpublished article.

Harvey, N., & 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., & 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., & 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., & Canellas, M. C. (2021). Decoding human behavior with big data? Critical, constructive input from the decision sciences. *AI Magazine*, in press.

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., Şimşek, Ö., Buckmann, M., & Gigerenzer, G. (2021). Transparent modeling of influenza incidence: Big data or a single data point from psychological theory? *International Journal of Forecasting*.

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.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of google flu: Traps in big data analysis. *Science*, *343*, 1203–1205.

Lessmann, S., Baesens, B., Seow, H. V., & 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., & Lovie, P. (1986). The flat maximum effect and linear scoring models for prediction. *Journal of Forecasting*, *5*(3), 159–168.

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

Moss, R. (2021). Because the model so: Commentary on transparent modeling of influenza incidence: Big data or single data point from psychological theory? *International Journal of Forecasting*, in press.

Reimers, S., & Harvey, N. (2011). Sensitivity to autocorrelation in judgmental time series forecasting. *International Journal of Forecasting*, *27*(4), 1196–1214.

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*.

Schurz, G. (2019). *Hume's problem solved: The optimality of metainduction*. 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.

Şimşek, O. (2013). Linear decision rule as aspiration for simple decision heuristics. *Vol. 26*, In *Advances in neural information processing systems* (pp. 2904–2912).

Stehn, S. J., Trivedi, K., Fawcett, N., Chaudhary, M., Incoronato, L., & Global Macro Research Team (2018). *The world cup and economics 2018*. New York, NY: Goldman Saschs Global Investment Research.

Taleb, N. N., Bar-Yam, Y., & Cirillo, P. (2020). On single point forecasts for fat-tailed variables. *International Journal of Forecasting*.

Todd, P. M., Gigerenzer, G., & the ABC Research Group (2012). *Ecological rationality: intelligence in the world*. New York, NY: Oxford University Press.