A Psychologically-Informed Approach to "Actuarial" Decision Making

Almog Simchon[a,b] and Michael Gilead[c]

[a]Department of Psychology, Ben-Gurion University of the Negev, Israel

[b]School of Psychological Science, University of Bristol, UK

[c]School of Psychological Science, Tel Aviv University, Israel

Acknowledgments

This research was supported by the Ministry of Science & Technology, Israel.

Abstract

Effective decision-making is crucial but often marred by human biases and limitations. Statistical prediction methods have consistently outperformed human judgment, especially in complex and uncertain domains. Recent advancements in machine learning offer further opportunities to improve statistical predictions. While the prospect of human obsolescence arises, we argue that a collaborative approach is still essential. This article reviews recent work emphasizing the integration of human expertise in the development of statistical models that support human judgment. Three key aspects are explored: informed feature extraction, informed priors, and informed data collection. By integrating human expertise, machine learning can produce superior predictive models, allowing for better decision support systems. Collaboration between humans and algorithms remains crucial in leveraging the strengths of both, advancing decision-making capabilities across various domains.

Effective decision-making is a crucial element in many aspects of life, encompassing personal choices and professional decisions that can have far-reaching effects. Nonetheless, humans often err in their decision-making processes, at times neglecting to utilize all available information to make well-informed choices. Even when people do gather and analyze information, they remain susceptible to various cognitive biases and limitations that can impact their decisions, leading to suboptimal decision-making and undesired consequences.

Since the pioneering work of Paul Meehl (1954), it has been firmly established that statistical prediction methods (or “actuarial prediction”) often outperform human judgment. In situations that involve complex data or substantial uncertainty, statistical models demonstrate superior accuracy and predictive power than human judgment. For instance, in finance, statistical models are widely employed to make investment decisions as they can swiftly process vast data and identify patterns that human analysts may not detect (Dixon et al. 2020). Similarly, in medicine, statistical models have been developed to aid disease diagnosis and predict patient outcomes, leading to enhanced accuracy in comparison to human diagnosis and prognosis (Rajkomar et al., 2019).

Recent advances in machine learning hold the potential to further improve actuarial predictions in domains of critical importance. New machine learning algorithms offer improvements over standard statistical models by allowing computers to learn patterns and make predictions from data without explicit programming. These algorithms can detect intricate patterns in big data, including non-linear relationships and interactions, which traditional statistical models may find challenging or impossible to identify. For example, machine learning models were shown to provide important gains in areas such as bail decisions (Kleinberg et al., 2018). Specifically, advanced models were shown to decrease crime recidivism rates without substantially increasing incarceration rates, potentially providing real benefits to human decision-makers.

These advances raise the possibility that human judgment and expertise could become less important in certain domains. As machine learning algorithms become more advanced and better able to learn from vast amounts of data, it is possible that they will eventually become capable of making decisions and predictions without the need for human input or oversight. This could lead to a future where the importance of domain knowledge significantly decreases, and the main thing needed to make accurate predictions and decisions is more data. However, we argue that such a scenario is still far off. In the meantime, there is much room for integration between human expertise and machine learning in developing better decision support systems.

In the current article, we review recent work that highlights the importance of incorporating human expertise in the development of statistical models that aid human judgment. Specifically, we focus on our work showing how psychological domain knowledge can be integrated with machine learning to produce better predictive models. In particular, we will argue for the utility of informed **feature extraction**, informed **priors**, and informed **data collection**.

Informed Feature Extraction

The reason prediction problems are so cumbersome is that the world is highly multidimensional; in light of this, it is often difficult to identify which dimensions of a given phenomenon are predictive of a given outcome. For example, consider the scenario of trying to identify whether a given job applicant is going to be a good worker. Each and every human has numerous character traits, qualifications, proficiencies, and limitations. A purely data-driven approach will use every bit of information that can be quantified about this person in order to predict their suitability for a job: their height, social security number, and favorite sandwich. When the amount of data is huge, an inductive algorithm may be able to separate the wheat from the chaff and identify those dimensions that contain information about future job success. However, in many realistic situations, even in relatively large datasets, the potential dimensionality of the data is so big, such that even a very large number of observations is not enough. This so-called "curse of dimensionality" often requires domain expertise in limiting the search space to theory-relevant dimensions rather than noise (e.g., trying to predict workplace suitability from a star sign).

The idea that expertise can be harnessed for better model performance has been exemplified in several lines of research and was first introduced by Einhorn (1972), who argued that while expert opinion is valuable for making informed decisions, their judgments may not be as accurate as a statistical model that combines their input. This observation has later been established and well-documented (e.g., Ganzach et al., 2000; Kuncel et al., 2013). In another line of research, attempting to predict human choice behavior in relatively small datasets has shown that models that rely on features that are derived from cognitive psychology theories can outperform purely inductive machine-learning models (Erev et al., 2017). However, research also shows that even more accurate predictions can be attained by using theoretically derived features as predictors for powerful algorithms (Plonsky et al., 2017).

Aside from modeling human choice behavior, such a synergy between humans and machines may also potentially help people in choice tasks such as personnel selection. For example, in recent work, Levitin et al. (in prep) tried to predict the degree to which NBA players are prone to on-court transgressive behaviors from their social media posts. Studying this highly unique population of elite athletes entailed that the number of data observations was relatively small (252 athletes)--which meant that a high-dimensional representation of the text data (e.g., using text embeddings; Pennington et al., 2014) would be especially prone to overfitting. In light of this, for this prediction problem, the authors utilized classic psychological theory concerning the structure of personality–The Five-Factor Model (FFM; McCrae & John, 1992). The authors used a language model that accurately identifies big-five personality characteristics from social media posts (previously trained on much larger samples; Park et al., 2015). This resulted in a parsimonious, five-dimensional representation of each player's personality, which can then be used as predictors of their tendency to commit technical fouls. The results showed that this approach generated significantly better than chance predictions (i.e., out-of-sample predictive performance; r = 182) despite the small dataset.

While the FFM is a highly useful representation of individuals' personalities, it has been argued that its predictive validity is modest (e.g., Morgeson et al., 2007). However, this does not mean that this five-dimensional representation of human personality provides an upper bound on the ability to predict life outcomes from a parsimonious personality representation. In recent work (Lavi et al., 2022), we relied on machine learning methods and big data to develop an alternative representation of personality that is geared towards prediction. This five-dimensional representation (i.e., the “predictive five”), was shown to outperform the FFM in a host of meaningful prediction tasks (e.g., prediction of IQ, depression).

While such representations as the FFM or the “predictive five” model may be useful, it is often the case that a higher-dimensional set of predictors is necessary in order to find a model that fits the data well. In such cases, psychologists can rely on their ability to engineer novel features based on their expertise, intuitions, past findings, and theories. Such ad-hoc engineered features can then be used as inputs for inductive models that rely on some form of regularization to separate the wheat from the chaff.

This was exemplified by Simchon and Gilead (2018). In the CLPsych Shared Task 2018 (Lynn et al., 2018), teams of computer scientists, linguists, and psychologists competed in a prediction problem. They were given essays of 11-year-old children, written more than 50 years ago, and were asked to predict their psychological distress in childhood and adulthood. Simchon and Gilead (2018) took a simple modeling approach (i.e., linear regression) but a theoretically elaborated approach to feature engineering. For example, whereas many teams simply corrected spelling errors and fed the inputs into neural networks, we counted the proportion of spelling errors and included them in the model under the assumption that they may reveal some useful information about learning disabilities that may foresee distress. This approach has proven itself comparable to systems leveraging neural networks (Lynn et al., 2018) and outperformed all other teams in the only subtask that required forecasting across people and time[1].

[1] The CLPsych 2018 Shared Task aimed to forecast mental health states based on essays written during childhood. Task A concentrated on identifying the psychological well-being of individuals at the time the essay was written (age 11); Task B focused on predicting the mental health outcomes of the participants at future ages—specifically 22, 33, 42, and 50—using the essays they composed when they were 11 years old. The age 50 scores were purposefully omitted to provide an out-of-sample evaluation across individuals and time.

Thus, a theoretically informed approach favoring explainability and simplicity can not only compare to much more algorithmically-sophisticated systems but can also top the charts in prediction problems.

Informed Priors

One of the reasons that actuarial decision-making can outperform human judgment is that such methods are free from humans' tendency for motivated reasoning and hypothesis confirmation bias. In many cases, people form a hypothesis in their mind (e.g., people who smile in the interview are going to be good team players) and then seek the observations that support this conclusion. The strength of actuarial approaches to human judgment is that they remain truthful to the data, whatever it may say.

However, the method of approaching every new dataset as if you were a *tabula rasa* is inconsistent with standard scientific practice. In classic scientific inquiry, we seek to build on prior findings and theories, and by "standing on the shoulders of giants" we are able to reach the stars. The successes of the classic scientific method in themselves should inform us that the reliance on prior findings and knowledge is nonetheless of merit. More specifically, by accumulating and aggregating evidence across different scientific contexts (e.g., labs, procedures, scientists, eras), we are able to identify generalizable truths that transcend a given dataset or experimental context.

Early research on linear models highlighted their robustness and how this quality can be beneficial in decision-making processes (Dawes & Corrigan, 1974). These "improper linear models" are weighted by intuition, or heuristics can be surprisingly good. For example, the integration of domain knowledge in deriving intuition-based weights was found to be superior compared to a single expert prediction (i.e., "clinical prediction"; Dawes, 1979). This reasoning was later echoed in other works, suggesting that exerting (limited) domain knowledge in the form of a simple heuristic outperforms move "advanced" computations (Katsikopoulos et al., 2010).

In traditional machine learning methods, the ability to generalize across the specific dataset is attained by "regularizing" our predictors; namely, by incurring a high burden of proof on each predictor and "shrinking" it in a way that maximizes across different folds of the data (Tibshirani, 1996). However, another less popular method of regularizing statistical models is applying Bayesian regularization methods (van Zwet & Gelman, 2022).

In the Bayesian framework, model parameters require setting prior probability distributions, which, together with the observed data, make the posterior probability distribution – the distribution from which conclusions are drawn. Effectively, in the Bayesian framework, the researcher incorporates prior expectations into their results. Incorporating expectations built on domain knowledge concerning the distribution of effect size in a given field (which, normatively, should be symmetric and heavily skewed centered around zero; van Zwet & Gelman, 2022) is an empirically informed way to know what types of effects are feasible, and which effects should be heavily regularized.

Moreover, the use of informed priors can allow us to avoid disregarding previous findings that were not borne out in our current dataset. In other words, if previous literature suggests that a given dimension (e.g., education) should be highly predictive of an outcome (e.g., job success), but this relationship is masked in our data because of some specific properties of these data, the use of these research-driven priors will magnify (rather than shrink) the value of the predictor, alleviating the danger that the model will under-fit future observations.

Such an approach to applying informed regularization was presented by (Gamoran et al., 2021). In the 2021 version of CLPsych Shared Task[2] (MacAvaney et al., 2021), teams competed in a prediction task of a sensitive nature, suicide attempts. The data included tweets of suicidal individuals donated to OurData-Helps.org. The tasks required identifying suicide attempts within 30 days and 180 days. With the same principles in mind as Simchon and Gilead (2018), the system described by Gamoran et al. (2021) used simple linguistic features and a general modeling approach (logistic regression). However, in addition to informed feature extraction, Gamoran and colleagues used the reported effect size of various linguistic features in the literature and as priors in their Bayesian model, which proved to be successful. The informed prior models ranked first in the False Acceptance Rate (FAR) across both tasks and came in second in the second task's main performance metric (F1), providing evidence for the utility of informed priors.

Surely, the reliance on informed priors of effect sizes can often decrease the effectiveness of predictions. Previous studies may suggest an effect where it does exist or needlessly shrink estimates of parameters that could have increased predictive accuracy. This is especially true when previously acquired estimates come from outdated or different samples. However, in the

[2] The CLPsych 2021 Shared Task was centered on predicting users' suicide attempts by analyzing their social media activity. For Task A, participants needed to make predictions based on the users' posts from the preceding 30 days, while Task B expanded the data set to include posts from the last 180 days for prediction.

long run, to the extent that one is seeking a highly generalizable model, the gradual integration of priors and evidence is an optimal prediction strategy.

Informed Data Collection

As noted above, the ability to generate a good predictive model is primarily reliant on the size and scope of the data. Huge datasets are the building blocks of incredible models like GPT (Brown et al., 2020). Often (but not always), big data protects from overfitting, minimizes bias, and enables the signal to surpass the noise. Nonetheless, we should consider the words of Peter Norvig: “More data beats clever algorithms, but better data beats more data”. Ensuring the quality of data is, of course, a substantial part of any data collection effort, but what constitutes “better” data?

The notion of “better” can be divided into two dimensions: lower “noise” (i.e., cleaner data) and higher “signal” (i.e., more useful information). Traditionally, data scientists take the amount of signal in the data as a given and focus on cleaning their data to reduce noise (e.g., identifying outliers, nonsensical responses, typos, and so forth). Such data cleansing can typically be done without much domain expertise; however, domain expertise is of essence whenever we are able to shape the process of data collection, such that it best captures a useful signal. Developing tasks that flesh out the behavior of interest *is* the expertise of experimental psychologists. Data collection, thus, can also be informed by psychological theory and domain knowledge. This nuance in approach—prioritizing informed data collection that maximizes relevant signal—might be obvious to psychologists. Nevertheless, for those in machine learning and related fields, this concept of enhancing data quality from the very start of data gathering might not be as evident.

Consider the cumbersome task of geopolitical forecasting. Over the past decade, the world has been confronted with a spectrum of phenomena, encompassing environmental calamities, transformative technological breakthroughs, and cultural and societal transformations. The capacity to foresee such events empowers society to proactively prepare and mitigate potential risks. In light of this, it is tempting to believe that machine learning algorithms could soon be used to predict geopolitical events. However, unlocking society's most pressing questions—requires a level of reasoning that machines have yet to master. Machine models heavily depend on copious "training data" to fine-tune their parameters and deliver precise predictions for uncharted territories. Yet, when faced with limited or chaotic data, the bedrock of machine learning crumbles (McAndrew et al., 2021).

In recent work (Shinitzky et al., 2023), we tried to address this question by developing a data collection process that may yield a meaningful signal upon which machine learning methods could work. Specifically, in early 2020, we collected geopolitical predictions from 153 individuals concerning 10 pressing geopolitical topics (e.g., when will a Covid vaccine become available; will the loser of the US election concede their defeat). We collected from each forecaster rich data concerning themselves and their predictions. The collected features included various psychological traits and tendencies that we associated in previous research with accurate forecasting (e.g., participants' ability to predict the consensus response; their level of open-mindedness). These data were given as inputs to machine learning models that sought to predict whether a given prediction was likely to be accurate. The resulting model had area under the curve (AUC) values of more than 0.8 in held-out training data, highlighting the potential of this approach.

Another example can be seen in a recent study by Simchon and Gilead (in prep), in which informed data collection was key to the prediction task. One psychological insight often disregarded is the concept of a *projective test*. The theory that underlies projective tests dates back to the late 19th century, the idea of revealing the subconscious through the interpretation of ambiguous stimuli or the spontaneous generation of free associations that may expose latent beliefs and mental representations of psychological importance (Rabin, 1968). While some researchers claim that these tools are effective for personality and psychopathology assessment (Viglione & Rivera, 2003), current implementations of projective tests in psychology have been shown to be inadequate in terms of their reliability, sensitivity, and specificity (Garb et al., 2002).

However, it is possible that the current limitations of the projective-test method can be overcome by leveraging new machine learning methods that may be more sensitive to the statistical regularities in the data. A recent study revealed that the semantic space trajectory of a free association task ("forward flow") was diagnostic in predicting creativity (Gray et al., 2019). Simchon and Gilead applied that logic to depression prediction. Using ten chains (cycles) of associations of length 10 (words), they build a model that predicts CES-D (Eaton et al., 2004) depression scores with $r = .28$. The data will be released to the scientific community with the intention to harness better algorithms, applied to better data, and outperform the current benchmark set by the authors.

Finally, with further advances in generative modeling, it is likely that future work could increasingly rely on theory-derived artificial data as part of the training process. In this case, instead of informing data collection, experts will inform the process of generating artificial data; the artificial data can then be used as a basis for training and fine-tuned by actual human data.

Indeed, recent research (Bourgin et al., 2019) suggests that such approaches may be promising, especially when human-generated datasets are relatively limited in size.

Conclusion

It has been seventy years since Meehl's (1954) seminal work highligyhing the superior value of statistical judgment over expert evaluations. While this insight has been incorporated into many areas of human decision-making, the use of expert evaluations still remains popular in numerous crucial contexts. Recent advances in statistical modeling and artificial intelligence are likely to increase humans' reliance on actuarial judgment and in many areas, improve it. Specifically, the ability of machine learning models to accurately predict human outcomes (e.g., what is the likelihood that candidate X will be successful? What is the likelihood that person Y is at risk?) can serve as inputs to decision algorithms that maximize the utility of decision-makers (e.g., should I hire this person? Should I invest time and effort in preventive treatments?

Despite the developments of predictive modeling in many domains of life, the predictability of psychological outcomes remains quite elusive (e.g., Joel et al., 2020; Salganik et al., 2020). We argue that the integration of psychological expertise and machine learning is important for developing better systems for making decisions concerning human behavior.

We suggest that psychological domain knowledge can contribute to the improvement of predictive models in three key ways. Firstly, through informed feature extraction, psychologists can leverage their expertise to identify theory-relevant dimensions and engineer novel features based on past findings and theories, enhancing the model's ability to capture meaningful predictors. Secondly, through informed priors, psychologists can incorporate prior expectations into the model by setting probability distributions based on domain knowledge, allowing for regularization that aligns with existing scientific knowledge and avoids underfitting or disregarding relevant predictors. Finally, through informed data collection, psychologists can shape the process of data collection to capture a more useful signal, drawing on psychological theory and experimental design expertise to develop tasks that elicit behavior of interest and yield higher-quality data for modeling purposes.

References

Bourgin, D. D., Peterson, J. C., Reichman, D., Russell, S. J., & Griffiths, T. L. (09--15 Jun 2019). Cognitive model priors for predicting human decisions. In K. Chaudhuri & R. Salakhutdinov (Eds.), *Proceedings of the 36th International Conference on Machine Learning* (Vol. 97, pp. 5133–5141). PMLR.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language Models are Few-Shot Learners. In *arXiv [cs.CL]*. arXiv. http://arxiv.org/abs/2005.14165

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *The American Psychologist*, *34*(7), 571–582.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, *81*(2), 95–106.

Dixon, M. F., Halperin, I., & Bilokon, P. (n.d.). *Machine Learning in Finance*. Springer International Publishing.

Eaton, W. W., Smith, C., Ybarra, M., Muntaner, C., & Tien, A. (2004). Center for Epidemiologic Studies

Depression Scale: review and revision (CESD and CESD-R). In M. E. Maruish (Ed.), *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment* (Vol. 3). Lawrence Erlbaum Associates Publishers.

Einhorn, H. J. (1972). Expert measurement and mechanical combination. *Organizational Behavior and Human Performance*, *7*(1), 86–106.

Erev, I., Ert, E., Plonsky, O., Cohen, D., & Cohen, O. (2017). From anomalies to forecasts: Toward a descriptive model of decisions under risk, under ambiguity, and from experience. *Psychological Review*, *124*(4), 369–409.

Gamoran, A., Kaplan, Y., Orr, R. I., Simchon, A., & Gilead, M. (2021). Using Psychologically-Informed Priors for Suicide Prediction in the CLPsych 2021 Shared Task. *Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access*, 103–109.

Ganzach, Y., Kluger, A. N., & Klayman, N. (2000). Making decisions from an interview: Expert measurement and mechanical combination. *Personnel Psychology*, *53*(1), 1–20.

Garb, H. N., Wood, J. M., & Lilienfeld, S. O. (2002). Effective use of projective techniques in clinical practice: Let the data help with selection and interpretation. *: Research and Practice*. https://psycnet.apa.org/record/2002-18346-004

Gray, K., Anderson, S., Chen, E. E., Kelly, J. M., Christian, M. S., Patrick, J., Huang, L., Kenett, Y. N., & Lewis, K. (2019). “Forward flow”: A new measure to quantify free thought and predict creativity. *The American Psychologist*, *74*(5), 539–554.

Joel, S., Eastwick, P. W., Allison, C. J., Arriaga, X. B., Baker, Z. G., Bar-Kalifa, E., Bergeron, S., Birnbaum, G. E., Brock, R. L., Brumbaugh, C. C., Carmichael, C. L., Chen, S., Clarke, J., Cobb, R. J., Coolsen, M. K., Davis, J., de Jong, D. C., Debrot, A., DeHaas, E. C., … Wolf, S. (2020). Machine learning uncovers the most robust self-report predictors of relationship quality across 43 longitudinal couples studies. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(32), 19061–19071.

Katsikopoulos, K. V., Schooler, L. J., & Hertwig, R. (2010). The robust beauty of ordinary information. *Psychological Review*, *117*(4), 1259–1266.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). HUMAN DECISIONS AND MACHINE PREDICTIONS. *The Quarterly Journal of Economics*, *133*(1), 237–293.

Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: a meta-analysis. *The Journal of Applied Psychology*, *98*(6), 1060–1072.

Lavi, G., Rosenblatt, J., & Gilead, M. (2022). A prediction-focused approach to personality modeling. *Scientific Reports*, *12*(1), 12650.

Lynn, V., Goodman, A., Niederhoffer, K., Loveys, K., Resnik, P., & Schwartz, H. A. (2018). CLPsych 2018 shared task: Predicting current and future psychological health from childhood essays. *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, 37–46.

MacAvaney, S., Mittu, A., Coppersmith, G., Leintz, J., & Resnik, P. (2021). Community-level Research on Suicidality Prediction in a Secure Environment: Overview of the CLPsych 2021 Shared Task. *Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access*, 70–80.

McAndrew, T., Wattanachit, N., Gibson, G. C., & Reich, N. G. (2021). Aggregating predictions from experts: a review of statistical methods, experiments, and applications. *Wiley Interdisciplinary Reviews. Computational Statistics*, *13*(2). https://doi.org/10.1002/wics.1514

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*(2), 175–215.

Meehl, P. E. (1954). Clinical versus statistical prediction: A theoretical analysis and a review of the evidence. In *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* (pp. x, 149 – x, 149). University of Minnesota Press. https://doi.org/10.1037/11281-000

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). ARE WE GETTING FOOLED AGAIN? COMING TO TERMS WITH LIMITATIONS IN THE USE OF PERSONALITY TESTS FOR PERSONNEL SELECTION. In *Personnel Psychology* (Vol. 60, Issue 4, pp. 1029–1049). https://doi.org/10.1111/j.1744-6570.2007.00100.x

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., & Seligman, M. E. P. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, *108*(6), 934–952.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.

Plonsky, O., Erev, I., Hazan, T., & Tennenholtz, M. (2017). Psychological Forest: Predicting Human Behavior. *Proceedings of the AAAI Conference on Artificial Intelligence*, *31*(1). https://doi.org/10.1609/aaai.v31i1.10613

Rabin, A. I. (1968). Projective Methods: An Historical Introduction. In A. I. Rabin (Ed.), *Projective Techniques in Personality Assessment* (pp. 3–17). Springer Berlin Heidelberg.

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine. *The New England Journal of Medicine*, *380*(14), 1347–1358.

Salganik, M. J., Lundberg, I., Kindel, A. T., Ahearn, C. E., Al-Ghoneim, K., Almaatouq, A., Altschul, D. M., Brand, J. E., Carnegie, N. B., Compton, R. J., Datta, D., Davidson, T., Filippova, A., Gilroy, C., Goode, B. J., Jahani, E., Kashyap, R., Kirchner, A., McKay, S., … McLanahan, S. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(15), 8398–8403.

Shinitzky, H., Shemesh, Y., Leiser, D., & Gilead, M. (2023). Improving geopolitical forecasts with 100 brains and one computer. *International Journal of Forecasting*. https://doi.org/10.1016/j.ijforecast.2023.08.004

Simchon, A., & Gilead, M. (2018). A Psychologically Informed Approach to CLPsych Shared Task 2018. *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, 113–118.

Simchon, A., & Gilead, M. (n.d). *Informed data collection for depression prediction* [Manuscript in preparation].

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, *58*(1), 267–288.

van Zwet, E., & Gelman, A. (2022). A Proposal for Informative Default Priors Scaled by the Standard Error of Estimates. *The American Statistician*, *76*(1), 1–9.

Viglione, D. J., & Rivera, B. (2003). Assessing Personality and Psychopathology with Projective Methods. In I. B. Weiner (Ed.), *Handbook of Psychology* (Vol. 6, p. 319). John Wiley & Sons, Inc.