Predictive and Causal-Explanatory Modeling in Psychology

A Review of Current Practices and an Illustrative Comparison and Integration of the Two Approaches

Kimberly Klugescheid

January 15th, 2024

Abstract

Traditionally, psychology has employed mainly causal-explanatory modeling. However, with the increasing prominence of predictive modeling in various domains of social sciences, new research questions have arisen and results uncovered. This paper delves into the differences between the two approaches in the modeling process while emphasizing distinct advantages and limitations associated with each approach. It also addresses challenges inherent in predictive modeling that are especially significant in the social science context, such as theory-dependent variables, limited sample sizes, and data reliability concerns. An integrative approach is proposed, which iteratively combines causal-explanatory and predictive modeling, to utilize the benefits of both approaches and minimize their limitations. Researchers are encouraged to weigh the advantages of predictive modeling in generating new insights against the challenges inherent in its application.

1 Introduction

From a statistical point of view, data modeling can be performed in a descriptive, causal-explanatory, or predictive manner. For reasons of simplicity, this paper focuses on the latter two. On the one hand, causal-explanatory modeling aims to shed light on causal relationships between predictor variables and an outcome variable. On the other hand, predictive modeling aims at predicting new observations that are not necessarily built on causal relationships but rather associations (Breiman, 2001).

While both modeling approaches strive for different goals and hence, should find use in different questions and tasks, it is common that a certain approach simply prevails in certain disciplines, largely ignoring the other one. For example, social sciences, including psychology, have almost exclusively relied on causal-explanatory modeling in their research. This is often rooted in the fact that these disciplines assume that explanation inherently facilitates accurate prediction (Shmueli, 2010). The idea is that by comprehensively cataloging the causes of behaviors, including moderating and mediating variables, one can measure these variables for a group of individuals and predict future behavior accurately.

However, from a statistical point of view, it is important to separate the two modeling approaches of causal-explanatory and predictive modeling as they differ in various forms (Breiman, 2001). Moreover, it cannot be assumed that a model with high explanatory power also encompasses high predictive power (Buchholz and Grote, 2023). This is due to the reason that causal-explanatory models are often built on observational data through association-based models like linear regression. In these cases, the causal relationship of the variables is only established through a theory that imposes such a causal structure. Additionally, the variables of interest in social sciences are often latent, meaning that they must be operationalized to be analyzed. The operationalization is usually also grounded in theory. Yet, this is problematic when the theory does not reflect reality and hence, leads to false operationalization and interpretation of the model. In such cases, the explanatory power of the model cannot be generalized to other populations and should not hold predictive value.

Theoretically, this shortcoming should be battled by increasing the usage of predictive modeling in the social sciences. Instead, as pointed out by Shmueli (2010), predictive modeling has been a rather rarely used approach in the social sciences for a long time. Furthermore, it was even viewed as unscientific due to its more applied usage in non-academic contexts. However, in the past years, with the increasing popularity of machine learning (ML), social sciences have started to use predictive modeling techniques. With this, new research questions and answers were investigated and found, as the two approaches differ in all steps of the statistical modeling process (Sainani, 2014).

In this paper, I would like to review the advances in social sciences through the usage of predictive modeling, taking psychology as an example. Moreover, I will identify in depth the differences between causal-explanatory and predictive modeling in the context of social scientific research. I will illustrate this by discussing how the single steps in the modeling process differ depending on the statistical modeling approach. I will exemplify this with my bachelor thesis, reconstructing each modeling step from the previously predictive approach to a causal-explanatory approach. Then, taking literature as well as my thesis into consideration, I will point out some limitations and problems of predictive modeling that are especially severe in the social science context and show why this approach cannot be applied as easily as in other domains. Finally, I will discuss an integrative approach that combines both modeling techniques.

2 Predictive Modeling in Psychology

Historically speaking, psychology has been mostly concerned with causal-explanatory modeling. Psychologists aim to uncover the causal relationships between psychological variables or traits and behaviors, beliefs, or other outcomes. While randomized and controlled experiments are the gold standard to find such relationships, psychology often time lacks the ability to create experimental data and has to opt for observational data. In order to still draw causal explanations from observational data, a strong theory is needed. While there is a lot of theory that aims to explain psychological mechanisms, the field could not often actually predict future behavior (Yarkoni and Westfall, 2017). However, since artificial intelligence has become more popular in many research fields, also psychological research has begun to utilize it. And with that, predictive modeling has started to emerge as a powerful tool in various domains of psychology.

2.1 Clinical Psychology

In clinical psychology, predictive modeling is used to anticipate patient outcomes, developments of mental disorders as well as individuals' probabilities for a clinical psychological condition (Meehan et al., 2022). Studies such as Schmaal et al. (2015) were able to predict depression trajectories for a period of two years through neural responses to various emotional faces with accuracies of up to 73 %. Through this, chronic patients can be identified and treated according to their depression trajectory. This is an example in which clinicians can be aided in prognostics and creating treatments more fit to their patients' future outcomes.

Similarly, Koutsouleris et al. (2018) were able to predict functional outcomes of people who had a first episode of psychosis with accuracies of over 70 %. In this case, predictive modeling can help to create personalized treatment plans on a patient level as well as enhance resource planning on a more institutional level. Finally, it can aid in identifying more complex associations and patterns that influence functional outcomes.

2.2 Educational Psychology

Also in educational psychology, predictive modeling has become more popular in recent years. In their review, Sghir et al. (2022) have pointed out several important applications of predictive modeling in this domain. First, it can be used to discover aspects related to learning outcomes, which would ultimately lead to the early detection of students with high risks of failure. Additionally, it could help in increasing the students' engagement and satisfaction with the learning process.

Predictive models have been able to analyze and learn the complex relationships between quiz scores, and psycho-social and other education-related variables. Through this, they can predict academic outcomes, such as educational attainment and academic success. For instance, Ahmad et al. (2015) were able to predict the academic success of first-year university students from their demographics, previous academic records, and family background with an accuracy of more than 70 %.

2.3 Personality Psychology

One landmark study in the realm of personality psychology is Youyou et al. (2015), who researched how well personality traits can be predicted through social media footprints such as likes, comments, and posts. They also compared their predictions to the judgments of close others or acquaintances of the participants. They showed that they were able to predict the participants' personalities more accurately than their close ones. However, not only did they outperform the judgments of the close ones but their predictions of the participants' personalities were also more accurate than the participants' judgments about their own personality traits.

Matz et al. (2017) went even one step further and utilized the prediction of personality traits to deploy psychologically tailored advertisements. As a result, advertisements matched to the individual's personality generated up to 40 % more clicks and up to 50 % more purchases compared to mismatched advertisements. While psychological targeting should be administered very carefully as it can be used for digital mass persuasion which is an ethical concern, it should be noted that it is an extremely powerful tool created through predictive modeling.

3 Differences

Having seen different psychological studies utilizing predictive modeling, I will now focus on exploring in depth the differences between predictive and causal-explanatory modeling. For this, each step in the statistical modeling process will be considered, from goal definition to analysis until interpretation. It will become evident that depending on the modeling approach, the procedure in each step differs (Sainani, 2014; Shmueli, 2010). For an overview, please refer to Figure 1. To illustrate the single modeling steps, I will utilize my bachelor thesis, where I have used predictive modeling in the psychological context. To point out the differences, I will highlight what I would have done differently in each step if I had followed a causal-explanatory approach instead.

3.1 Goal Definition

The first step is to define the goal or the research question. In my bachelor thesis, I asked how well I could predict romantic desire. Additionally, I investigated which features are the most important ones in this prediction, meaning that I wanted to find out which features my model relied the most on when predicting the outcome variable for a previously unknown data point.

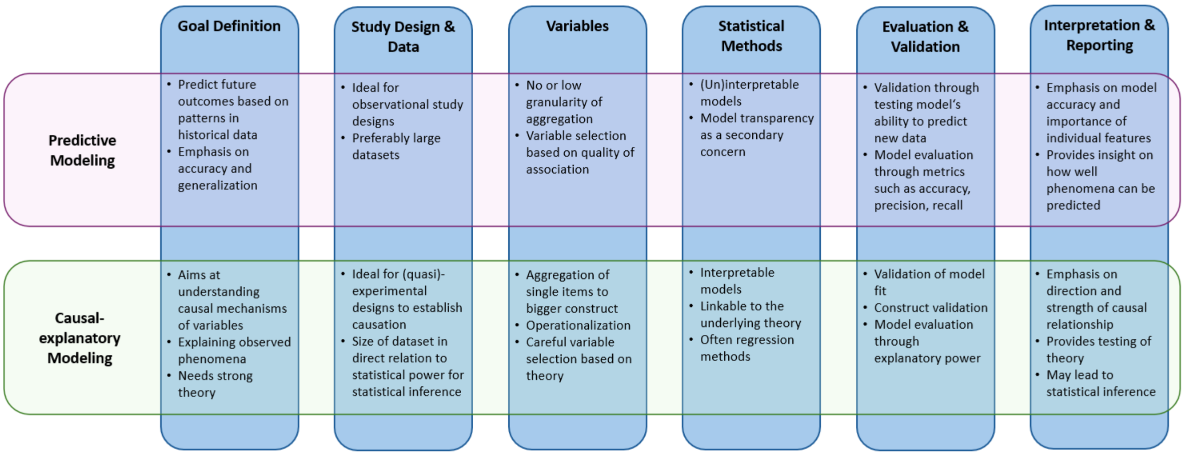

Figure 1: Overview of the statistical modeling process in predictive and causal-explanatory modeling.

If my study had been causal-explanatory, my research question would have differed. First of all, I would have needed a strong theory that already suggests causal relationships between predictor variables and romantic desire. Based on these theories, I would have singled out a smaller amount of predictor variables and would have created psychological as well as statistical hypotheses, which I would either confirm or reject in my later analysis. Generally, my analysis would have aimed at finding evidence for or against the theory by showing which variables have a causal relation with romantic desire.

3.2 Study Design and Data

As I did not collect the data myself but rather used secondary data from a sample of 187 university students, the sample size was not under my influence. However, it should be noted that this particular sample has been used before in causal-explanatory and predictive modeling. Then again, it should also be mentioned that in general, a sample of 187 data points is small for both modeling approaches, but unfortunately often the harsh reality in psychology. It is therefore more than difficult to reach the needed statistical power to use the results for inference in causal-explanatory modeling, but also problematic for predictive modeling as bias and variance may be quite high.

Additionally, the study design is observational which per Shmueli (2010) is preferable for predictive modeling. However, it is also common practice in psychology to use observational data for causal-explanatory modeling as long as an underlying theory establishes a causal relationship between the observed predictor variables and the outcome. In this case, we assume that a strong theory exists that establishes this causal relationship.

3.3 Variables

The feature and outcome variables were collected through self-report questionnaires as well as behavioral measures. The self-report questionnaires were carefully constructed in psychometrical processes. This ensures that they truly measure the underlying constructs or latent variables. Due to this, the variables derived from the self-report questionnaires can also be used for a causal-explanatory approach, as proper operationalization of underlying constructs is given. However, what differs depending on the modeling approach, is the granularity, or level of aggregation. I decided to use the feature variables in my predictive modeling approach on an item level, meaning without aggregation of the items to a bigger psychological scale. This is important because aggregation of feature variables can potentially lead to information loss when non-linear models are used (Seeboth and Mõttus, 2018).

If we now imagine a causal-explanatory approach, it would be common practice to aggregate single items to a bigger scale to gain broader knowledge on constructs that cannot be defined by one single item. At the same time, this would increase interpretability and allow a connection to the underlying theories which usually work with constructs and not single items.

Furthermore, a different number of variables would be considered. While I used in my thesis a total of 269 features to predict romantic desire, the number would drastically decrease in causal-explanatory modeling because of (a) aggregation, and (b) variable selection based on theory. The latter point means that I would only research the causal relationship of a fraction of the variables in the dataset. Specifically, those variables would be the ones named in the underlying theory, as only for them a theoretically established causal relationship with the outcome variable exists. The additional variables may be considered in the analysis as confounding variables, however, their actual relationship with the outcome variable would be left unexplored.

3.4 Statistical Methods

In my thesis, I used an ensemble ML method, namely random forest regressors. Random forests have two key advantages that are very useful for my research question. First of all, random forests can handle a huge number of predictor variables which was needed since I did not rely on a specific theory but rather wanted to explore how well romantic desire could be predicted at all. Second, random forests are non-parametric, so complex relationships can be identified because the model does not impose a particular structure on the data (Joel et al., 2017).

A switch to a causal-explanatory modeling approach would have brought quite many differences in this step. First of all, the model of choice would have likely been a linear regression as it is so often in psychology. The advantages of the random forest would have vanished: linear regression would impose a clear relationship between the variables. However, ultimately this is also what causal-explanatory modeling aims at: causal relationships to confirm or disprove our underlying theory. In addition, the assumption that linear regression makes would have needed to be tested: multicollinearity, heteroskedasticity, normality, etc. Given that none of these assumptions are violated, the linear regression could have been performed, leaping coefficients for the predictor variables.

3.5 Evaluation and Validation

Depending on the modeling approach, evaluation and validation differ. In my thesis, I cross-validated the random forest model while simultaneously tuning its hyperparameters. Through this, I found that the predictive power varies quite strongly over the single folds. This limited the generalizability of my findings, as the predictive performance was not as strong and consistent in the cross-validation.

If I had used a causal-explanatory approach instead, I would have checked for reliability and validity of my predictor variables to ensure that my model uses adequately operationalized variables that fit the underlying theory. Moreover, I would have applied goodness-of-fit tests to confirm that my model fits the data. Finally, I would have evaluated the explanatory power of the model by checking the statistical significance of the predictor variables, which would then influence the interpretation of the results.

3.6 Interpretation and Reporting

The final results would have drastically differed. My thesis gave insight into how well romantic desire could be predicted on previously new and unseen data points. A causal-explanatory modeling approach, on the other hand, would have shown which variables cause romantic desire but the model would most likely not have had much predictive value due to limited sample size and statistical inference. Again, this highlights the strong differences between the two approaches. While causal-explanatory modeling aims at uncovering the predictor variables causing a certain effect retrospectively, predictive modeling focuses on how well we can predict the effect prospectively.

4 Limitations of Predictive Modeling in Psychology

Having pointed out examples of predictive modeling in psychology and having contrasted them with causal-explanatory modeling, I would now like to point out some limitations of predictive modeling that can make its usage in certain domains of social sciences problematic. As defined by Buchholz and Grote (2023), I would like to discuss some success conditions for applying predictive modeling in social science and then discuss why these conditions may not always be met in psychology.

4.1 Theory-dependent Variables

First of all, the prediction task has to be well-defined in such a way that the outcome variable that is being predicted and the latent construct that ought to be predicted are virtually the same (Buchholz and Grote, 2023). In general, the outcome variables that are interesting for psychology are big latent constructs that are often operationalized through theory. The whole process of operationalization is a research field in itself and is not as intuitive as in fields like medicine, physics, or biology.

In many domains of psychology, it is therefore impossible to use predictive modeling without at least "a bit" of theory. This becomes clear when taking again Youyou et al. (2015) at hand, where ML algorithms predicted personality traits, to be more specific, the so-called *Big 5*. The Big 5 are the theoretic standard model of personality research and are therefore often used when assessing personality. It becomes obvious that it is imperative to have a theoretical model that operationalizes the large concept of personality into predictable outcome variables, even if we want to use a predictive model.

I therefore argue that in many domains of psychology, even data-driven predictive modeling approaches need theory-driven operationalizations. This implies two problems (1) it may be in most cases impossible to find the exact manifest correlate to the latent construct that is being researched, and (2) predictive modeling in most domains of psychology needs theory, unlike in many other fields.

4.2 Sample Size

Additionally, predictive modeling needs a large set of high-quality training data (Buchholz and Grote, 2023). As predictive modeling aims to generalize to new and unseen data, a sufficiently large training sample is needed to accurately represent the underlying population. However, compared to other sciences, psychological studies often only provide a very limited sample size. This may be due to resource constraints as studies are often time-consuming and expensive but also due to the complexity of human subjects and specialized participant criteria that are often applied.

Altogether this means that predictive modeling is usually carried out on much smaller sample sizes than in other sciences. Ultimately, this leads to a problem that psychology has been fighting with for a long time: the issue of replicating results on other samples. In causal-explanatory modeling, this issue becomes clear when assessing statistical power and effect sizes. In predictive modeling, for example, cross-validation would uncover limited generalizability that likely also stems from the small sample size. Taking my bachelor thesis as an example again, the cross-validation uncovered a large variance among the single folds' performances. While sample size is an important factor in both approaches, it is even more important in predictive modeling as the model is built from the data, which requires a larger amount of data to achieve lower bias and variance (Shmueli, 2010). It can therefore be concluded, that predictive modeling will face even larger problems than causal-explanatory modeling due to psychology's rather small sample sizes.

4.3 Unreliable Data

Furthermore, it is important to consider the predictor variables that are being used in psychology. Most variables in psychology are latent, as the usage of psycho-social predictor variables is much more common than the usage of genetic predictor variables. Latent variables cannot be measured directly but have to be operationalized in some way, which often leads to measurement error (Stachl et al., 2020). While the effect of unreliable variables due to measurement errors is quite well understood for linear models in causal-explanatory psychological research(Lord et al., 1968), the effect of measurement errors is often not enough assessed in predictive psychological research that often uses non-linear ML models.

Jacobucci and Grimm (2020) manipulated the reliability of a predictor variable and then fed it to linear and non-linear predictive models. While the non-linear models performed better on perfectly reliable data, their performance saw a more drastic drop when the data was rather unreliable. The linear model, on the other hand, was more stable but still decreased in performance with more unreliable predictors. However, the ranking of performance changed so far that the linear model outperformed the non-linear one on more unreliable data. This result persisted, even if the authors formulated non-linear relationships in the data. They conclude that with unreliably measured data, non-linear models are unable to capture those relationships and cannot predict the outcome variable accurately.

What has to be highlighted is that unfortunately psychology often has unreliably measured variables and even psychological scales tested for their reliability can be unreliable if they are used in the wrong way. However, this is a major limitation of predictive modeling in psychology as more complex non-linear relationships that could be uncovered through non-linear ML models, still fail to surface due to unreliable variables. We therefore have to consider this shortcoming when using predictive modeling in psychology and, if possible, ensure the reliable measurement of our variables and, if not possible, acknowledge that these methods may simply not be a good fit for our domain or research question.

5 Integrative Approach

Finally, I would like to discuss an integrative approach that uses both causal-explanatory and predictive modeling to maximize the utility. What Hofman et al. (2021) described as particularly promising is the iterative alternation between predictive and causal-explanatory modeling. Through this, important associations can be identified, new theories constructed and theoretically established causal relations tested.

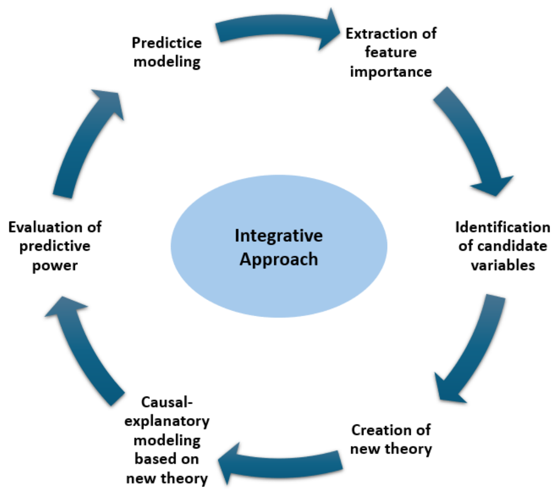

Figure 2: Illustration of an integrative approach featuring both predictive and causal-explanatory modeling.

The proposed idea will again be illustrated with the example of my bachelor thesis. For a graphical illustration, please refer to Figure 2. One additional step that I took in my predictive modeling approach was the extraction of permutation feature importance across the outer folds in the nested cross-validation. Through this, I wanted to gain at least some insight into which features the random forest relied most heavily on in its prediction. In an integrative approach, a first predictive model like mine showed (1) how well romantic desire can be predicted with various predictor variables and (2) which predictor variables the model relied most heavily on. Through the latter point, it is possible to identify potential candidate variables and create a theory, ideally also by consulting existing literature, that establishes a causal relationship between the candidate variables and the outcome variable.

For example, the features with the highest feature importance for the male romantic partner desire, meaning how attractive a man was perceived on average by his speed-dates, were physical attractiveness, flirting behavior, rejection sensitivity, and vitality. However, the predictive model was unable to point out any causal relation between these features and romantic desire but merely noted an association. Moreover, permutation feature importance is unable to reveal the direction of the association, meaning that for example, it is unknown whether flirting behavior is positively or negatively associated with romantic desire. This is exactly where the integrative approach comes into action. Out of the exploratory predictive model, a theory can be formed about how these features cause an effect in a certain direction in romantic desire. An overly simplified theory could state that the named four features have a causal effect on male romantic partner desire, with higher physical attractiveness, more flirting behavior, and higher vitality increasing romantic desire linearly, while a higher rejection sensitivity decreases romantic desire due to such individuals being perceived as less secure.

Based on the newly formulated theory, I would create psychological and statistical hypotheses for each predictor variable. Next, the hypotheses would be tested through a causal-explanatory modeling approach, such as linear regression. Through this analysis, it would hopefully become clear whether a significant causal relationship between romantic desire and the predictor variables exists or not. In an iterative process, the model can be adjusted, for example by excluding non-significant predictor variables. Finally, the model's explanatory power should be evaluated and it should be checked if the properties of the underlying population can be inferred from this specific sample. In addition, the predictive properties of the final model can be tested. For this, previously unseen data would be used to see how accurately the linear regression can predict the outcome variable on new data.

6 Conclusion

This paper aimed to investigate the two modeling approaches predictive and causal-explanatory modeling in the context of psychology, taking it as an example for all social sciences. It was highlighted that even though social sciences have been known for mostly using causal-explanatory modeling, in recent years, they have drastically developed their research by also incorporating predictive modeling. The evolution of predictive modeling in psychology was illustrated through examples in personality psychology, clinical psychology, and educational psychology. It was shown that psychologists were able to make accurate predictions and uncover patterns that were not easily discernible through causal-explanatory methods.

To further point out how drastic this change is, I illustrated how each step of the statistical modeling process has to be adapted when changing the modeling approach. From this it can be inferred that a different modeling approach serves an entirely different research question, asks different variables in the modeling process and needs different methods of evaluation and validation. While psychological research traditionally asked questions regarding the causal nature of variables, it has become more interested in predicting outcomes. What needs to be clear is, that both approaches serve distinct types of questions and one is not able to answer the other's questions. Therefore, both are needed in social science research and cannot be replaced by the other one.

However, the rise of predictive modeling in the social sciences has not only brought new possibilities but also challenges. To discuss the latter, the paper outlined several limitations and challenges associated with the usage of predictive modeling in psychology. Issues such as theory-dependent variables, small sample sizes, and unreliable data were discussed as factors that may decrease the effectiveness of predictive modeling in certain psychological domains and also other social sciences. In light of these considerations, it is suggested that psychologists should take a balanced and thoughtful approach when it comes to modeling. I encourage researchers to carefully weigh the benefits of predictive modeling in gaining new insights, while also acknowledging the difficulties that come with its application.

Furthermore, the paper proposed an integrative approach utilizing both predictive and causal-explanatory modeling to use the benefits of both of them. That is, predictive modeling can be used to explore associations between the predictor and outcome variables, and with that identify possibly important variables for the prediction. Based on this, new theories can be created or added to existing ones. This new theory can then be tested through a causal-explanatory model. Finally, not only the explanatory but also the predictive power of that model can be assessed. Overall, the integrative approach should contribute to a more comprehensive and detailed understanding of complex psychological phenomena.

Ultimately, the driving force behind all research, regardless if it is predictive or causal-explanatory, is the pursuit of scientific advancement. With this goal in mind, researchers should keep the differences between the two approaches in mind while utilizing both of them when appropriate or even finding combined solutions. Through this, researchers are well equipped to advance our knowledge in different scientific fields.

References

Ahmad, F., Ismail, N. H., and Aziz, A. A. (2015). The prediction of students' academic performance using classification data mining techniques. *Applied Mathematical Sciences*, 9:6415–6426.

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3).

Buchholz, O. and Grote, T. (2023). Predicting and explaining with machine learning models: Social science as a touchstone. *Studies in History and Philosophy of Science Part A*, 102(C):60–69.

Hofman, J. M., Watts, D. J., Athey, S., Garip, F., Griffiths, T. L., Kleinberg, J., Margetts, H., Mullainathan, S., Salganik, M. J., Vazire, S., and et al. (2021). Integrating explanation and prediction in computational social science. *Nature*, 595(7866):181–188.

Jacobucci, R. and Grimm, K. J. (2020). Machine learning and psychological research: The unexplored effect of measurement. *Perspectives on Psychological Science*, 15(3):809–816.

Joel, S., Eastwick, P. W., and Finkel, E. J. (2017). Is romantic desire predictable? machine learning applied to initial romantic attraction. *Psychological Science*, 28(10):1478–1489.

Koutsouleris, N., Kambeitz-Ilankovic, L., Ruhrmann, S., Rosen, M., Ruef, A., Dwyer, D. B., Paolini, M., Chisholm, K., Kambeitz, J., Haidl, T., and et al. (2018). Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression. *JAMA Psychiatry*, 75(11):1156.

Lord, F. M., Novick, M. R., and Birnbaum, A. (1968). *Statistical theories of mental test scores*. Addison-Wesley.

Matz, S. C., Kosinski, M., Nave, G., and Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences*, 114(48):12714–12719.

Meehan, A. J., Lewis, S. J., Fazel, S., Fusar-Poli, P., Steyerberg, E. W., Stahl, D., and Danese, A. (2022). Clinical prediction models in psychiatry: A systematic review of two decades of progress and challenges. *Molecular Psychiatry*, 27(6):2700–2708.

Sainani, K. L. (2014). Explanatory versus predictive modeling. *PMR*, 6(9):841–844.

Schmaal, L., Marquand, A. F., Rhebergen, D., van Tol, M.-J., Ruhé, H. G., van der Wee, N. J., Veltman, D. J., and Penninx, B. W. (2015). Predicting the naturalistic course of major depressive disorder using clinical and multimodal neuroimaging information: A multivariate pattern recognition study. *Biological Psychiatry*, 78(4):278–286.

Seeboth, A. and Mõttus, R. (2018). Successful explanations start with accurate descriptions: Questionnaire items as personality markers for more accurate predictions. *European Journal of Personality*, 32(3):186–201.

Sghir, N., Adadi, A., and Lahmer, M. (2022). Recent advances in predictive learning analytics: A decade systematic review (2012–2022). *Education and Information Technologies*, 28(7):8299–8333.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25(3).

Stachl, C., Au, Q., Schoedel, R., Gosling, S. D., Harari, G. M., Buschek, D., Völkel, S. T., Schuwerk, T., Oldemeier, M., Ullmann, T., and et al. (2020). Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences*, 117(30):17680–17687.

Yarkoni, T. and Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, 12(6):1100–1122.

Youyou, W., Kosinski, M., and Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4):1036–1040.