Statistical foundations of ecological rationality

Henry Brighton

Abstract

If we reassess the rationality question under the assumption that the uncertainty of the natural world is largely unquantifiable, where do we end up? In this article the author argues that we arrive at a statistical, normative, and cognitive theory of ecological rationality. The main casualty of this rebuilding process is optimality. Once we view optimality as a formal implication of quantified uncertainty rather than an ecologically meaningful objective, the rationality question shifts from being axiomatic/probabilistic in nature to being algorithmic/predictive in nature. These distinct views on rationality mirror fundamental and long-standing divisions in statistics

(Published in Special Issue Bio-psycho-social foundations of macroeconomics)

Authors

Henry Brighton, Department of Cognitive Science and Artificial Intelligence, Tilburg University, The Netherlands; Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Berlin, Germany, brighton@uvt.nl

1   Introduction

> How do human beings reason when the conditions for rationality postulated
> by the model of neoclassical economics are not met – for example, when no
> one can define the appropriate utility function? (Simon, 1989, p. 377)

> It is sometimes *more* rational to admit that one does not have sufficient infor-
> mation for probabilistic beliefs than to pretend that one does. (Gilboa et al.,
> 2012, p. 28)

If we reassess the rationality question under the assumption that the uncertainty of
the natural world is largely unquantifiable, where do we end up? Organisms rely on an
ability to make accurate inferences from limited observations of complex, uncertain, and
unstable environments, and an answer to the rationality question should formulate both
the nature of this problem and its solution. The orthodox answer is that the nature of the
problem is probabilistic inference and the nature of the solution is probabilistic optimality.
If we approach the rationality question with this view in mind then the prospects appear
slim for a formally viable, normative theory of rational action when probabilities can't
be quantified. In this article I argue that this view neglects alternative formulations of
the problem of statistical inference, and the statistical theory of ecological rationality in
particular (Brighton and Gigerenzer, 2007; Gigerenzer and Brighton, 2009; Brighton and
Gigerenzer, 2012; Brighton, 2018). By setting out the statistical foundations of ecological
rationality, my argument will undermine the claim that bounded and ecological rational-
ity offer no normative challenge to orthodox rationality. The principle casualty in this
statistical reassessment of the rationality question is optimality. Once we view optimality
as a formal implication of quantified uncertainty rather than an ecologically meaningful
objective, the rationality question shifts from being axiomatic/probabilistic in nature to
being algorithmic/predictive in nature. In short, by reassessing the rationality question
in this way, we end up not with a revised solution to the same statistical problem but a
theory of rationality that responds to a different statistical problem.

I will first argue that the implications of deepened uncertainty are often obscured by a
cluster of commonly held statistical intuitions collectively termed the bias bias (Brighton
and Gigerenzer, 2015). The bias bias leads many researchers in several disciplines to
neglect the relationship between two critical and controllable components of prediction
error, bias and variance. Bias is widely understood, at least intuitively, and reflects the
ability of a model to accurately capture systematic regularities in observations. The
variance component of prediction error reflects the sensitivity of a model's predictions

to different observations of the same problem, such as a different sample from the same population. These two components additively contribute to expected prediction error (O'Sullivan, 1986; Geman et al., 1992; Hastie et al., 2001; Bishop, 2006). Unlike bias, the role of variance is less intuitive and often neglected, causing a bias bias in statistical thinking that simplifies the role of uncertainty, masks a wide range of predictive models from consideration, and leads to the development of questionable theories and poorly justified policies. To illustrate the bias bias I will examine examples of social and economic systems that pose problems without optimal solutions.

Once the statistical pathologies associated with the bias bias have been clarified, I argue that the bias bias also characterizes the orthodox formulation of rational decision making under uncertainty. This second stage of my argument appeals to Breiman's (2001) distinction between two cultures of statistical modeling known as data modeling and algorithmic modeling. Data modeling characterizes much of traditional statistical inquiry and proceeds by conjecturing a data generating model. Algorithmic modeling proceeds under the assumption that the data generating model is indeterminable or non-existent, and does so by analyzing the relative ability of competing learning algorithms to incur low prediction error. This distinction echoes deeper distinctions in statistics and information theory (Geisser, 1993; Rissanen, 2007; Shmueli, 2010). I then argue that these distinct modes of statistical inquiry map directly onto the practices of orthodox rationality and ecological rationality, and, crucially, justify different kinds of rationality claim (Brighton, 2018). Claims of ecological rationality refer to the ability cognitive algorithms, such as simple heuristics, to incur low prediction error relative to alternative algorithms. Such claims neither imply optimal functioning, nor do they require that an optimal response be determinable in order to be established, explained, or justified.

The strategy of extending the categories of uncertainty assumed when formulating the rationality question, and then examining how robust established rationality principles are to these revised uncertainty conditions, is familiar but neglected (Binmore, 2009). A central concern for Savage (1954) was the category small worlds that he assumed when formulating the foundations of Bayesian decision theory. Small worlds are those that can modeled using a decision matrix defining a mutually exclusive and exhaustive set of states of the world, consequences, and actions that map between them. If we assume, as many do, that "the worlds of macroeconomics and high finance certainly don't fall into this category" (Binmore, 2009, p. 2) then this raises the question of which categories these worlds do fall into. In this article I focus on problems characterized not only by unmeasurable uncertainty (Knight, 1921; Kozyreva and Hertwig, 2019) or probabilistic ambiguity under the subjective interpretation of probability (e.g., Gilboa and Schmeidler, 1989; Machina and

Schmeidler, 1992; Epstein, 1999), but by unquantifiable uncertainty under any definition of probability, and to the extent that it is more rational to regard generating distributions as a non-existent than to view them as the foundation for formulating the rationality question. As the following example illustrates, worlds of this kind are not hard to find.

2   Bias, Variance, and the Bias/Variance Dilemma

Consider a serial offender who has committed a series of crimes while living at a single home location. Given only the locations of these crimes, how accurately can the location of the offender's home be predicted? I will use this well-studied problem in geographical criminal profiling to illustrate the relationship between bias, variance, and prediction error. Figure 1(a) depicts a map covering an area of roughly 30km^2 of Baltimore county, Maryland, USA. Superimposed on this map are 15 blue circles identifying the locations of a series of burglaries committed by a serial burglar residing at a single address[1]. In additional to these locations, Figure 1(b) plots the true home location of the offender and the home locations predicted by two geographical profiling models. The first model is the centroid method that predicts the "center of gravity" of the crime locations, which is simply the mean x-coordinate and mean y-coordinate of the observed crimes. The second model computes a probability surface over the crime area using an exponential decay function. Given a crime area divided into an array of cells, this model estimates the probability that each cell contains the home address of the offender. As the distance d from a crime location increases, the exponential decay function models the finding that the probability that offenders commit crimes a distance d from their homes decreases as a negative exponential function of d. By integrating the probabilities calculated from each crime location, the model estimates a probability surface over the crime area, shown in Figure 1(c). For this offender, the cell with greatest probability predicts the true home location with nearly zero error. The centroid method, on the other hand, predicts a point roughly 1.5km north-west of the true home location.

Is the probability surface model more accurate in general, or might this near perfect prediction be a lucky guess? One way of addressing this question is to evaluate the two models over a range of serial criminals (e.g., Block and Bernasco, 2009; Leitner and Kent, 2009; Levine and Block, 2011). In addition, further insights into this specific example can be gained by deepening the uncertainty conditions under which predictions are made. For example, how would the two models perform when, rather than predicting the home

---

[1]The data for this offender, labeled TS15D, can be found in the dataset supplied as part of the CrimeStat (Levine, 2010) package available from `https://www.icpsr.umich.edu/CrimeStat/`.

location of the offender from all 15 crime locations, they only had 5 locations to work with? Figure 1(d) examines this question by plotting the prediction error of the two models as a function of the number of crimes sampled. Specifically, r crimes are repeatedly sampled without replacement from the complete series of 15 crimes. For each such sample we
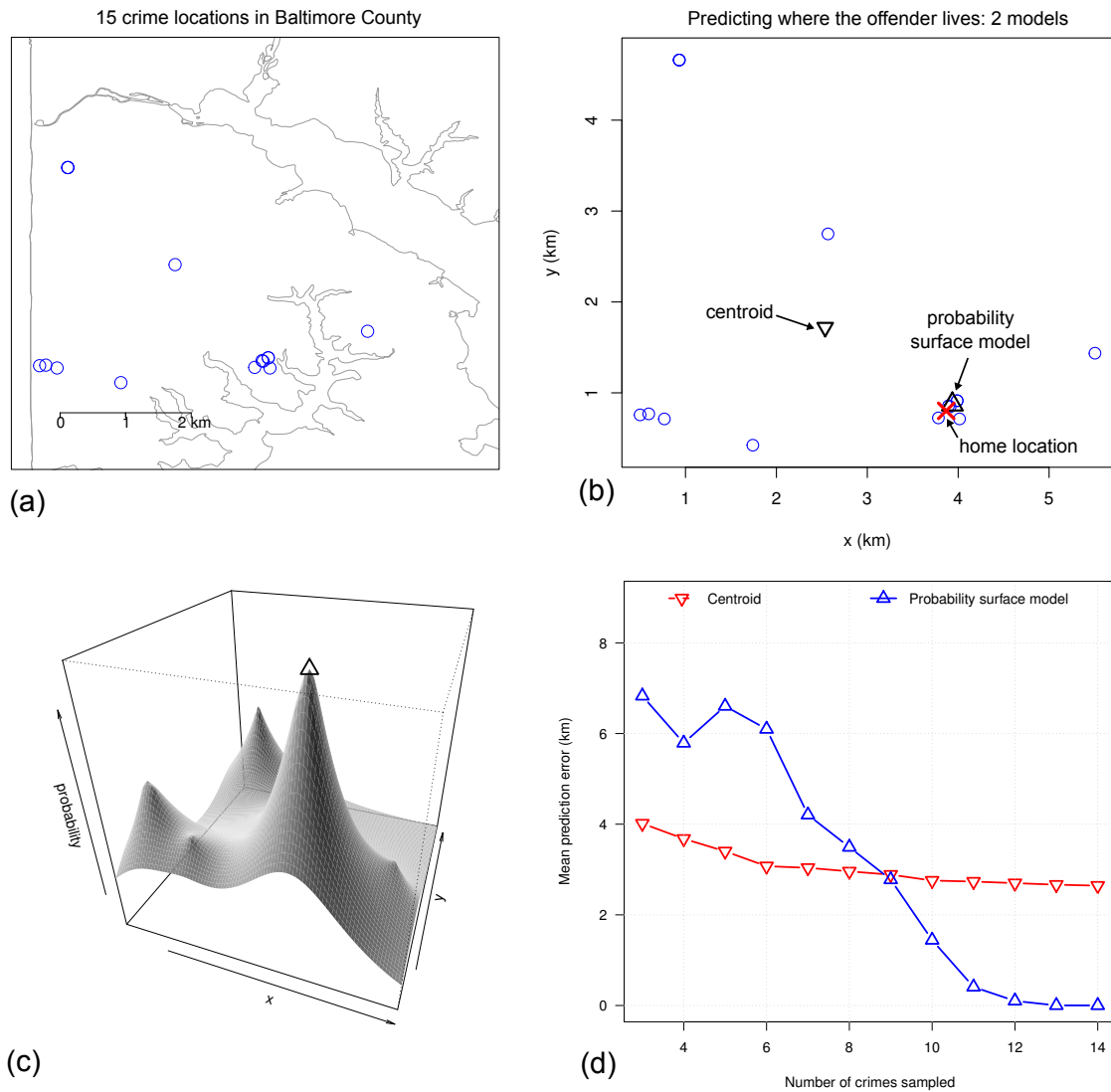
Figure 1 Panel (a) plots the locations of 15 burglaries committed by a single offender in Baltimore County. Panel (b) plots the true home location of the offender and the home locations predicted by the centroid method and a probability surface model with an exponential decay function depicted in (c). Panel (d) plots the prediction error of the two models as a function of the number of crimes sampled. The relative superiority of the two models inverts after 9 crimes have been observed. Note: Crime locations visually overlap in 5 cases.

measure the prediction error of both models, and then report the mean prediction error of both models as a function of r. What is striking about Figure 1(d) is that when fewer than 9 crimes are observed the centroid method outperforms the probability surface model. In other words, when the uncertainty conditions are deepened the relative performance of the two models inverts. Our next step is to understand how the concepts of bias and variance help to explain why.

2.1   Decomposing Prediction Error Into Bias and Variance

Suppose we are given n crime locations $\mathbf{c}_1, \ldots, \mathbf{c}_n$ where each crime location $\mathbf{c}_i = (x_i, y_i)$ is a point in a compact 2-dimensional Euclidean space covering the crime area. For simplicity I will assume that the offender's home location $\mathbf{h}$ also lies within the crime area. The task of the two models is to then map a series of crime locations to a point prediction of the offender's home location. Because the area that law enforcement may need to search before locating the offender's home will grow as a squared function of the prediction error, I will consider squared loss, and specifically, the squared Euclidean distance between the predicted and true home location, $||\mathbf{h} - \mathbf{f}||^2$. Now, at a given sample size r, analyses of the kind shown in Figure 1(d) sample r crimes k times to yield an ensemble of k predictions denoted $\mathbf{f}^{(1)}, \ldots, \mathbf{f}^{(k)}$. For a given sample size the expected error of a model can be decomposed as follows:

$$\text{Expected error} = (\text{bias})^2 + \text{variance} + \text{noise}. \tag{1}$$

The derivation of this decomposition for squared loss can be found in most machine learning textbooks (e.g., Duda et al., 2001; Hastie et al., 2001; Bishop, 2006) as well as a landmark article by Geman et al. (1992). For this problem the bias is the distance between the mean prediction of the ensemble defined above and true location of offender's residence. The mean prediction of the ensemble is simply the centroid of $\mathbf{f}^{(1)}, \ldots, \mathbf{f}^{(k)}$, denoted $\bar{\mathbf{f}}$. Thus, bias is given by:

$$(\text{bias})^2 = ||\bar{\mathbf{f}} - \mathbf{h}||^2. \tag{2}$$

Figure 2(a) provides a visual illustration of the relationship between an ensemble of 5 example predictions, the true home location, and bias. Variance is then the degree to which the individual model predictions – the members of the ensemble – vary about the mean prediction $\bar{\mathbf{f}}$:

$$\text{variance} = \frac{1}{k} \sum_{i=1}^{k} ||\bar{\mathbf{f}} - \mathbf{f}^{(i)}||^2. \tag{3}$$

Figure 2(b) illustrates the relationship between the same ensemble of 5 predictions above, the mean prediction $\bar{\mathbf{f}}$, and variance. Finally, the noise term in Equation 1 plays no role in this example, but could represent the additional and irreducible error we would incur if, say, an adversary added some normally distributed error to our predictions.

2.2 Using Bias and Variance to Analyze and Explain Relative Prediction Error

To better understand the relative performance of the two models shown in Figure 1(d), Figure 2(c) decomposes their prediction errors into bias and variance. Focusing on bias first we see that the centroid model suffers from high bias across all sample sizes because the assumption that the offender's home lies at the centroid of the crime locations fails to hold for this offender. The probability surface model also has high bias at low sample sizes but its bias steadily decreases as the sample size increases, eventually approaching zero. If bias were the only concern then the probability surface model is clearly superior. Turning to variance, we see that the "trick" behind the centroid method is that it incurs remarkably low variance until roughly 14 crimes are observed, at which point the variance of both models approaches zero. The key insight here is that although the probability surface model is unbiased and makes a near perfect prediction, this is only true under complete information. When uncertainty is increased due to fewer crimes being observed, the probability surface model suffers from high variance, the upshot being that the biased, low-variance model centroid method incurs lower total prediction error. Thus, the ability of a model to match, represent, or capture the underlying structure of the problem – its potential to incur low bias – is only one determinant of low prediction error. How robust the predictions of the model are to different realizations of the problem – its potential to incur low variance – will often prove critical.

2.3 The Bias/Variance Dilemma

There will always be an infinite number of explanations that are consistent with our observations. Models code the assumptions needed to select which explanations we entertain in their functional form, their parameters, and in the constraints they impose on what values these parameters can be assigned. Assumptions are like bets. For example, we could hedge our bets by deploying a nonparametric model such as a multilayer neural network capable of representing a wide range of systematic patterns. The hope is that the network will incur low bias, but due to the large number of parameters needed to model layers of the network, we run the risk of incurring high variance. Another approach is
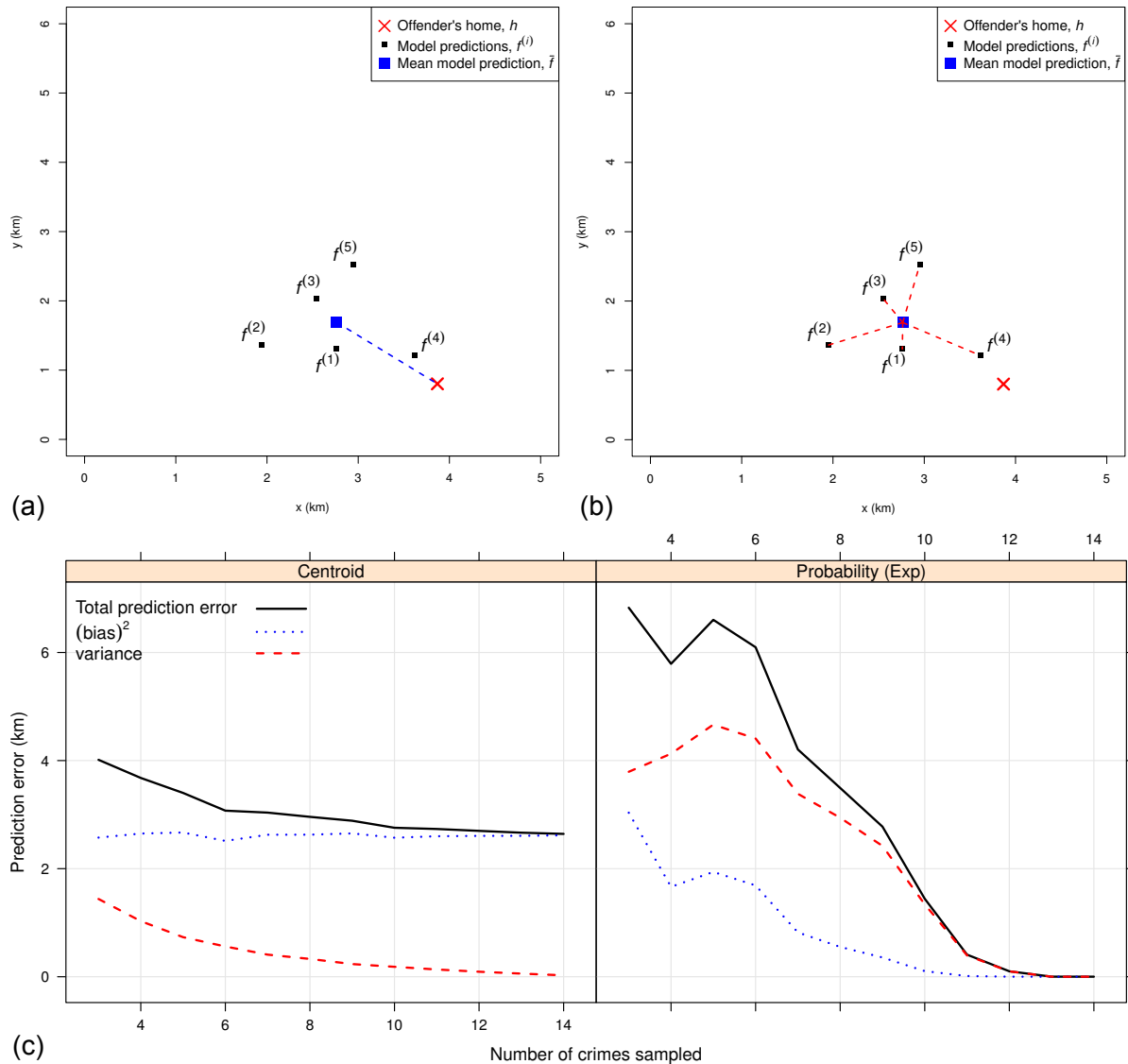
Figure 2 Panels (a) and (b) illustrate the calculation of bias and variance for an ensemble of 5 predictions, $\mathbf{f}^{(1)}, \ldots, \mathbf{f}^{(5)}$. Bias, shown in (a), is the difference between the mean prediction of the ensemble, $\bar{\mathbf{f}}$, shown as a blue square, and the true home location of the offender, $\mathbf{h}$, shown using a red cross. Variance, shown in (b), is variation among the individual members of ensemble relative to the mean prediction. Panel (c) decomposes prediction errors of the model comparison shown in Figure 1(d) into bias and variance.

to deploy a parametric method such as a linear model in the hope that the systematic patterns in the observations are governed by a linear function, or something close. This risk might be worth taking, given that the fewer parameters needed to specify the linear model could keep variance within acceptable bounds. An even bolder approach is to deploy a parameterless model with no free parameters. Here, we ignore the observations

completely to guarantee zero variance. Unless this bold conjecture turns to be correct, or close to correct, the problem is that the model will incur high bias.

These issues clarify that whenever uncertainty surrounds our choice of model, we face a bias/variance dilemma (Geman et al., 1992). The dilemma arises because techniques for reducing variance tend to increase bias, and techniques for reducing bias tend to increase variance. As we have seen, the number of observations also play critical role in determining the bias and variance incurred by a model. All told, predictive inference involves complex interactions between models, observations, the data generating process, and prediction error. These complexities are often glossed over, and the bias bias explains to some extent why.

3   The Bias Bias

To suffer from the bias bias is to develop, deploy, or prefer models that are likely to achieve low bias, while simultaneously paying little or no attention to models with low variance (Brighton and Gigerenzer, 2015). The bias bias manifests itself in a range of statistical intuitions and practices which over-simplify or ignore the complexities of the bias/variance dilemma. In this sense the bias bias approximates the problem of statistical inference. This approximation is justified when the data generating machinery is known with a high degree of certainty, or when the number of observations asymptotes. The following four examples illustrate the dangers of the bias bias when dealing with the uncertainty of the natural world, and social and economic systems in particular.

3.1   Four Examples of the Bias Bias

Example 1: Geographical Criminal Profiling.   A guiding concern for geographical criminal profiling is achieving low prediction error, and the field progresses in large part through the competitive testing of diverse models ranging from the centroid method to sophisticated probabilistic models (e.g., Rossmo, 1999; Canter et al., 2000; Snook et al., 2005; Levine and Block, 2011). This seems like an unlikely field to harbor a bias bias, yet the bias/variance perspective appears to be unfamiliar to researchers in this area[2]. Unlike the analysis above, I have failed to find any studies that decompose prediction error into bias and variance or systematically examine, for a specific offender, prediction error as a function of the number of crimes sampled. While it is typical for studies to

---

[2]Levine (2009, pp. 177-181) discusses the related concepts of accuracy and precision, and argues that both should considered in evaluating models, but doesn't draw a connection to the bias and variance, or the bias/variance dilemma.

compare predictive models over a range of serial offenders who have committed varying numbers of crimes, this can only offer a limited insight into the potential of low variance models. Reversals in relative model performance, like we see in Figure 1(d), are masked by standard practices of model evaluation, and this is likely to steer model development away from techniques that incur low variance under conditions of heightened uncertainty.

Failure to explicitly consider and analyze bias and variance does not necessarily imply a bias bias, although it often goes hand in hand with a focus on bias reduction as the driving concern. This is reflected in the tendency of recent work to use increasingly complex probabilistic methods to model the factors thought to drive offender behavior (Block and Bernasco, 2009; Leitner and Kent, 2009; Levine and Block, 2011). I use the terms "simple" and "complex" here in a non-technical sense to refer to the degree to which one attempts to model the geographical, psychological, social, and economic factors likely to drive where and how many offenses an individual commits. In short, there is an argument to made that the field of geographical criminal profiling suffers from a mild bias bias, principally because the concept of variance is unfamiliar, and common techniques for reducing variance appear to play an increasingly minor role in model development. On the other hand, this field scores highly on model diversity and there is a broad recognition that simple models are often hard to improve on. As the following example illustrates, the influence of the bias bias is often far stronger.

Example 2: Regulating International Banks. How can financial crises be avoided? Regulating international banks is one option, but what rules should the regulators impose? If systematic regularities in past bank failures can be identified then these regularities should help to predict future failures, and then the problem can viewed, at least in part, as a prediction problem. Focusing on the Basel accord, Haldane and Madouros (2012) point out that the regulatory rules introduced by Basel I in 1988, and their subsequent development into Basel III introduced in 2010, have undergone a striking increase in complexity. In numbers, the Basel I agreement was 30 pages long and required banks to calculate only five risk weights, while the post-2008 financial crisis Basel III was 616 pages long and involved calculating several million risk weights. Has this increase in complexity made international banks more resilient and financial crises less likely? Based on a sample of 8,500 US banks, 442 of which entered receivership, Haldane and Madouros found that all five CAMELS indicators – a standard set of indicators used to summarize the condition of a bank – contributed to the in-sample fit of a linear model used to explain bank failure. Turning to the more challenging problem of out-of-sample prediction, they then estimated

the predictive accuracy of the full CAMELS model, a number of single indicator models, and constant model.

With model parameters estimated from small samples of banks, and hence conditions of increased uncertainty, this test of out-of-sample predictive accuracy revealed that a single liquidity indicator model achieved the lowest prediction error, outperforming the full CAMELS model over a range of uncertainty conditions. These results follow a similar pattern to the criminal profiling example. Models that integrate several factors that should, intuitively, improve predictive accuracy can in fact reduce predictive accuracy under conditions of high uncertainty. Appealing to further analyses, and the concepts of bias and variance, Haldane and Madouros argue that the regulators have mistakenly assumed that complex problems like regulating an international banking system require complex solutions. This strongly implicates a bias bias in the intuitions of the architects of Basel III, and the benefits of simple, low variance models found by Haldane and Madouros support this conclusion (see also Aikman et al., 2014). These architects appear to have overlooked basic principles of out-of-sample prediction error and competitive model testing, and consequently overlooked effective responses to the bias/variance dilemma.

Example 3: The Mean-Variance Portfolio. From banks to individuals, economic agents frequently face the problem of investing money into N funds. Harry Markowitz's Nobel prizing winning contributions to portfolio theory included the mean-variance model, an optimal solution to this problem (Markowitz, 1959). An overarching concern for this discussion is the nature of the assumptions needed to formulate optimality results. Under conditions of unquantifiable uncertainty these assumptions are likely to be violated, the upshot being that the status of "optimal" solutions must be relegated to "just another model." DeMiguel et al. (2009) shed some light on this issue by comparing the performance of 15 models on seven investment problems. As well as the mean-variance model they considered a range of probabilistic models and a simple heuristic, $1/N$, which allocates money equally to the N funds. The task was to predict the performance of each portfolio in the following month based on parameters estimated from the preceding 10 years of stock data. Sliding this 10-year window forward one month at a time, DeMiguel et al. compared the accumulated performance of the 15 models and found the $1/N$ heuristic ranked first on certainty equivalent returns, second on turnover, and fifth on the Sharpe ratio. Notice that $1/N$ embodies an extreme response to the bias/variance dilemma. Its lack of free parameters means ignoring the observations altogether and opting for zero variance at a potential cost of high bias, and the success of $1/N$ tells us that the seven investment problems considered by DeMiguel et al. represent close to ideal uncertainty conditions for

exposing a bias bias. These conditions are not exceptional. Jagannathan and Ma (2003), to take another example, examined the role of bias and variance in the performance of portfolios constructed using regularization techniques that impose "wrong constraints", those that violate statistical characteristics of the population. These regularized portfolios, like $1/N$, often improved investment performance. We either view such results as to some extent foreseeable or accept that our intuitions suffer from a bias bias.

Example 4: Modeling Consumer Behavior. Managers in the retail industry often need to distinguish between active and inactive customers. One strategy is to use observations of past customer activity to estimate the parameters of a probabilistic model detailing the processes thought to drive customer behavior. For example, the Pareto/NBD model estimates the parameters of a Poisson process modeling customer purchasing behavior and the parameters of exponential distribution modeling customer dropout rates (Schmittlein et al., 1987). Combined with further probabilistic assumptions about the heterogeneity of customers within the population, categorization decisions are then made using a maximum likelihood calculation (Fader et al., 2005). An alternative strategy is to deploy a simple hiatus rule where customers who have not made a purchase within a hiatus period of, say, 9 months are classified as inactive, and all other customers are categorized as active. These two approaches differ in how they prioritize either bias reduction, the ability to accurately model the regularities of the data-generating processes, or variance reduction, the ability to limit the instability of predictions. Pitting these two strategies against each other, Wübben and Wangenheim (2008) compared simple hiatus rules and the Pareto/NBD model using transaction data from the apparel, airline, and music industries. First, they used 40 weeks of customer transaction data to estimate the parameters of the Pareto/NBD model. Using transaction data for the subsequent 40 weeks, they then estimated how accurately each model predicted future customer activity. For the apparel, airline, and music customers, the Pareto/NBD model achieved predictive accuracies of 75%, 74%, and 77%. Hiatus rules with cutoff periods recommended by experienced managers, on the other hand, predicted customer activity with accuracies of 83%, 77%, and 77%. Again, we see that prioritizing bias at the expense of variance appears to be the wrong choice, and the intuition that accurate probabilistic models of the data generating process will result in accurate predictions is another manifestation of the bias bias.

3.2 Categories of Uncertainty

We err when making predictions, forecasts, and decisions because the world is uncertain and models provide only a distorted representation of the complex processes that determine

events of interest. There are a number of ways the uncertain nature of our understanding can be formulated and categorized. Relative to a given model, the most basic category of uncertainty is stochasticity, also termed noise, randomness, or irreducible error. For example, we will err when attempting to predict the outcomes of a series of fair coin tosses. Underspecification is another source of uncertainty that occurs when there are insufficient observations to reliably estimate model parameters. For example, if we have only observed a limited number of bank failures then however sophisticated and well-justified our model of bank failure might be, uncertainty will surround the parameter values. Model misspecification, on the other hand, arises from discrepancies between model and the system being modelled. For example, if serial burglars tend to avoid offending in the immediate vicinity of their homes then modeling their journey to crime behavior using an exponential decay function will only approximate this behavior. Stochasticity, underspecification, and misspecification correspond to the noise, variance, bias terms of Equation 1 (Brighton and Gigerenzer, 2012).

Additional uncertainties arise when we consider that many systems are non-stationary. Sources of uncertainty when modeling how systems change over time include covariate shift and predictive shift (Shimodaira, 2000; Quionero-Candela et al., 2009). Covariate shift occurs when the distribution over the independent variables changes, and, say, the observations used to estimate model parameters and the observations we attempt to predict may be drawn from different regions of the observation space. Predictive shift occurs when the functional relationship between independent and dependent variables changes. For example, bank failures could occur for new and unforeseen reasons after the implementation of new regulations. It is relative to these categories of uncertainty, which are by no means comprehensive, that the role of the bias bias should considered. The statistical pathologies associated with the bias bias, recall, stem from viewing bias reduction as the primary means for reducing prediction error. This is equivalent to viewing model misspecification as key, which justifies a focus on high fidelity, causal models of the system being studied. Notice, though, that the bias bias does not imply an argument against the development of potentially complex causal models. Rather, it is an argument against the idea that we necessarily discover more predictive models by doing so. Indeed, some degree of causal thinking is needed to develop models at all. Haldane and Madouros (2012), for instance, decided to use a single liquidity indicator as a predictor of bank failure. This choice can have a causal justification even though they chose to disregard other potential causes of bank failure.

3.3   Modeling Under Uncertainty: Bias, Variance, and Optimality

When seen through the lens of the bias bias, how might Examples 1-4 above shed light on the rationality question? Before making a direct connection to considerations of rationality, three points need to be considered:

1. None of the four preceding examples have optimal solutions. The social and economic processes at work are unstable, complex, largely latent, and by any measure deeply uncertain. Of course, there is nothing to stop us from quantifying subjective probabilities and then using probability theory to derive and justify an optimality claim. But in doing so there is a sense in which we resort to changing the problem rather than exploring the space of potential solutions. Indeed, formulating optimal solutions to the problems of criminal profiling, bank regulation, portfolio investment, and customer profiling would say more about our modeling techniques than the systems being modeled.

2. This brings us to the issue of the relationship between bias, variance, and the existence of an assumed data generating function. Given that the meaning and measurement of bias requires that the underlying data generating function is known, how can I claim that the bias/variance perspective offers an insight into problems without optimal solutions? My claim is not that the bias/variance perspective undermines the goal of optimality. Rather, it brings into focus the need to understand how, when, and why statistical models incur low prediction error *outside* of their implied optimality conditions, and *relative* to other models. The meaning and measurement of variance, though, does not presuppose or require knowledge of the data generating function. Examples of abound of both justifying and explaining the performance of probabilistic models in terms on their relative ability to reduce variance outside of their optimality conditions (Hand and Yu, 2001; Domingos and Pazzani, 1997; Friedman, 1997; Ng and Jordan, 2002; Van Der Putten and Van Someren, 2004; Webb et al., 2005).

3. The preceding examples are not meant to imply that simple models necessarily incur lower prediction error. I selected these examples to underscore the point that variance reduction is a fundamental part of modeling under uncertainty, and that simplicity has proven to be a highly productive heuristic for discovering low variance models. They offer points of contrast to the popular assumption that advances in statistical modeling arise from greater sophistication and complexity (Hand, 2006). Other perspectives on variance reduction include regularization (Chen and Haykin,

2002) and ensemble methods (Seni and Elder, 2010), the latter being an example of how averaging the predictions of potentially complex models can also reduce variance.

In summary, the outcomes of exploratory statistical analyses tend not to be optimality claims, the bias/variance trade-off is a familiar concept in statistics, machine learning, forecasting, and econometrics, and the potential benefits of simplicity in statistical modeling have long been known. My argument has on occasion been misinterpreted as a claim that these three points are novel, whereas my argument is in fact that these three points are novel considerations when formulating the *rationality* question. Put differently, one may be familiar with and subscribe to the statistical approach I have presented when modeling social, economic, and other systems, but somehow regard them as being irrelevant to the problem of formulating rational action in these *same* contexts. The next stage of my argument details how the preceding insights, and the bias bias in particular, can and should guide a reassessment of the rationality question under conditions of unquantifiable uncertainty.

4   Towards Ecological Rationality

By any measure the bias/variance perspective has failed to penetrate the orthodox study of decision making under uncertainty. Consider first the case of decision theory, where rational actors are seen as Bayesian maximizers of expected utility operating over a mutually exclusive and exhaustive set of future states of the world, consequences, and the actions that map between them (Savage, 1954). On this view rationality is a primarily a subjective matter of axiomatic coherence rather than one of making rational inferences about the world. The cognitive and biological sciences tend to adopt a different attitude toward the rationality question by viewing rational decision makers as optimal Bayesian decision makers defined relative to a probabilistic model of the task environment (McNamara and Houston, 1980; Anderson, 1991b; Chater and Oaksford, 1999; Chater et al., 2006; Griffiths and Tenenbaum, 2006; Knill and Pouget, 2004; Gershman et al., 2015). On this view rational decisions are probabilistically justified decisions, or equivalently, predictions whose accuracy will depend on the fidelity of the assumed model of the environment.

This second notion of rationality, one centered on environmental correspondence rather than axiomatic coherence, is of greater relevance to the decisions of criminal profilers, bank regulators, portfolio managers, and marketing executives given that any loss their decisions incur will be determined by latent or future states of the task environment

(Arkes et al., 2016). When seen in this way, we cannot escape the bias/variance dilemma when formulating the rationality question because all claims flow from an assumed ability to model the data generating process. On recognizing this, I will argue that orthodox rationality – the view that rational decisions under uncertainty are optimal Bayesian decisions – can be seen as falling foul of the bias bias. Using a different but related notion of bias, Gigerenzer (2018) argues that behavioral economics also suffers from a bias bias in its formulation of the rationality question. However, to focus the discussion, this second stage of my argument will center on a contrast between Bayesian optimality modeling and ecological rationality. To establish this contrast it is first necessary to take a step back and consider how these two approaches to formulating the rationality question rest on distinct forms of statistical inquiry.

4.1 Constructing the Rationality Problem

The statistician Leo Breiman (2001) characterized two cultures of statistical modeling illustrated schematically in Figure 3. Consider first the scenario where we start with observations, each relating a set of independent variables to a dependent variable. The environment can be seen as a black box containing data-generating machinery that determines the joint distribution over the inputs to the black box (independent variables) and the output (the dependent variable) shown in Figure 3(a). Much of traditional statistical inquiry, and this includes the orthodox study of Bayesian optimality considered here, requires that we make a conjecture about the contents of this black box, depicted in Figure 3(b). We might, for instance, formulate an hypothesis space, prior distribution, and various parameters that we fit using the available observations. Breiman termed this approach data modeling and its defining characteristic is that at some point a conjecture is made about the contents of the black box.

An alternative to data modeling is what Breiman termed algorithmic modeling. When algorithmic modeling we refrain from making a conjecture about the contents of the black box and instead try to predict its behavior. As shown in Figure 3(c), the observations are used to estimate the predictive accuracy of competing models of inductive inference, which in practice usually means comparing machine learning algorithms using the principles of exploratory data analysis, much like the modeling approaches detailed in my discussion of the bias bias. Crucially, learning algorithms and the probabilistic assumptions they imply tend not to be seen as models or properties of the environment, but rather inductive biases likely to introduce model infidelities in order to reduce variance. Among the algorithms being considered, one or more will achieve the lowest prediction error. Such
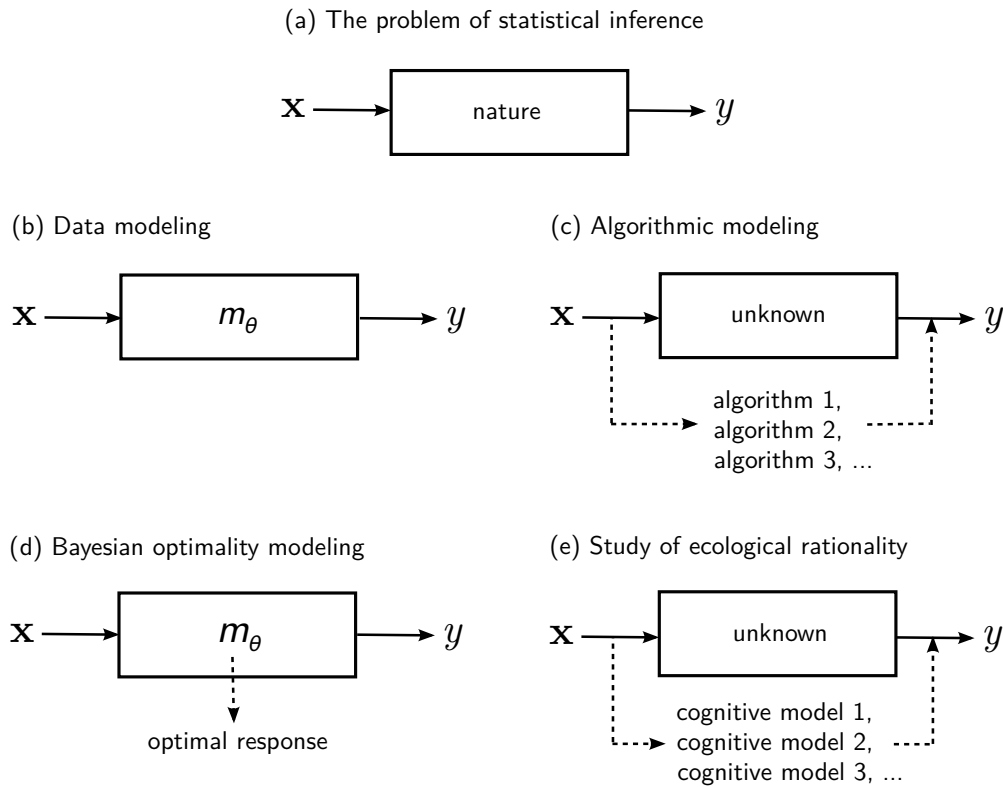
(a) The problem of statistical inference

$\mathbf{x} \longrightarrow$ [nature] $\longrightarrow y$

(b) Data modeling

$\mathbf{x} \longrightarrow$ [m_θ] $\longrightarrow y$

(c) Algorithmic modeling

$\mathbf{x} \longrightarrow$ [unknown] $\longrightarrow y$

algorithm 1,
algorithm 2,
algorithm 3, ...

(d) Bayesian optimality modeling

$\mathbf{x} \longrightarrow$ [m_θ] $\longrightarrow y$

optimal response

(e) Study of ecological rationality

$\mathbf{x} \longrightarrow$ [unknown] $\longrightarrow y$

cognitive model 1,
cognitive model 2,
cognitive model 3, ...

Figure 3 How should the problem of statistical inference be formulated? We start with observations of an unknown functional relationship between input vectors $\mathbf{x}$ (the independent variables) and output y (the dependent variable) determined by some aspect of nature, shown in (a). Following Breiman (2001), the statistical culture of data modeling views the conjecture of a stochastic model $y = m_\theta(\mathbf{x})$ of nature's black box an essential step in statistical inference, shown in (b). Rather than attempt to model the contents of nature's black box, algorithmic modeling is an incremental search for learning algorithms that can, to varying degrees, accurately predict the input-output relationship, shown in (c). Bayesian optimality modeling conducts data modeling in order to define an optimal response (d), while the study of ecological rationality conducts algorithmic modeling and interprets predictive models as potential cognitive models, shown in (e). Diagrams (a-c) adapted from Breiman (2001).

findings in no way license an optimality claim. They merely provide an indication of the kinds of algorithmic design decisions or statistical techniques that reduce prediction error, thereby suggesting further algorithms worth evaluating. Algorithmic modeling is exploratory, yields a functional understanding of the algorithms being considered, yet in no way invokes the concept of optimality to explain model performance. In short, algorithmic modeling seeks a relative understanding of the ability of competing algorithms to reduce prediction error.

Defining Optimality: Data Modeling

Because the claims of Bayesian optimality being considered here are made relative to a probabilistic model of the task environment they necessarily adhere to Breiman's notion of data modeling. To be clear on this point, probabilistic models of this kind are not (initially) seen as subjective models personal to the actor, but the outcome of an ecological analysis conducted by the theorist seeking to make the rationality claim. In his pioneering work on the Bayesian analysis of cognition, for example, Anderson (1991a) states that "the structure of such a theory is concerned with the outside world rather than what is inside the head" (p. 410). The decisions of this rational actor will be probabilistically optimal with respect to the model, but not necessarily the environment being modeled. This raises two concerns. First, how can this model be formulated when we lack sufficient knowledge or the observations needed to probabilistically quantify the relevant uncertainties? If we have observed only a limited sample of banking failures over one or two crisis cycles, or observed only a small number of offenders who commit a certain kind of serial crime, how should we proceed?

The orthodox solution is that whatever uncertainties we face, they can and should be probabilistically quantified using, say, uninformed priors, second-order probabilities, or imprecise probabilities. Alternatively, we could recall the second of the two epigraphs that began this discussion, the proposal that "it is sometimes *more* rational to admit that one does not have sufficient information for probabilistic beliefs than to pretend that one does" (Gilboa et al., 2012, p. 28). But what is this "*more* rational" alternative, and how might it be justified? The second concern is that by conjecturing an explanatory, causal, or high fidelity probabilistic model of the environment we run the risk of succumbing to the bias bias. This modeling goal will often diverge from the goal of predictive modeling because a simpler, possibly regularized, and likely biased model incorporating known representational inaccuracies may incur lower prediction error. How can cases where, as Haldane and Madouros (2012) put it, "simple does not just defeat complex; it trumps the truth" (p. 17) be reconciled with the goal of developing probabilistic models of the environment? My claim is that ecological rationality represents a "more rational" response that avoids the bias bias, and it proceeds through algorithmic modeling.

Exploring Ecological Rationality: Algorithmic Modeling

The study of ecological rationality considers the adaptive fit between organisms and the structure of natural environments (Gigerenzer et al., 1999; Gigerenzer and Selten, 2001; Gigerenzer et al., 2011; Todd et al., 2012). It proceeds by examining the interaction

between three components: (1) algorithmic models of how organisms make inductive inferences, with a particular focus on simple heuristics; (2) the properties of natural environments whose probabilistic structure is either uncertain or unknown; and (3), a formulation of the problem of statistical inference that defines and quantifies the meaning of an adaptive fit. A defining characteristic of simple heuristics is that they ignore information, and the overarching hypothesis is that these heuristics are a vital part of how organisms successfully cope with the uncertainty of the natural world. Results supporting this hypothesis are termed less-is-more effects, and they detail how minimalist processing strategies improve the accuracy of decisions relative to more complex and supposedly sophisticated strategies commonly assumed in the cognitive sciences (e.g., Gigerenzer and Goldstein, 1996; Czerlinski et al., 1999; Goldstein and Gigerenzer, 2002; Brighton, 2006; Gigerenzer and Brighton, 2009; Şimşek and Buckmann, 2015).

Because the overarching concern here is the contrast between orthodox and ecological rationality, I will sidestep a detailed discussion of models of ecological rationality and experimental studies focusing on their use by humans and other animals. I will instead focus on how component (3) of the interaction above typically assumes the perspective algorithmic modeling. Because of this, rationality claims are made relative to alternative models rather than an assumed data model. On finding that one model achieves lower prediction error than the alternatives, this model is regarded as more ecologically rational than the alternatives. Optimality plays no role (Brighton and Olsson, 2009). In Figure 1(d), for instance, the centroid method outperformed the probability surface model at low sample sizes not because is it optimal, and not because the centroid assumption holds for this offender. Similarly, when Haldane and Madouros (2012) found that a single indicator model outperformed the full CAMELS model, it was not because single indicator model was optimal, and not because the "true" economic processes determining bank receivership reduce to a single measure of liquidity. In both cases the simpler model was more ecologically rational than the competitors, and in both cases this was due to knowingly biased and "incorrect" models incurring low variance. The same statistical arguments justify the use of simple cognitive heuristics (Brighton and Gigerenzer, 2007; Gigerenzer and Brighton, 2009).

4.2 Orthodox Rationality and the Bias Bias

Advocates of ecological rationality have always maintained that the concept is incompatible with orthodox rationality. At the same time, several commentators have argued that while models of ecological rationality may generate interesting and insightful findings,

these findings not only fail to challenge orthodox rationality but require established rationality principles to be explained (Chater et al., 2003; Chater and Oaksford, 1999; Oaksford and Chater, 2009; Gintis, 2012; Jones and Love, 2011). I have argued that a fundamental incompatibility does exist, and to establish it requires that the relationship be considered at the level of the assumed statistical problem. Previous critiques, in contrast, have assumed that the terms of the relationship can be established by considering the algorithmic properties of models of ecological rationality alone. Algorithms alone, though, do not fully specify the problem they attempt to solve, leaving critics of ecological rationality free to assume optimality as the assumed goal. The onus is therefore on advocates of ecological rationality to not only specify component (3) of the interaction above, but explain why it leads to an incompatibility with orthodox rationality. I will now revisit previous critiques of ecological rationality in the light of my proposed response to this challenge.

A recurring critique is that when a heuristic works, one still needs a rational explanation for why it works, and that the concept of ecological rationality must ultimately appeal to orthodox rationality principles when formulating such an explanation (Chater et al., 2003). And this explanation should be Bayesian in nature. This view extends to critics of Bayesian optimality modeling, such as Jones and Love (2011), who argue that the two approaches are "highly compatible" because "any inference algorithm implicitly embodies a prior expectation about the environment" (p. 186). I interpret this point to mean that a heuristic (or any learning algorithm) is rational to the extent that this prior expectation coincides the probabilistic structure of the environment. The problem with this line of reasoning is that it requires that the probabilistic structure of the environment be known in order to establish and explain instances of success. As I have shown, this assumption is incompatible with and challenged by ecological rationality. Furthermore, even if we assume a probabilistic model of the task environment, this kind of explanation remains problematic. First, the relative rationality claims of ecological rationality are made relative not only to an estimate of prediction error, but relative to the alternative models being considered. These relative claims highlight that models can and typically do succeed outside of the their implied optimality conditions, and for reasons not easily explained in terms of a discrepancy between these conditions and the structure of the environment. Indeed, from bias/variance perspective, they can succeed *because* of this discrepancy[3].

---

[3] A classic example is the naïve Bayes classifier, a simple learning algorithm that makes a strong assumption that the features (i.e., cues, attributes, or independent variables) are conditionally independent of each other, given the class. Despite this assumption being extremely unlikely to be met in practice the naïve Bayes classifier has a long history of performing surprising well, particularly when learning from

A recurring theme among compatibility arguments is the idea that any simple heuristic can be reformulated as a probabilistic model operating with respect to a set of optimality conditions (e.g., Parpart et al., 2018). A compatibility between formulations of the *rationality question*, though, cannot be established at the level of specific models. One has to consider the issue at the level of the assumed statistical problem and uncertainty conditions characterising this problem. Gintis (2012) applies this reformulation argument slightly differently by noting that advocates of bounded and ecological have failed to appreciate that any algorithm with consistent preferences can be reformulated as maximizing an objective function. Therefore, assuming consistent preferences, instances of heuristic success can always be seen as solutions to an optimization problem (see also Boland, 1981). What this argument fails to consider is that we can also establish and justify cases of relative success in task environments where the optimal solution is undefined, and therefore in situations where we have no basis on which to claim that optimality has been achieved. The goal of reducing prediction error does not imply that we know what the minimum achievable prediction error is, and hence what the optimal response is. Gintis' argument would carry weight if the incompatibility between ecological and orthodox rationality centered exclusively on the issue of incompatible *algorithmic* properties, and specifically, a contrast between algorithms that don't explicitly optimize and those that do.

Finally, these concerns return us to the issue of the bias bias when we consider Oaksford and Chater's (2009) claim that Bayesian optimality modeling "cannot be replaced by, but seeks to explain, ecological rationality" (p. 110). Seen at the level of the assumed statistical problem, this claim implies that Breiman's notion of algorithmic modeling is reducible to data modeling. I have argued that two are incompatible, but there is also an argument to made that Oaksford and Chater's claimed relationship should be reversed, and that data modeling is more accurately seen a special case of algorithmic modeling. Specifically, data modeling is a special case of algorithmic modeling where we assign a single model the authority of assumed truth and we interpret all other models as approximations. It is relative to this "true" data model that optimality claims are then made. Crucially, the authority of this data model stems not from an analysis of its prediction error relative to alternative models, but faith in our ability to formally integrate the observations and our probabilistic beliefs about the environment. This is a form of Bayesianism, and the resulting model is assumed to be unbiased, and by conjecturing it rather than inferring it, the variance component of prediction is rendered irrelevant. On this view, it makes little sense to make an optimality claim relative to a model that sacrifices bias in order to achieve

sparse data (Hand and Yu, 2001; Domingos and Pazzani, 1997; Friedman, 1997; Ng and Jordan, 2002; Webb et al., 2005).

a greater reduction in variance, and this is why orthodox formulations of rationality can be seen as another instantiation of the bias bias.

5   Discussion

> You have a big approximation and a small approximation. The big approximation is your approximation to the problem you want to solve. The small approximation is involved in getting the solution to the approximate problem.
> (attributed to George Box [e.g., Fieberg et al., 2010, p. 10; Hand, 2014])

Individuals, groups, and organizations make decisions based on limited observations of complex, uncertain, and unstable environments. I take this to be the overarching problem that theories of rationality formulate and provide a normative solution to. Despite the controversial status of optimality in scientific inquiry (Dupré, 1987; Schoemaker, 1991), the idea that rationality implies optimality is so widely assumed as to seem barely worth discussing. Optimal Bayesian decision makers in the cognitive sciences, optimal foragers in biology, and Bayesian maximizers of expected utility in economics are different faces of the same interdisciplinary orthodoxy. My goal has been to reassess this view by first recognizing that in formulating the rationality problem we necessarily make a big approximation. Specifically, an optimal probabilistic response is a solution, a type of small approximation, to a specific kind of big approximation that presupposes all relevant uncertainties can and should be quantified. Given the uncertainty of the natural world I consider this to be an approximation worth questioning, and the examples I used underscore this point. There exist no optimal solutions to the problems of locating serial burglars, regulating international banks, managing portfolios, or identifying active customers. It is not the concept of rationality that implies optimality, but the big approximation of quantified uncertainty that implies optimality.

For each of the problems I considered, some models incur lower prediction error than others. For a specific task environment, it is therefore reasonable to regard some models as being more rational than others. The concepts of bias, variance, and the bias/variance dilemma help to disentangle the often complex relationship between the number of available observations, properties of models being considered, and their relative performance. Moreover, these concepts allow us to understand how biased, low variance models can outperform more "principled" models under conditions of heightened uncertainty, such as in the criminal profiling example detailed in Figure 2(c). Previous analyses of the problems I considered exhibited the negative consequences of the bias bias, a cluster of statistical

intuitions that neglect or ignore the role of variance. As such, the bias bias is the manifestation of a big approximation to the problem of statistical inference that masks the discovery of predictive, low variance models. If we then run with the idea that rationality claims can be relative claims, that these relative claims can be made in contexts where optimality is indeterminable and can be explained in terms of variance reduction, then orthodox rationality can be seen as falling foul of the bias bias. This is because optimality claims flow from an assumed ability to profitably conjecture an accurate, unbiased, probabilistic model of the data generating process. Variance plays no role in formulating conjectures of this kind. This is the point at which the substantive contrast between orthodox rationality and ecological rationality begins, and the contrast centers not on specific models or competing theories of cognitive processing, but on the assumed nature of the statistical problem. It is a clash of big approximations.

The study of ecological rationality makes relative rationality claims that refer to the ability of one cognitive mechanism to incur lower prediction error relative to other cognitive mechanisms in a given task environment. And because ecological rationality proceeds by conjecturing and analyzing the performance of cognitive mechanisms in environments with unknown or non-existent generating distributions, it is fundamentally exploratory. This is the big approximation of ecological rationality, and I used to Breiman's (2001) distinction between data modeling and algorithmic modeling to locate this clash of big approximations within the broader context of statistical inquiry. Although this distinction has not previously been related to the study of rationality, I argued that Breiman's distinction maps directly onto the big approximations of ecological and orthodox rationality. Furthermore, this contrast is not specific to Breiman but reflects long-standing, deep divisions in statistics (e.g., Tukey, 1962; Geisser, 1993; Vapnik, 1998; Shmueli, 2010). For example, the foundations of both algorithmic modeling and ecological rationality share those of the Minimum Description Length (MDL) principle developed by Rissanen (1978, 1986, 1989) which rests on the idea that "no assumption of a 'true' data-generating distribution is needed. This changes the objective and foundation for all model building" (Rissanen, 2007, p. 6). This general mode of statistical inquiry, I propose, also changes the foundation and objective for theories of rationality.

I began with a question: If we reassess the rationality question under the assumption that the uncertainty of the natural world is largely unquantifiable, where do we end up? We have arrived at the statistical, cognitive, and normative theory ecological rationality. Other arrival points are undoubtedly feasible, but it is worth noting that this reassessment of the rationality question provides a converging line of argument in support of Simon's (1956, 1978) bounded rationality. In his critique of *Homo economicus*, Simon stressed

that bounded rationality is shaped not only by the computational and cognitive bounds of the decision maker, but also informational bounds. In cognitive science and economics there is a tendency to focus on the first aspect of Simon's proposal, and view bounded rationality as reducing to the claim that optimization is infeasible due to cognitive and computational limitations, and this renders optimal responses either out of reach or in need of redefinition (e.g., Boland, 1981; Gintis, 2012; Griffiths et al., 2015; Gershman et al., 2015). This view neglects the potential for informational bounds to undermine the assumed objective of optimality and reshape the rationality question. In contrast, my reassessment of the rationality question stems entirely from a consideration of informational bounds, and specifically the impact of unquantifiable uncertainty. Yet this statistical reassessment converges on the same conclusion. The conclusion is that rationality in an uncertain world is algorithmic/predictive in nature rather than axiomatic/probabilistic in nature. Reinhard Selten (2001) argued that "bounded rationality cannot be precisely defined. It is a problem that needs to be explored" (p. 15). I agree, and my argument elaborates on Selten's point by concluding that once we accept that the uncertainty of the natural world is largely unquantifiable, from a statistical standpoint, the rational response is to explore algorithms capable of reducing uncertainty rather than seek to define a probabilistic model that attempts to fully quantify all relevant uncertainties.

Acknowledgements

References

Aikman, D., Galesic, M., Gigerenzer, G., Kapadia, S., Katsikopoulos, K. V., Kothiyal, A., Murphy, E., and Neumann, T. (2014). Taking uncertainty seriously: Simplicity versus complexity in financial regulation. *Bank of England Financial Stability Paper No. 28*. URL https://www.bankofengland.co.uk/financial-stability-paper/2014/taking-uncertainty-seriously-simplicity-versus-complexity-in-financial-regulation.

Anderson, J. R. (1991a). The adaptive nature of human categorization. *Psychological Review*, 98(3): 409–429. URL https://doi.org/10.1037/0033-295X.98.3.409.

Anderson, J. R. (1991b). Is human cognition adaptive? *Behavioral and Brain Sciences*, 14: 471–517. URL https://doi.org/10.1017/S0140525X00070801.

Arkes, H. R., Gigerenzer, G., and Hertwig, R. (2016). How bad is incoherence? *Decision*, 3(1): 20–39. URL https://doi.org/10.1037/dec0000043.

Binmore, K. (2009). *Rational decisions*. Princeton: Princeton University Press.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.

Block, R., and Bernasco, W. (2009). Finding a serial burglar's home using distance decay and conditional origin-destination patterns: A test of empirical Bayes journey-to-crime estimation in the Hague. *Journal of Investigative Psychology and Offender Profiling*, 6(3): 187–211. DOI 10.1002/jip.108. URL https://doi.org/10.1002/jip.108.

Boland, L. A. (1981). On the futility of criticizing the neoclassical maximization hypothesis. *The American Economic Review*, 71: 1031–1036.

Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*, 16: 199–231. URL https://projecteuclid.org/download/pdf_1/euclid.ss/1009213726.

Brighton, H. (2006). Robust inference with simple cognitive models. In C. Lebiere, and R. Wray (Eds.), *Between a Rock and a Hard Place: Cognitive Science Principles Meet AI-Hard Problems (AAAI Tehcnical Report SS-02-06)*, pages 189–211. Menlo Park, CA: AAAI Press.

Brighton, H. (2018). Rationality without optimality: Bounded and ecological rationality from a Marrian perspective. PsyArXiv Preprint: https://psyarxiv.com/m2sz5/. URL https://psyarxiv.com/m2sz5/. To appear in: R. Viale (Ed.) *Routledge Handbook of Bounded Rationality*.

Brighton, H., and Gigerenzer, G. (2007). Bayesian brains and cognitive mechanisms: Harmony or dissonance? In N. Chater, and M. Oaksford (Eds.), *The Probabalistic Mind: Prospects for Bayesian Cognitive Science*, pages 189–208. Cambridge: Cambridge University Press.

Brighton, H., and Gigerenzer, G. (2012). Are rational actor models "rational" outside small worlds? In S. Okasha, and K. Binmore (Eds.), *Evolution and Rationality: Decisions, Co-operation and Strategic Behaviour*, pages 84–109. Cambridge: Cambridge University Press.

Brighton, H., and Gigerenzer, G. (2015). The bias bias. *Journal of Business Research*, 68: 1772–1784. URL https://doi.org/10.1016/j.jbusres.2015.01.061.

Brighton, H., and Olsson, H. (2009). Identifying the optimal response is not a necessary step toward explaining function. *Behavioral and Brain Sciences*, 32: 85–86. URL https://doi.org/10.1017/S0140525X09000302.

Canter, D., Coffey, T., Huntley, M., and Missen, C. (2000). Predicting serial killers' home base using a decision support system. *Journal of Quantitative Criminology*, 16: 457–478. URL https://doi.org/10.1023/A:1007551316253.

Chater, N., and Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences*, 3(2): 57–65. URL https://www.ncbi.nlm.nih.gov/pubmed/10234228.

Chater, N., Oaksford, M., Nakisa, R., and Redington, M. (2003). Fast, frugal, and rational: How rational norms explain behavior. *Organizational Behavior and Human Decision Processes*, 90: 63–86. URL https://doi.org/10.1016/S0749-5978(02)00508-3.

Chater, N., Tenenbaum, J. B., and Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, 10(7): 287–291. URL https://www.ncbi.nlm.nih.gov/pubmed/16807064.

Chen, Z., and Haykin, S. (2002). On different facets of regularization theory. *Neural Computation*, 14(12): 2791–2846. URL https://www.ncbi.nlm.nih.gov/pubmed/12487794.

Şimşek, O., and Buckmann, M. (2015). Learning from small samples: An analysis of simple decision heuristics. In D. D. Lee, M. Sugiyama, and R. Garnett (Eds.), *Advances in Neural Information Processing Systems (NIPS) 28*.

Czerlinski, J., Gigerenzer, G., and Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd, and The ABC Research Group (Eds.), *Simple Heuristics That Make Us Smart*, pages 119–140. Oxford: Oxford University Press.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, 22: 1915–1953. URL https://doi.org/10.1093/rfs/hhm075.

Domingos, P., and Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, 29: 103–130. URL http://dblp.uni-trier.de/db/journals/ml/ml29.html#DomingosP97.

Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern classification.* New York: Wiley.

Dupré, J. (Ed.) (1987). *The latest on the best: Essays on evolution and optimality.* Cambridge, MA: MIT Press.

Epstein, L. G. (1999). A definition of uncertainty aversion. *Review of Economic Studies*, 66: 579–608. URL https://www.jstor.org/stable/2567015.

Fader, P. S., Hardie, B. G. S., and Lee, K. L. (2005). "Counting your customers" the easy way: An alternative to the Pareto/NBD model. *Marketing Science*, 24: 275–284. URL https://doi.org/10.1287/mksc.1040.0098.

Fieberg, J. R., Shertzer, K. W., Conn, P. B., Noyce, K. V., and Garshelis, D. L. (2010). Integrated population modeling of black bears in Minnesota: Implications for monitoring and management. *PLOS One*, 5: e12114. URL https://doi.org/10.1371/journal.pone.0012114.

Friedman, J. H. (1997). On bias, variance, 0/1-loss, and the curse-of-dimensionality. *Data Mining and Knowledge Discovery*, 1: 55–77. URL https://doi.org/10.1023/A:1009778005914.

Geisser, S. (1993). *Predictive inference: An introduction.* New York: Chapman and Hall.

Geman, S., Bienenstock, E., and Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, 4: 1–58. URL https://doi.org/10.1162/neco.1992.4.1.1.

Gershman, S. J., Horvitz, E. J., and Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, 349(6245): 273–278. ISSN 0036–8075. DOI 10.1126/science.aac6076. URL http://science.sciencemag.org/content/349/6245/273.

Gigerenzer, G. (2018). The bias bias in behavioral economics. *Review of Behavioral Economics*, 5(3–4): 303–336. URL http://dx.doi.org/10.1561/105.00000092.

Gigerenzer, G., and Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, 1: 107–143. URL https://www.ncbi.nlm.nih.gov/pubmed/25164802.

Gigerenzer, G., and Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded Rationality. *Psychological Review*, 103(4): 650–669. URL https://doi.org/10.1037/0033-295X.103.4.650.

Gigerenzer, G., Hertwig, R., and Pachur, T. (Eds.) (2011). *Heuristics: The foundations of adaptive behavior*. Oxford: Oxford University Press.

Gigerenzer, G., and Selten, R. (2001). *Bounded rationality: The adaptive toolbox*. Cambridge, MA: MIT Press.

Gigerenzer, G., Todd, P. M., and The ABC Research Group (1999). *Simple heuristics that make us smart*. Oxford: Oxford University Press.

Gilboa, I., Postlewaite, A., and Schmeidler, D. (2012). Rationality of belief or: Why savage's axioms are neither necessary or sufficient for rationality. *Synthese*, 187: 11–31. URL https://doi.org/10.1007/s11229-011-0034-2.

Gilboa, I., and Schmeidler, D. (1989). Maxmin expected utility with a non-unique prior. *Journal of Mathematical Economics*, 18: 141–153. URL https://doi.org/10.1016/0304-4068(89)90018-9.

Gintis, H. (2012). An evolutionary perspective on the unification of the behavioral sciences. In S. Okasha, and K. Binmore (Eds.), *Evolution and Rationality: Decisions, Co-operation and Strategic Behaviour*, pages 213–245. Cambridge: Cambridge University Press.

Goldstein, D. G., and Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, 109: 75–90. URL https://doi.org/10.1037/0033-295X.109.1.75.

Griffiths, T. L., Lieder, F., and Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, 7(2): 217–229. URL https://www.ncbi.nlm.nih.gov/pubmed/25898807.

Griffiths, T. L., and Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science*, 17(9): 767–773. URL https://www.ncbi.nlm.nih.gov/pubmed/16984293.

Haldane, A., and Madouros, V. (2012). The dog and the frisbee. Speech given at the Federal Reserve Bank of Kansas City's 36th Economic Policy Symposium "The Changing Policy Landscape", Jackson Hole, Wyoming, USA.

Hand, D. J. (2006). Classifier technology and the illusion of progress. *Statistical Science*, 21: 1–14. URL https://projecteuclid.org/euclid.ss/1149600839.

Hand, D. J. (2014). Hand writing: Solving the right problem. IMS Bulletin Online, 2014-3-14 (http://bulletin.imstat.org/2014/03/hand-writing-solving-the-right-problem/). URL http://bulletin.imstat.org/2014/03/hand-writing-solving-the-right-problem/.

Hand, D. J., and Yu, K. (2001). Idiot's Bayes: Not so stupid after all? *International Statistical Review*, 69: 385–398. URL https://www.jstor.org/stable/1403452.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The elements of statistical learning: data mining, Inference, and Prediction.* New York: Springer.

Jagannathan, R., and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58: 1651–1683. URL https://www.jstor.org/stable/3648224.

Jones, M., and Love, B. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, 34: 169–231. URL https://www.ncbi.nlm.nih.gov/pubmed/21864419.

Knight, F. H. (1921). *Risk, uncertainty, and profit.* Boston, MA: Houghton Mifflin Company.

Knill, D. C., and Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12): 712–719. URL https://www.ncbi.nlm.nih.gov/pubmed/15541511.

Kozyreva, A., and Hertwig, R. (2019). The interpretation of uncertainty in ecological rationality. *Synthese.* URL https://doi.org/10.1007/s11229-019-02140-w. Advance online publication. doi:10.1007/s11229-019-02140-w.

Leitner, M., and Kent, J. (2009). Bayesian journey-to-crime modelling of single and multiple crime-type series in Baltimore County, MD. *Journal of Investigative Psychology and Offender Profiling*, 6(3): 213–236. URL https://doi.org/10.1002/jip.109.

Levine, N. (2009). Introduction to the special issue on Bayesian journey-to-crime modelling. *Journal of Investigative Psychology and Offender Profiling*, 6(3): 167–185. URL https://doi.org/10.1002/jip.107.

Levine, N. (2010). CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 3.3). Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC. URL https://www.icpsr.umich.edu/icpsrweb/NACJD/studies/2824.

Levine, N., and Block, R. (2011). Bayesian journey-to-crime estimation: An improvement in geographic profiling methodology. *The Professional Geographer*, 63: 213–229. URL https://doi.org/10.1080/00330124.2010.547152.

Machina, M. J., and Schmeidler, D. (1992). A more robust definition of subjective probability. *Econometrica*, 60: 745–780. URL https://www.jstor.org/stable/2951565.

Markowitz, H. (1959). *Portfolio selection: Efficient diversification of investments*. New York: John Wiley & Sons.

McNamara, J. M., and Houston, A. (1980). The application of statistical decision theory to animal behaviour. *Journal of Theoretical Biology*, 85: 673–690. URL https://doi.org/10.1016/0022-5193(80)90265-9.

Ng, A. Y., and Jordan, M. I. (2002). On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. In T. Dietterich, S. Becker, and Z. Ghahramani (Eds.), *Advances in Neural Information Processing Systems (NIPS) 14*.

Oaksford, M., and Chater, N. (2009). The uncertain reasoner: Bayes, logic, and rationality. *Behavioral and Brain Sciences*, 32: 105–120. URL https://doi.org/10.1017/S0140525X0900051X.

O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statistical Science*, 1: 502–518. URL https://projecteuclid.org/download/pdf_1/euclid.ss/1177013525.

Parpart, P., Jones, M., and Love, B. (2018). Heuristics as Bayesian inference under extreme priors. *Cognitive Psychology*, 102: 127–144. URL https://www.ncbi.nlm.nih.gov/pubmed/29500961.

Quionero-Candela, J., Sugiyama, M., Schwaighofer, A., and Lawrence, N. D. (Eds.) (2009). *Dataset shift in machine learning*. Cambridge, MA: MIT Press.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14: 465–471. URL https://doi.org/10.1016/0005-1098(78)90005-5.

Rissanen, J. (1986). Stochastic complexity and modeling. *The Annals of Statistics*, 14: 1080–1100. URL https://projecteuclid.org/euclid.aos/1176350051.

Rissanen, J. (1989). *Stochastic complexity in statistical inquiry*. Singapore: World Scientific Publishing.

Rissanen, J. (2007). *Information and complexity in statistical modeling*. New York: Springer.

Rossmo, D. K. (1999). *Geographic profiling*. Boca Raton, FL: CRC Press.

Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.

Schmittlein, D. C., Morrison, D. G., and Colombo, R. (1987). Counting your customers: Who are they and what will they do next? *Management Science*, 33: 1–24. URL https://www.jstor.org/stable/2631608.

Schoemaker, P. J. H. (1991). The quest for optimality: A positive heuristic of science? *Behavioral And Brain Sciences*, 14: 205–215. URL https://doi.org/10.1017/S0140525X00066140.

Selten, R. (2001). What is bounded rationality? In G. Gigerenzer, and R. Selten (Eds.), *Bounded rationality: The adaptive toolbox*, pages 13–36. Cambridge, MA: MIT Press.

Seni, G., and Elder, J. F. (2010). *Ensemble methods in data mining: Improving accuracy through combining predictions*. San Francisco: Morgan & Claypool.

Shimodaira, H. (2000). Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference*, 90: 227–244. URL https://doi.org/10.1016/S0378-3758(00)00115-4.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25(3): 289–310. URL https://arxiv.org/pdf/1101.0891.pdf.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63: 129–138. URL https://doi.org/10.1037/h0042769.

Simon, H. A. (1978). Rationality as process and as product of thought. *American Economic Review*, 68: 1–16. URL https://www.jstor.org/stable/1816653.

Simon, H. A. (1989). The scientist as problem solver. In D. Klahr, and K. Kotovsky (Eds.), *Complex information processing: The Impact of Herbert A. Simon*, pages 373–398. Hillsdale, NJ: Erlbaum.

Snook, B., Zito, M., Bennell, C., and Taylor, P. J. (2005). On the complexity and accuracy of geographical profiling strategies. *Journal of Quantitative Criminology*, 21: 1–25. URL https://doi.org/10.1007/s10940-004-1785-4.

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

Tukey, J. W. (1962). The future of data analysis. *The Annals of Mathematical Statistics*, 33: 1–67. URL https://projecteuclid.org/euclid.aoms/1177704711.

Van Der Putten, P., and Van Someren, M. (2004). A bias-variance analysis of a real world learning problem: The CoIL challenge 2000. *Machine Learning*, 57: 177–195. URL https://doi.org/10.1023/B:MACH.0000035476.95130.99.

Vapnik, V. N. (1998). *Statistical learning theory*. New York: Wiley.

Webb, G. I., Boughton, J. R., and Wang, Z. (2005). Not so naive Bayes: Aggregating one dependence estimators. *Machine Learning*, 58: 5–24. URL https://doi.org/10.1007/s10994-005-4258-6.

Wübben, M., and Wangenheim, F. v. (2008). Instant customer base analysis: Managerial heuristics often "get it right". *Journal of Marketing*, 72: 82–93. URL https://doi.org/10.1509/jmkg.72.3.082.

Economics
The Open-Access, Open-Assessment E-Journal

Please note:

You are most sincerely encouraged to participate in the open assessment of this article. You can do so by either recommending the article or by posting your comments.

Please go to:

http://dx.doi.org/10.5018/economics-ejournal.ja.2020-2

The Editor