3

Prediction, Judgment, and Complexity

A Theory of Decision-Making and Artificial Intelligence

Ajay Agrawal, Joshua Gans, and Avi Goldfarb

3.1 Introduction

There is widespread discussion regarding the impact of machines on employment (see Autor 2015). In some sense, the discussion mirrors a long-standing literature on the impact of the accumulation of capital equipment on employment; specifically, whether capital and labor are substitutes or complements (Acemoglu 2003). But the recent discussion is motivated by the integration of software with hardware and whether the role of machines goes beyond physical tasks to mental ones as well (Brynjolfsson and McAfee 2014). As mental tasks were seen as always being present and essential, human comparative advantage in these was seen as the main reason why, at least in the long term, capital accumulation would complement employment by enhancing labor productivity in those tasks.

The computer revolution has blurred the line between physical and men-

Ajay Agrawal is the Peter Munk Professor of Entrepreneurship at the Rotman School of Management, University of Toronto, and a research associate of the National Bureau of Economic Research. Joshua Gans is professor of strategic management and holder of the Jeffrey S. Skoll Chair of Technical Innovation and Entrepreneurship at the Rotman School of Management, University of Toronto (with a cross appointment in the Department of Economics), and a research associate of the National Bureau of Economic Research. Avi Goldfarb holds the Rotman Chair in Artificial Intelligence and Healthcare and is professor of marketing at the Rotman School of Management, University of Toronto, and is a research associate of the National Bureau of Economic Research.

Our thanks to Andrea Prat, Scott Stern, Hal Varian, and participants at the AEA (Chicago), NBER Summer Institute (2017), NBER Economics of AI Conference (Toronto), Columbia Law School, Harvard Business School, MIT, and University of Toronto for helpful comments. Responsibility for all errors remains our own. The latest version of this chapter is available at joshuagans.com. For acknowledgments, sources of research support, and disclosure of the authors' material financial relationships, if any, please see http://www.nber.org/chapters/c14010.ack.

tal tasks. For instance, the invention of the spreadsheet in the late 1970s fundamentally changed the role of bookkeepers. Prior to that invention, there was a time-intensive task involving the recomputation of outcomes in spreadsheets as data or assumptions changed. That human task was substituted by the spreadsheet software that could produce the calculations more quickly, cheaply, and frequently. However, at the same time, the spreadsheet made the jobs of accountants, analysts, and others far more productive. In the accounting books, capital was substituting for labor, but the mental productivity of labor was being changed. Thus, the impact on employment critically depended on whether there were tasks the "computers cannot do."

These assumptions persist in models today. Acemoglu and Restrepo (2017) observe that capital substitutes for labor in certain tasks while at the same time technological progress creates new tasks. They make what they call a "natural assumption" that only labor can perform the new tasks as they are more complex than previous ones.[1] Benzell et al. (2015) consider the impact of software more explicitly. Their environment has two types of labor—high-tech (who can, among other things, code) and low-tech (who are empathetic and can handle interpersonal tasks). In this environment, it is the low-tech workers who cannot be replaced by machines while the high-tech ones are employed initially to create the code that will eventually displace their kind. The results of the model depend, therefore, on a class of worker who cannot be substituted directly for capital, but also on the inability of workers themselves to substitute between classes.

In this chapter, our approach is to delve into the weeds of what is happening currently in the field of artificial intelligence (AI). The recent wave of developments in AI all involve advances in machine learning. Those advances allow for automated and cheap prediction; that is, providing a forecast (or nowcast) of a variable of interest from available data (Agrawal, Gans and Goldfarb 2018b). In some cases, prediction has enabled full automation of tasks—for example, self-driving vehicles where the process of data collection, prediction of behavior and surroundings, and actions are all conducted without a human in the loop. In other cases, prediction is a standalone tool—such as image recognition or fraud detection—that may or may not lead to further substitution of human users of such tools by machines. Thus far, substitution between humans and machines has focused mainly on cost considerations. Are machines cheaper, more reliable, and more scalable (in their software form) than humans? This chapter, however, considers the role of prediction in decision-making explicitly and from that examines the complementary skills that may be matched with prediction within a task.

1. To be sure, their model is designed to examine how automation of tasks causes a change in factor prices that biases innovation toward the creation of new tasks that labor is more suited to.

Our focus, in this regard, is on what we term *judgment*. While judgment is a term with broad meaning, here we use it to refer to a very specific skill. To see this, consider a decision. That decision involves choosing an action, x, from a set, X. The payoff (or reward) from that action is defined by a function, $u(x, \theta)$ where θ is a realization of an uncertain state drawn from a distribution, $F(\theta)$. Suppose that, prior to making a decision, a *prediction* (or signal), s, can be generated that results in a posterior, $F(\theta|s)$. Thus, the decision maker would solve

$$\max_{x \in X} \int u\left(x, \theta\right) dF\left(\theta \middle| s\right).$$

In other words, a standard problem of choice under uncertainty. In this standard world, the role of prediction is to improve decision-making. The payoff, or utility function, is known.

To create a role for judgment, we depart from this standard set-up in statistical decision theory and ask how a decision maker comes to know the function, $u(x, \theta)$? We assume that this is not simply given or a primitive of the decision-making model. Instead, it requires a human to undertake a costly process that allows the mapping from (x, θ) to a particular payoff value, u, to be discovered. This is a reasonable assumption given that beyond some rudimentary experimentation in closed environments, there is no current way for an AI to impute a utility function that resides with humans. Additionally, this process separates the costs of providing the mapping for each pair, (x, θ). (Actually, we focus, without loss in generality, on situations where $u(x, \theta) \neq u(x)$ for all θ and presume that if a payoff to an action is state independent that payoff is known.) In other words, while prediction can obtain a signal of the underlying state, judgment is the process by which the payoffs from actions that arise based on that state can be determined. We assume that this process of determining payoffs requires human understanding of the situation: it is not a prediction problem.

For intuition on the difference between prediction and judgment, consider the example of credit card fraud. A bank observes a credit card transaction. That transaction is either legitimate or fraudulent. The decision is whether to approve the transaction. If the bank knows for sure that the transaction is legitimate, the bank will approve it. If the bank knows for sure that it is fraudulent, the bank will refuse the transaction. Why? Because the bank knows the payoff of approving a legitimate transaction is higher than the payoff of refusing that transaction. Things get more interesting if the bank is uncertain about whether the transaction is legitimate. The uncertainty means that the bank also needs to know the payoff from refusing a legitimate transaction and from approving a fraudulent transaction. In our model, judgment is the process of determining these payoffs. It is a costly activity, in the sense that it requires time and effort.

As the new developments regarding AI all involve making prediction more readily available, we ask, how does judgment and its endogenous appli-

cation change the value of prediction? Are prediction and judgment substitutes or complements? How does the value of prediction change monotonically with the difficulty of applying judgment? In complex environments (as they relate to automation, contracting, and the boundaries of the firm), how do improvements in prediction affect the value of judgment?

We proceed by first providing supportive evidence for our assumption that recent developments in AI overwhelmingly impact the costs of prediction. We then use the example of radiology to provide a context for understanding the different roles of prediction and judgment. Drawing inspiration from Bolton and Faure-Grimaud (2009), we then build the baseline model with two states of the world and uncertainty about payoffs to actions in each state. We explore the value of judgment in the absence of any prediction technology, and then the value of prediction technology when there is no judgment. We finish the discussion of the baseline model with an exploration of the interaction between prediction and judgment, demonstrating that prediction and judgment are complements as long as judgment isn't too difficult. We then separate prediction quality into prediction frequency and prediction accuracy. As judgment improves, accuracy becomes more important relative to frequency. Finally, we examine complex environments where the number of potential states is large. Such environments are common in economic models of automation, contracting, and boundaries of the firm. We show that the effect of improvements in prediction on the importance of judgment depend a great deal on whether the improvements in prediction enable automated decision-making.

3.2 AI and Prediction Costs

We argue that the recent advances in artificial intelligence are advances in the technology of prediction. Most broadly, we define prediction as the ability to take known information to generate new information. Our model emphasizes prediction about the state of the world.

Most contemporary artificial intelligence research and applications come from a field now called "machine learning." Many of the tools of machine learning have a long history in statistics and data analysis, and are likely familiar to economists and applied statisticians as tools for prediction and classification.[2] For example, Alpaydin's (2010) textbook *Introduction to Machine Learning* covers maximum likelihood estimation, Bayesian estimation, multivariate linear regression, principal components analysis, clustering, and nonparametric regression. In addition, it covers tools that may be less familiar, but also use independent variables to predict outcomes:

2. We define prediction as known information to generate new information. Therefore, classification techniques such as clustering are prediction techniques in which the new information to be predicted is the appropriate category or class.

regression trees, neural networks, hidden Markov models, and reinforcement learning. Hastie, Tibshirani, and Friedman (2009) cover similar topics. The 2014 *Journal of Economic Perspectives* symposium on big data covered several of these less familiar prediction techniques in articles by Varian (2014) and Belloni, Chernozhukov, and Hansen (2014).

While many of these prediction techniques are not new, recent advances in computer speed, data collection, data storage, and the prediction methods themselves have led to substantial improvements. These improvements have transformed the computer science research field of artificial intelligence. The Oxford English Dictionary defines artificial intelligence as "[t]he theory and development of computer systems able to perform tasks normally requiring human intelligence." In the 1960s and 1970s, artificial intelligence research was primarily rules-based, symbolic logic. It involved human experts generating rules that an algorithm could follow (Domingos 2015, 89). These are not prediction technologies. Such systems became very good chess players and they guided factory robots in highly controlled settings; however, by the 1980s, it became clear that rules-based systems could not deal with the complexity of many nonartificial settings. This led to an "AI winter" in which research funding artificial intelligence projects largely dried up (Markov 2015).

Over the past ten years, a different approach to artificial intelligence has taken off. The idea is to program computers to "learn" from example data or experience. In the absence of the ability to predetermine the decision rules, a data-driven prediction approach can conduct many mental tasks. For example, humans are good at recognizing familiar faces, but we would struggle to explain and codify this skill. By connecting data on names to image data on faces, machine learning solves this problem by predicting which image data patterns are associated with which names. As a prominent artificial intelligence researcher put it, "Almost all of AI's recent progress is through one type, in which some input data (A) is used to quickly generate some simple response (B)" (Ng 2016). Thus, the progress is explicitly about improvements in prediction. In other words, the suite of technologies that have given rise to the recent resurgence of interest in artificial intelligence use data collected from sensors, images, videos, typed notes, or anything else that can be represented in bits to fill in missing information, recognize objects, or forecast what will happen next.

To be clear, we do not take a position on whether these prediction technologies really do mimic the core aspects of human intelligence. While Palm Computing founder Jeff Hawkins argues that human intelligence is—in essence—prediction (Hawkins 2004), many neuroscientists, psychologists, and others disagree. Our point is that the technologies that have been given the label artificial intelligence are prediction technologies. Therefore, in order to understand the impact of these technologies, it is important to assess the impact of prediction on decisions.

3.3 Case: Radiology

Before proceeding to the model, we provide some intuition of how prediction and judgment apply in a particular context where prediction machines are expected to have a large impact: radiology. In 2016, Geoff Hinton—one of the pioneers of deep learning neural networks—stated that it was no longer worth training radiologists. His strong implication was that radiologists would not have a future. This is something that radiologists have been concerned about since 1960 (Lusted 1960). Today, machine-learning techniques are being heavily applied in radiology by IBM using its Watson computer and by a start-up, Enlitic. Enlitic has been able to use deep learning to detect lung nodules (a fairly routine exercise)[3] but also fractures (which is more complex). Watson can now identify pulmonary embolism and some other heart issues. These advances are at the heart of Hinton's forecast, but have also been widely discussed among radiologists and pathologists (Jha and Topol 2016). What does the model in this chapter suggest about the future of radiologists?

If we consider a simplified characterization of the job of a radiologist, it would be that they examine an image in order to characterize and classify that image and return an assessment to a physician. While often that assessment is a diagnosis (i.e., "the patient has pneumonia"), in many cases, the assessment is in the negative (i.e., "pneumonia not excluded"). In that regard, this is stated as a predictive task to inform the physician of the likelihood of the state of the world. Using that, the physician can devise a treatment.

These predictions are what machines are aiming to provide. In particular, it might provide a differential diagnosis of the following kind:

> *Based on Mr Patel's demographics and imaging, the mass in the liver has a 66.6 percent chance of being benign, 33.3 percent chance of being malignant, and a 0.1 percent of not being real.*[4]

The action is whether some intervention is needed. For instance, if a potential tumor is identified in a noninvasive scan, then this will inform whether an invasive examination will be conducted. In terms of identifying the state of the world, the invasive exam is costly but safe—it can deduce a cancer with certainty and remove it if necessary. The role of a noninvasive exam is to inform whether an invasive exam should be forgone. That is, it is to make physicians more confident about abstaining from treatment and further analysis. In this regard, if the machine improves prediction, it will lead to fewer invasive examinations.

3. "You did not go to medical school to measure lung nodules." http://www.medscape.com/viewarticle/863127#vp_2.

4. http://www.medscape.com/viewarticle/863127#vp_3.

Judgment involves understanding the payoffs. What is the payoff to conducting a biopsy if the mass is benign, malignant, or not real? What is the payoff to not doing anything in those three states? The issue for radiologists in particular is whether a trained specialist radiologist is in the best position to make this judgment or will it occur further along the chain of decision-making or involve new job classes that merge diagnostic information such as a combined radiologist/pathologist (Jha and Topol 2016). Next, we formalize these ideas.

3.4 Baseline Model

Our baseline model is inspired by the "bandit" environment considered by Bolton and Faure-Grimaud (2009), although it departs significantly in the questions addressed and base assumptions made. Like them, in our baseline model, we suppose there are two states of the world, $\{\theta_1,\theta_2\}$ with prior probabilities of $\{\mu,1-\mu\}$. There are two possible actions: a state independent action with known payoff of S (safe) and a state dependent action with two possible payoffs, R or r, as the case may be (risky).

As noted in the introduction, a key departure from the usual assumptions of rational decision-making is that the decision maker does not know the payoff from the risky action in each state and must apply *judgment* to determine that payoff.[5] Moreover, decision makers need to be able to make a judgment for each state that might arise in order to formulate a plan that would be the equivalent of payoff maximization. In the absence of such judgment, the ex ante expectation that the risky action is optimal in any state is v (which is independent between states). To make things more concrete, we assume $R > S > r$.[6] Thus, we assume that v is the probability in any state that the risky payoff is R rather than r. This is not a conditional probability of the state. It is a statement about the payoff, given the state.

In the absence of knowledge regarding the specific payoffs from the risky action, a decision can only be made on the basis of prior probabilities. Then the safe action will be chosen if

$$\mu\big(vR + (1-v)r\big) + \big(1-\mu\big)\big(vR + (1-v)r\big) = vR + (1-v)r \leq S.$$

5. Bolton and Faure-Grimaud (2009) consider this step to be the equivalent of a thought experiment where thinking takes time. To the extent that our results can be interpreted as a statement about the comparative advantage of humans, we assume that only humans can do judgment.

6. Thus, we assume that the payoff function, u, can only take one of three values, $\{R, r, S\}$. The issue is which combinations of state realization and action lead to which payoffs. However, we assume that S is the payoff from the safe action regardless of state and so this is known to the decision maker. As it is the relative payoffs from actions that drive the results, this assumption is without loss in generality. Requiring this property of the safe action to be discovered would just add an extra cost. Implicitly, as the decision maker cannot make a decision in complete ignorance, we are assuming that the safe action's payoff can be judged at an arbitrarily low cost.

So that the payoff is: $V_0 = \max\{vR + (1 - v)r, S\}$. To make things simpler, we will focus our attention on the case where the safe action is—in the absence of prediction or judgment—the default. That is, we assume that

(A1) **(Safe Default)** $vR + (1 - v)r \leq S$.

This assumption is made for simplicity only and will not change the qualitative conclusions.[7] Under (A1), in the absence of knowledge of the payoff function or a signal of the state, the decision maker would choose S.

3.4.1 Judgment in the Absence of Prediction

Prediction provides knowledge of the state. The process of judgment provides knowledge of the payoff function. Judgment therefore allows the decision maker to understand which action is optimal for a given state should it arise. Suppose that this knowledge is gained without cost (as it would be assumed to do under the usual assumptions of economic rationality). In other words, the decision maker has knowledge of optimal action in a given state. Then the risky action will be chosen (a) if it is the preferred action in both states (which arises with probability v^2); (b) if it is the preferred action in θ_1 but not θ_2 and $\mu R + (1 - \mu)r > S$ (with probability $v(1 - v)$); or (c) if it is the preferred action in θ_2 but not θ_1 and $\mu r + (1 - \mu)R > S$ (with probability $v(1 - v)$). Thus, the expected payoff is

$$v^2 R + v(1-v)\max\left\{\mu R + (1-\mu)r, S\right\} + v(1-v)\max\left\{\mu r + (1-\mu)R, S\right\} + (1-v)^2 S.$$

Note that this is greater than V_0. The reason for this is that, when there is uncertainty, judgment is valuable because it can identify actions that are dominant or dominated—that is, that might be optimal across states. In this situation, any resolution of uncertainty does not matter as it will not change the decision made.

A key insight is that judgment itself can be consequential.

Result 1: *If* $\max\{\mu R + (1 - \mu)r, \mu r + (1 - \mu)R\} > S$, *it is possible that judgment alone can cause the decision to switch from the default action (safe) to the alternative action (risky).*

As we are motivated by understanding the interplay between prediction and judgment, we want to make these consequential. Therefore, we make the following assumption to ensure prediction always has some value:

(A2) **(Judgment Insufficient)** $\max\{\mu R + (1 - \mu)r, \mu r + (1 - \mu)R\} \leq S$.

Under this assumption, if different actions are optimal in each state and this is known, the decision maker will not change to the risky action. This, of course, implies that the expected payoff is

7. Bolton and Faure-Grimaud (2009) make the opposite assumption. Here, as our focus is on the impact of prediction, it is better to consider environments where prediction has the effect of reducing uncertainty over riskier actions.

$$v^2R + (1 - v^2)S.$$

Note that, absent any cost, full judgment improves the decision maker's expected payoff.

Judgment does not come for free. We assume here that it takes time (although the formulation would naturally match with the notion that it takes costly effort). Suppose the discount factor is $\delta < 1$. A decision maker can spend time in a period determining what the optimal action is for a particular state. If they choose to apply judgment with respect to state θ_i, then there is a probability λ_i that they will determine the optimal action in that period and can make a choice based on that judgment. Otherwise, they can choose to apply judgment to that problem in the next period.

It is useful, at this point, to consider what judgment means once it has been applied. The initial assumption we make here is that the knowledge of the payoff function depreciates as soon as a decision is made. In other words, applying judgment can delay a decision (and that is costly) and it can improve that decision (which is its value) but it cannot generate experience that can be applied to other decisions (including future ones). In other words, the initial conception of judgment is the application of *thought* rather than the gathering of *experience*.[8] Practically, this reduces our examination to a static model. However, in a later section, we consider the experience formulation and demonstrate that most of the insights of the static model carry over to the dynamic model.

In summary, the timing of the game is as follows:

1. At the beginning of a decision stage, the decision maker chooses whether to apply judgment and to what state or whether to simply choose an action without judgment. If an action is chosen, uncertainty is resolved and payoffs are realized and we move to a new decision stage.

2. If judgment is chosen, with probability, $1 - \lambda_i$, they do not find out the payoffs for the risky action in that state, a period of time elapses and the game moves back to 1. With probability λ_i, the decision maker gains this knowledge. The decision maker can then take an action, uncertainty is resolved and payoffs are realized, and we move to a new decision stage (back to 1). If no action is taken, a period of time elapses and the current decision stage continues.

3. The decision maker chooses whether to apply judgment to the other state. If an action is chosen, uncertainty is resolved and payoffs are realized and we move to a new decision stage (back to 1).

4. If judgment is chosen, with probability, $1 - \lambda_{-i}$, they do not find out the payoffs for the risky action in that state, a period of time elapses and the game moves back to 1. With probability λ_{-i}, the decision maker gains this knowledge. The decision maker then chooses an action, uncertainty

8. The experience frame is considered in Agrawal, Gans, and Goldfarb (2018a).

Table 3.1 **Model parameters**

Parameter	Description
S	Known payoff from the safe action
R	Potential payoff from the risky action in a given state
r	Potential payoff from the risky action in a given state
θ_i	Label of state $i \in \{1,2\}$
μ	Probability of state 1
v	Prior probability that the payoff in a given state is R
λ_i	Probablilty that decision maker learns the payoff to the risky action θ_i if judgment is applied for one period
δ	Discount factor

is resolved and payoffs are realized, and we move to a new decision stage (back to 1).

When prediction is available, it will become available prior to the beginning of a decision stage. The various parameters are listed in table 3.1.

Suppose that the decision maker focuses on judging the optimal action (i.e., assessing the payoff) for θ_i. Then the expected present discount payoff from applying judgment is

$$\lambda_i\left(vR+(1-v)S\right)+(1-\lambda_i)\delta\lambda_i\left(vR+(1-v)S\right)+\sum_{t=2}^{\infty}(1-\lambda_i)^t\delta^t\lambda_i\left(vR+(1-v)S\right)$$
$$=\frac{\lambda_i}{1-(1-\lambda_i)\delta}\left(vR+(1-v)S\right).$$

The decision maker eventually can learn what to do and will earn a higher payoff than without judgment, but will trade this off against a delay in the payoff.

This calculation presumes that the decision maker knows the state—that θ_i is true—prior to engaging in judgment. If this is not the case, then the expected present discounted payoff to judgment on, say, θ_1 alone is

$$\frac{\lambda_1}{1-(1-\lambda_1)\delta}\left(\max\left\{v\left(\mu R+(1-\mu)\left(vR+(1-v)r\right)\right)+(1-v)\left(\mu r+(1-\mu)\left(vR+(1-v)r\right)\right),S\right\}\right)$$
$$=\frac{\lambda_1}{1-(1-\lambda_1)\delta}\left(\max\left\{v\left(\mu R+(1-\mu)\left(vR+(1-v)r\right)\right),S\right\}+(1-v)S\right),$$

where the last step follows from equation (A1). To make exposition simpler, we suppose that $\lambda_1 = \lambda_2 = \lambda$. In addition, let $\hat{\lambda} = \lambda / (1-(1-\lambda)\delta)$; $\hat{\lambda}$ can be given a similar interpretation to λ, the quality of judgment.

If the strategy were to apply judgment on one state only and then make a decision, this would be the relevant payoff to consider. However, because judgment is possible in both states, there are several cases to consider.

First, the decision maker might apply judgment to both states in sequence. In this case, the expected present discounted payoff is

$$\hat{\lambda}^2\Big(v^2R + v(1-v)\max\{\mu R + (1-\mu)r, S\} + v(1-v)\max\{\mu r + (1-\mu)R, S\} + (1-v)^2 S\Big) = \hat{\lambda}^2\big(v^2R + (1-v^2)S\big),$$

where the last step follows from equation (A1).

Second, the decision maker might apply judgment to, say, θ_1 first and then, contingent on the outcome there, apply judgment to θ_2. If the decision maker chooses to pursue judgment on θ_2 if the outcome for θ_1 is that the risky action is optimal, the payoff becomes

$$\hat{\lambda}\Big(v\hat{\lambda}\big(vR + (1-v)\max\{\mu R + (1-\mu)r, S\}\big) + (1-v)\max\{\mu r + (1-\mu)(vR + (1-v)r), S\}\Big) = \hat{\lambda}\Big(v\hat{\lambda}\big(vR + (1-v)S\big) + (1-v)S\Big).$$

If the decision maker chooses to pursue judgment on θ_2 after determining that the outcome for θ_1 is that the safe action is optimal, the payoff becomes

$$\hat{\lambda}\Big(v\max\{\mu R + (1-\mu)(vR + (1-v)r), S\} + (1-v)\hat{\lambda}\big(v\max\{\mu r + (1-\mu)R, S\} + (1-v)S\big)\Big) = \hat{\lambda}\Big(v\max\{\mu R + (1-\mu)(vR + (1-v)r), S\} + (1-v)\hat{\lambda}S\Big).$$

Note that this is option is dominated by not applying further judgment at all if the outcome for θ_1 is that the safe action is optimal.

Given this we can prove the following:

Proposition 1: *Under (A1) and (A2), and in the absence of any signal about the state, (a) judging both states and (b) continuing after the discovery that the safe action is preferred in a state are never optimal.*

Proof: Note that judging two states is optimal if

$$\hat{\lambda} > \frac{S}{v\max\{\mu r + (1-\mu)R, S\} + (1-v)S}$$

$$\hat{\lambda} > \frac{\mu R + (1-\mu)\big(vR + (1-v)r\big)}{vR + (1-v)\max\{\mu R + (1-\mu)r, S\}}.$$

As (A2) implies that $\mu r + (1 - \mu)R \leq S$, the first condition reduces to $\hat{\lambda} > 1$. Thus, (a) judging two states is dominated by judging one state and continuing to explore only if the risk is found to be optimal in that state.

Turning to the strategy of continuing to apply judgment only if the safe action is found to be preferred in a state, we can compare this to the payoff from applying judgment to one state and then acting immediately. Note that

$$\hat{\lambda}\Big(v\max\big\{\mu R+(1-\mu)(vR+(1-v)r),S\big\}+(1-v)\hat{\lambda}S\Big)$$
$$>\hat{\lambda}\Big(v\max\big\{\mu R+(1-\mu)(vR+(1-v)r),S\big\}+(1-v)S\Big).$$

This can never hold, proving that (b) is dominated.

The intuition is similar to Propositions 1 and 2 in Bolton and Faure-Grimaud (2009). In particular, applying judgment is only useful if it is going to lead to the decision maker switching to the risky action. Thus, it is never worthwhile to unconditionally explore a second state as it may not change the action taken. Similarly, if judging one state leads to knowledge the safe action continues to be optimal in that state, in the presence of uncertainty about the state, even if knowledge is gained of the payoff to the risky action in the second state, that action will never be chosen. Hence, further judgment is not worthwhile. Hence, it is better to choose immediately at that point rather than delay the inevitable.

Given this proposition, there are only two strategies that are potentially optimal (in the absence of prediction). One strategy (we will term here J1) is where judgment is applied to one state and if the risky action is optimal, then that action is taken immediately; otherwise, the safe default is taken immediately. The state where judgment is applied first is the state most likely to arise. This will be state 1 if $\mu > 1/2$. This strategy might be chosen if

$$\hat{\lambda}\Big(v\max\big\{\mu R+(1-\mu)(vR+(1-v)r),S\big\}+(1-v)S\Big)>S$$
$$\Rightarrow \hat{\lambda}>\hat{\lambda}_{J1}\equiv\frac{S}{v\max\big\{\mu R+(1-\mu)(vR+(1-v)r),S\big\}+(1-v)S},$$

which clearly requires that $\mu R + (1 - \mu)(vR + (1 - v)r) > S$.

The other strategy (we will term here J2) is where judgment is applied to one state and if the risky action is optimal, then judgment is applied to the next state; otherwise, the safe default is taken immediately. Note that J2 is preferred to J1 if

$$\hat{\lambda}\Big(v\hat{\lambda}(vR+(1-v)S)+(1-v)S\Big)$$
$$>\hat{\lambda}\Big(v\max\big\{\mu R+(1-\mu)(vR+(1-v)r),S\big\}+(1-v)S\Big)$$
$$\Rightarrow \hat{\lambda}v(vR+(1-v)S)>v\max\big\{\mu R+(1-\mu)(vR+(1-v)r),S\big\}$$
$$\Rightarrow \hat{\lambda}>\frac{\max\big\{\mu R+(1-\mu)(vR+(1-v)r),S\big\}}{vR+(1-v)S}.$$

This is intuitive. Basically, it is only when the efficiency of judgment is sufficiently high that more judgment is applied. However, for this inequality to be relevant, J2 must also be preferred to the status quo yielding a payoff of S. Thus, J2 is not dominated if

$$\hat{\lambda} > \hat{\lambda}_{J2} \equiv \max\left\{ \frac{\max\left\{\mu R + (1-\mu)\left(vR+(1-v)r\right), S\right\}}{vR+(1-v)S}, \frac{\sqrt{S\left(4v^2R+S(1+2v-3v^2)\right)} - (1-v)S}{2v\left(vR+(1-v)S\right)} \right\},$$

where the first term is the range where J2 dominates J1, while the second term is where J2 dominates S alone; so for J2 to be optimal, it must exceed both. Note also that as $\mu \to (S-r)/(R-r)$ (its highest possible level consistent with [A1] and [A2]), then $\hat{\lambda}_{J2} \to 1$.

If $\mu R + (1-\mu)(vR + (1-v)r) > S$, note that

$$\hat{\lambda}_{J2} > \hat{\lambda}_{J1} \Rightarrow \frac{\mu R + (1-\mu)\left(vR+(1-v)r\right)}{vR+(1-v)S} > \frac{S}{v\left(\mu R+(1-\mu)\left(vR+(1-v)r\right)\right)+(1-v)S}$$

$$\Rightarrow (1-v)S\left(\mu R+(1-\mu)\left(vR+(1-v)r\right)-S\right) > v\left(RS-\left(\mu R+(1-\mu)\left(vR+(1-v)r\right)\right)^2\right),$$

which may not hold for v sufficiently high. However, it can be shown that when $\hat{\lambda}_{J2} + \hat{\lambda}_{J1}$, then the two terms of $\hat{\lambda}_{J2}$ are equal and the second term exceeds the first when $\hat{\lambda}_{J2}\ \hat{\lambda}_{J1}$. This implies that in the range where $\hat{\lambda}_{J2} < \hat{\lambda}_{J1}$, J2 dominates J1.

This analysis implies there are two types of regimes with judgment only. If $\hat{\lambda}_{J2} > \hat{\lambda}_{J1}$, then easier decisions (with high $\hat{\lambda}$) involve using J2, the next tranche of decisions use J1 (with intermediate $\hat{\lambda}$) while the remainder involves no exercise of judgment at all. On the other hand, if $\hat{\lambda}_{J2} < \hat{\lambda}_{J1}$, then the easier decisions involve using J2 while the remainder do not involve judgment at all.

3.4.2 Prediction in the Absence of Judgment

Next, we consider the model with prediction but no judgment. Suppose that there exists an AI that can, if deployed, identify the state prior to a decision being made. In other words, prediction, if it occurs, is perfect; an assumption we will relax in a later section. Initially, suppose there is no judgment mechanism to determine what the optimal action is in each state.

Recall that, in the absence of prediction or judgment, (A1) ensures that the safe action will be chosen. If the decision maker knows the state, then the risky action in a given state is chosen if

$$vR + (1-v)r > S.$$

This contradicts (A1). Thus, the expected payoff is

$$V_P = S,$$

which is the same outcome if there is no judgment or prediction.

3.4.3 Prediction and Judgment Together

Both prediction and judgment can be valuable on their own. The question we next wish to consider is whether they are complements or substitutes.

While perfect prediction allows you to choose an action based on the

actual rather than expected state, it also affords the same opportunity with respect to judgment. As judgment is costly, it is useful not to waste considering what action might be taken in a state that does not arise. This was not possible when there was no prediction. But if you receive a prediction regarding the state, you can then apply judgment exclusively to actions in relation to that state. To be sure, that judgment still involves a cost, but at the same time does not lead to any wasted cognitive resources.

Given this, if the decision maker were the apply judgment after the state is predicted, their expected discounted payoff would be

$$V_{PJ} = \max\left\{\hat{\lambda}\left(vR + (1-v)S\right), S\right\}.$$

This represents the highest expected payoff possible (net of the costs of judgment). A necessary condition for both prediction and judgment to be optimal is that: $\hat{\lambda} \geq \hat{\lambda}_{PJ} \equiv s/[vR + (1 - v)S]$. Note that $\hat{\lambda}_{PJ} \leq \hat{\lambda}_{J1}, \hat{\lambda}_{J2}$.

3.4.4 Complements or Substitutes?

To evaluate whether prediction and judgment are complements or substitutes, we adopt the following parameterization for the effectiveness of prediction: we assume that with probability *e* an AI yields a prediction, while otherwise, the decision must be made in its absence (with judgment only). With this parameterization, we can prove the following:

Proposition 2: *In the range of* λ *where* $\hat{\lambda} < \hat{\lambda}_{J2}$, *e and* λ *are complements, otherwise they are substitutes.*

Proof: Step 1. Is $\hat{\lambda}_{J2} > R/[2(vR + (1 - v)S)]$? First, note that

$$\frac{\max\left\{\mu R + (1-\mu)\left(vR + (1-v)r\right), S\right\}}{vR + (1-v)S} > \frac{R}{2\left(vR + (1-v)S\right)}$$

$$\Rightarrow \max\left\{\mu R + (1-\mu)\left(vR + (1-v)r\right), S\right\} > \frac{1}{2}R.$$

Note that by (A2) and since $\mu > (1/2)$, $S > \mu R + (1 - \mu)r > (1/2)R$ so this inequality always holds.

Second, note that

$$\frac{\sqrt{S\left(4v^2R + S(1 + 2v - 3v^2)\right)} - (1-v)S}{2v\left(vR + (1-v)S\right)} > \frac{R}{2\left(vR + (1-v)S\right)}$$

$$\Rightarrow S\left(4v^2R + S(1 + 2v - 3v^2)\right) > \left(vR + (1-v)S\right)^2$$

$$\Rightarrow S(S - 2R) > v(R^2 - 6RS + S^2),$$

which holds as the left-hand side is always positive while the right-hand side is always negative.

Step 2: Suppose that $\mu R + (1 - \mu)(vR + (1 - v)r) \leq S$; then J1 is never optimal. In this case, the expected payoff is

$$eV_{PJ} + (1-e)V_{J2} = e\hat{\lambda}(vR + (1-v)S) + (1-e)\hat{\lambda}\left(v\hat{\lambda}(vR + (1-v)S) + (1-v)S\right).$$

This mixed partial derivative with respect to (e, $\hat{\lambda}$) is $v(R - 2\hat{\lambda}(vR + (1 - v)S))$. This is positive if $R/[2(vR + (1 - v)S)] \geq \hat{\lambda}$. By Step 1, this implies that for $\hat{\lambda} < \hat{\lambda}_{J2}$, prediction and judgment are complements; otherwise, they are substitutes.

Step 3: Suppose that that $\mu R + (1 - \mu)(vR + (1 - v)r) > S$. Note that for $\hat{\lambda}_{J1}$ $\hat{\lambda} < \hat{\lambda}_{J2}$, J1 is preferred to J2. In this case, the expected payoff to prediction and judgment is

$$e\hat{\lambda}(vR + (1-v)S) + (1-e)\hat{\lambda}\left(v \max\left\{\mu R + (1-\mu)(vR + (1-v)r), S\right\} + (1-v)S\right).$$

This mixed partial derivative with respect to (e,$\hat{\lambda}$) is $v(R - \max\{\mu R + (1 - \mu)(vR + (1 - v)r), S\}) > 0$. By Step 1, this implies that for $\hat{\lambda} < \hat{\lambda}_{J2}$, prediction and judgment are complements; otherwise, they are substitutes.

The intuition is as follows. When $\hat{\lambda} < \hat{\lambda}_{J2}$, then, in the absence of prediction either no judgment is applied or, alternatively, strategy J1 (with one round of judgment) is optimal; *e* parameterizes the degree of difference between the expected value with both prediction and judgment and the expected value without prediction with an increase in λ, increasing both. However, with one round of judgment, the increase when judgment is used alone is less than that when both are used together. Thus, when $\hat{\lambda} < \hat{\lambda}_{J2}$, prediction and judgment are complements.

By contrast, when $\hat{\lambda} > \hat{\lambda}_{J2}$, then strategy J2 (with two rounds of judgment) is used in the absence of prediction. In this case, increasing λ increases the expected payoff from judgment alone disproportionately more because judgment is applied on both states, whereas under prediction and judgment it is only applied on one. Thus, improving the quality of judgment reduces the returns to prediction. And so, when $\hat{\lambda} > \hat{\lambda}_{J2}$, prediction and judgment are substitutes.

3.5 Complexity

Thus far, the model illustrates the interplay between knowing the reward function (judgment) and prediction. While those results show that prediction and judgment can be substitutes, there is a sense in which they are more naturally complements. The reason is this: what prediction enables is a form of state-contingent decision-making. Without a prediction, a decision maker is forced to make the same choice regardless of the state that might arise. In the spirit of Herbert Simon, one might call this a heuristic. And in the absence of prediction, the role of judgment is to make that choice. Moreover, that choice is easier—that is, more likely to be optimal—when there exists dominant (or "near dominant") choices. Thus, when either the state space or the action space expand (as it may in more complex situations), it is

less likely that there will exist a dominant choice. In that regard, faced with complexity, in the absence of prediction, the value of judgment diminishes and we are more likely to see decision makers choose default actions that, on average, are likely to be better than others.

Suppose now we add a prediction machine to the mix. While in our model such a machine, when it renders a prediction, can perfectly signal the state that will arise, let us consider a more convenient alternative that may arise in complex situations: the prediction machine can perfectly signal some states (should they arise), but for other states no precise prediction is possible except for the fact that one of those states is the correct one. In other words, the prediction machine can sometimes render a fine prediction and otherwise a coarse one. Here, an improvement in the prediction machine means an increase in the number of states in which the machine can render a fine prediction.

Thus, consider an N-state model where the probability of state i is μ_i. Suppose that states $\{1, \ldots, m\}$ can be finely predicted by an AI, while the remainder cannot be distinguished. Suppose that in the states that cannot be distinguished applying judgment is not worthwhile so that the optimal choice is the safe action. Also, assume that when a prediction is available, judgment is worthwhile; that is, $\hat{\lambda} \geq s/[vR + (1 - v)S]$. In this situation, the expected present discounted value when both prediction and judgment are available is

$$V_{PJ} = \hat{\lambda}\sum_{i=1}^{m}\mu_i\left(vR + (1-v)S\right) + \sum_{i=m+1}^{N}\mu_i S.$$

Similarly, it is easy to see that $V_P = V_J = S = V_0$ as $vR + (1 - v)r \leq S$. Note that as m increases (perhaps because the prediction machine learns to predict more states), then the marginal value of better judgment increases. That is, $\hat{\lambda}\mu_m\left(vR + (1-v)S\right) - \mu_m S$ is increasing in $\hat{\lambda}$.

What happens as the situation becomes more complex (that is, N increases)? An increase in N will weakly lead to a reduction in μ_i for any given i. Holding m fixed (and so the quality of the prediction machine does not improve with the complexity of the world), this will reduce the value of prediction and judgment as greater weight is placed on states where prediction is unavailable; that is, it is assumed that the increase in complexity does not, ceteris paribus, create a state where prediction is available. Thus, complexity appears to be associated with *lower* returns to both prediction and judgment. Put differently, an improvement in prediction machines would mean m increases with N fixed. In this case, the returns to judgment rise as greater weight is put on states where prediction is available.

This insight is useful because there are several places in the economics literature where complexity has interacted with other economic decisions. These include automation, contracting, and firm boundaries. We discuss each of these in turn, highlighting potential implications.

3.5.1 Automation

The literature on automation is sometimes synonymous with AI. This arises because AI may power new robots that are able to operate in open environments thanks to machine learning. For instance, while automated trains have been possible for some time since they run on tracks, automated cars are new because they need to operate in far more complex environments. It is prediction in those open environments that has allowed the emergence of environmentally flexible capital equipment. Note that leads to the implication that as AI improves, tasks in more complex environments can be handled by machines (Acemoglu and Restrepo 2017).

However, this story masks the message that emerges from our analysis that recent AI developments are all about prediction. Why prediction enables automated vehicles is because it is relatively straightforward to describe (and hence, program) what those vehicles should do in different situations. In other words, if prediction enables "state contingent decisions," then automated vehicles arise because someone knows what decision is optimal in each state. In other words, automation means that judgment can be encoded in machine behavior. Prediction added to that means that automated capital can be moved into more complex environments. In that respect, it is perhaps natural to suggest that improvements in AI will lead to a substitution of humans for machines as more tasks in more complex environments become capable of being programmed in a state-contingent manner.

That said, there is another dimension of substitution that arises in complex environments. As noted above, when states cannot be predicted (something that for a given technology is more likely to be the case in more complex environments), then the actions chosen are more likely to be defaults or the results of heuristics that perform, on average, well. Many, including Acemoglu and Restrepo (2017), argue that it is for more complex tasks that humans have a comparative advantage relative to machines. However, this is not at all obvious. If it is known that a particular default or heuristic should be used, then a machine can be programmed to undertake this. In this regard, the most complex tasks—precisely because little is known regarding how to take better actions given that the prediction of the state is coarse—may be more, not less, amenable to automation.

If we had to speculate, imagine that states were ordered in terms of diminished likelihood (i.e., $\mu_i \geq \mu_j$ for all $i < j$). The lowest index states might be ones that, because they arrive frequently, there is knowledge of what the optimal action is in each and so they can be programmed to be handled by a machine. The highest index states similarly, because the optimal action that cannot be determined can also be programmed. It is the intermediate states that arise less frequently but not infrequently where, if a reliable prediction existed, could be handled by humans applying judgment when those states arose. Thus, the payoff could be written

$$V_{PJ} = \sum_{i=1}^{k} \mu_i \left(vR + (1-v)S \right) + \hat{\lambda} \sum_{i=k+1}^{m} \mu_i \left(vR + (1-v)S \right) + \sum_{i=m+1}^{N} \mu_i S,$$

where tasks 1 through k are automated using prediction because there is knowledge of the optimal action. If this was the matching of tasks to machines and humans, then it is not at all clear whether an increase in complexity would be associated with more or less human employment.

That said, the issue for the automation literature is not subtleties over the term "complex tasks," but as AI becomes more prevalent, where might the substitution of machines for humans arise. As noted above, an increase in AI increases m. At this margin, humans are able to come into the marginal tasks and, because a prediction machine is available, use judgment to conduct state-contingent decisions in those situations. Absent other effects, therefore, an increase in AI is associated with more human labor on any given task. However, as the weight on those marginal tasks is falling in the level of complexity, it may not be the more complex tasks that humans are performing more of. On the other hand, one can imagine that in a model with a full labor market equilibrium that an increase in AI that enables more human judgment at the margin may also create opportunities to study that judgment to see if it can be programmed into lower index states and be handled by machines. So, while the AI does not necessarily cause more routine tasks to be handled by machines, it might create the economic conditions that lead to just that.

3.5.2 Contracting

Contracting shares much with programming. Here is Jean Tirole (2009, 265) on the subject:

> Its general thrust goes as follows. The parties to a contract (buyer, seller) initially avail themselves of an available design, perhaps an industry standard. This design or contract is the best contract under existing knowledge. The parties are unaware, however, of the contract's implications, but they realize that something may go wrong with this contract; indeed, they may exert cognitive effort in order to find out about what may go wrong and how to draft the contract accordingly: put differently, a *contingency* is foreseeable (perhaps at a prohibitively high cost), but not necessarily foreseen. To take a trivial example, the possibility that the price of oil increases, implying that the contract should be indexed on it, is perfectly foreseeable, but this does not imply that parties will think about this possibility and index the contract price accordingly.

Tirole argues that contingencies can be planned for in contracts using cognitive effort (akin to what we have termed here as judgment), while others may be optimally left out because the effort is too costly relative to the return given, say, the low likelihood that contingency arises.

This logic can assist us in understanding what prediction machines might

do to contracts. If an AI becomes available then, in writing contracts, it is possible, because fine state predictions are possible, to incur cognitive costs to determine what the contingencies should be if those states should arise. For other states, the contract will be left incomplete—perhaps for a default action or alternatively some renegotiation process. A direct implication of this is that contracts may well become less incomplete.

Of course, when it comes to employment contracts, the effects may be different. As Herbert Simon (1951) noted, employment contracts differ from other contracts precisely because it is often not possible to specify what actions should be performed in what circumstance. Hence, what those contracts often allocate are different decision rights.

What is of interest here is the notion that contacts can be specified clearly—that is, programmed—but also that prediction can activate the use of human judgment. That latter notion means that actions cannot be easily contracted—by definition, contractibility is programming and needing judgment implies that programming was not possible. Thus, as prediction machines improve and more human judgment is optimal, then that judgment will be applied outside of objective contract measures—including objective performance measures. If we had to speculate, this would favor more subjective performance processes, including relational contracts (Baker, Gibbons, and Murphy 1999).[9]

3.5.3 Firm Boundaries

We now turn to consider what impact AI may have on firm boundaries (that is, the make or buy decision). Suppose that it is a buyer (B) who receives the value from a decision taken—that is, the payoff from the risky or safe action as the case may be. To make things simple, let's assume that $\mu_i = \mu$ for all i, so that $V = k\big(vR + (1-v)S\big) + \hat{\lambda}(m-k)\big(vR + (1-v)S\big) + (N-m)S$.

We suppose that the tasks are undertaken by a seller (S). The tasks $\{1, \ldots, k\}$ and $\{m+1, \ldots, N\}$ can be contracted upon, while the intermediate tasks require the seller to exercise judgment. We suppose that the cost of providing judgment is a function $c(\hat{\lambda})$, which is nondecreasing and convex. (We write this function in terms of $\hat{\lambda}$ just to keep the notation simple.) The costs can be anticipated by the buyer. So if one of the intermediate states arises, the buyer can choose to give the seller a fixed price contract (and bear none of the costs) or a cost-plus contract (and bear all of them).

Following Tadelis (2002), we assume that the seller market is competitive and so all surplus accrues to the buyer. In this case, the buyer return is

9. A recent paper by Dogan and Yildirim (2017) actually considers how automation might impact on worker contracts. However, they do not examine AI per se, and focus on how it might change objective performance measures in teams moving from joint performance evaluation to more relative performance evaluation.

$$k(vR+(1-v)S)+\max\left\{\hat{\lambda}(m-k)(vR+(1-v)S),S\right\}+(N-m)S-p-zc(\hat{\lambda}),$$

while the seller return is: $p - (1 - z)c(\hat{\lambda})$. Here $p + zc(\hat{\lambda})$ is the contract price and z is 0 for a fixed price contract and 1 for a cost-plus contract. Note that only with a cost-plus contract does the seller exercise any judgment. Thus, the buyer chooses a cost-plus over a fixed price contract if

$$k(vR+(1-v)S)+\max\left\{\hat{\lambda}(m-k)(vR+(1-v)S),S\right\}+(N-m)S-c(\hat{\lambda}) > k(vR+(1-v)S)+(N-k)S.$$

It is easy to see that as m rises (i.e., prediction becomes cheaper), a cost-plus contract is more likely to be chosen. That is, incentives fall as prediction becomes more abundant.

Now we can consider the impact of integration. We assume that the buyer can choose to make the decisions themselves, but at a higher cost. That is, $c(\hat{\lambda},I) > c(\hat{\lambda})$ where I denotes integration. We also assume that $\partial c(\hat{\lambda},I)/\partial\hat{\lambda} > (\partial c(\hat{\lambda}) / \partial\hat{\lambda})$. Under integration, the buyer's value is

$$k(vR+(1-v)S)+\hat{\lambda}^{*}(m-k)(vR+(1-v)S)+(N-m)S-c(\hat{\lambda}^{*},I)$$

where $\hat{\lambda}^{*}$ maximizes the buyer payoff in this case. Given this, it can easily be seen that as m increases, the returns to integration rise.

By contrast, notice that as k increases, the incentives for a cost-plus contract are diminished and the returns to integration fall. Thus, the more prediction machines allow for the placement of contingencies in a contract (the larger m-k), the higher powered will seller incentives be and the more likely there is to be integration.

Forbes and Lederman (2009) showed that airlines are more likely to vertically integrate with regional partners when scheduling is more complex: specifically, where bad weather is more likely to lead to delays. The impact of prediction machines will depend on whether they lead to an increase in the number of states where the action can be automated in a state-contingent manner (k) relative to the increase in the number of states where the state becomes known but the action cannot be automated (m). If the former, then we will see more vertical integration with the rise of prediction machines. If the latter, we will see less. The difference is driven by the need for more costly judgment in the vertically integrated case as m-k rises.

3.6 Conclusions

In this chapter, we explore the consequences of recent improvements in machine-learning technology that have advanced the broader field of artificial intelligence. In particular, we argue that these advances in the ability of machines to conduct mental tasks are driven by improvements in machine prediction. In order to understand how improvements in machine prediction will impact decision-making, it is important to analyze how the payoffs of the model arise. We label the process of learning payoffs "judgment."

By modeling judgment explicitly, we derive a number of useful insights into the value of prediction. We show that prediction and judgment are generally complements, as long as judgment is not too difficult. We also show that improvements in judgment change the type of prediction quality that is most useful: better judgment means that more accurate predictions are valuable relative to more frequent predictions. Finally, we explore the role of complexity, demonstrating that, in the presence of complexity, the impact of improved prediction on the value of judgment depends on whether improved prediction leads to automated decision-making. Complexity is a key aspect of economic research in automation, contracting, and the boundaries of the firm. As prediction machines improve, our model suggests that the consequences in complex environments are particularly fruitful to study.

There are numerous directions research in this area could proceed. First, the chapter does not explicitly model the form of the prediction—including what measures might be the basis for decision-making. In reality, this is an important design variable and impacts on the accuracy of predictions and decision-making. In computer science, this is referred to as the choice of surrogates, and this appears to be a topic amenable for economic theoretical investigation. Second, the chapter treats judgment as largely a human-directed activity. However, we have noted that it can else be encoded, but have not been explicit about the process by which this occurs. Endogenising this—perhaps relating it to the accumulation of experience—would be an avenue for further investigation. Finally, this is a single-agent model. It would be interesting to explore how judgment and prediction mix when each is impacted upon by the actions and decisions of other agents in a game theoretic setting.

References

Acemoglu, Daron. 2003. "Labor- and Capital-Augmenting Technical Change." *Journal of the European Economic Association* 1 (1): 1–37.

Acemoglu, Daron, and Pascual Restrepo. 2017. "The Race between Machine and Man: Implications of Technology for Growth, Factor Shares, and Employment." Working paper, Massachusetts Institute of Technology.

Agrawal, Ajay, Joshua S. Gans, and Avi Goldfarb. 2018a. "Human Judgment and AI Pricing." *American Economic Association: Papers & Proceedings*, 108:58–63.

———. 2018b. *Prediction Machines: The Simple Economics of Artificial Intelligence.* Boston, MA: Harvard Business Review Press.

Alpaydin, Ethem. 2010. *Introduction to Machine Learning*, 2nd ed. Cambridge, MA: MIT Press.

Autor, David. 2015. "Why Are There Still So Many Jobs? The History and Future of Workplace Automation." *Journal of Economic Perspectives* 29 (3): 3–30.

Baker, George, Robert Gibbons, and Kevin Murphy. 1999. "Informal Authority in Organizations." *Journal of Law, Economics, and Organization* 15:56–73.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "High-

Dimensional Methods and Inference on Structural and Treatment Effects." *Journal of Economic Perspectives* 28 (2): 29–50.
Benzell, Seth G., Laurence J. Kotlikoff, Guillermo LaGarda, and Jeffrey D. Sachs. 2015. "Robots Are Us: Some Economics of Human Replacement." NBER Working Paper no. 20941, Cambridge, MA.
Bolton, P., and A. Faure-Grimaud. 2009. "Thinking Ahead: The Decision Problem." *Review of Economic Studies* 76:1205–38.
Brynjolfsson, Erik, and Andrew McAfee. 2014. *The Second Machine Age*. New York: W. W. Norton.
Dogan, M., and P. Yildirim. 2017. "Man vs. Machine: When Is Automation Inferior to Human Labor?" Unpublished manuscript, The Wharton School of the University of Pennsylvania.
Domingos, Pedro. 2015. *The Master Algorithm.* New York: Basic Books.
Forbes, Silke, and Mara Lederman. 2009. "Adaptation and Vertical Integration in the Airline Industry." *American Economic Review* 99 (5): 1831–49.
Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York: Springer.
Hawkins, Jeff. 2004. *On Intelligence*. New York: Times Books.
Jha, S., and E. J. Topol. 2016. "Adapting to Artificial Intelligence: Radiologists and Pathologists as Information Specialists." *Journal of the American Medical Association* 316 (22): 2353–54.
Lusted, L. B. 1960. "Logical Analysis in Roentgen Diagnosis." *Radiology* 74:178–93.
Markov, John. 2015. *Machines of Loving Grace*. New York: HarperCollins Publishers.
Ng, Andrew. 2016. "What Artificial Intelligence Can and Can't Do Right Now." *Harvard Business Review Online*. Accessed Dec. 8, 2016. https://hbr.org/2016/11/what-artificial-intelligence-can-and-cant-do-right-now.
Simon, H. A. 1951. "A Formal Theory of the Employment Relationship." *Econometrica* 19 (3): 293–305.
Tadelis, S. 2002. "Complexity, Flexibility and the Make-or-Buy Decision." *American Economic Review* 92 (2): 433–37.
Tirole, J. 2009. "Cognition and Incomplete Contracts." *American Economic Review* 99 (1): 265–94.
Varian, Hal R. 2014. "Big Data: New Tricks for Econometrics." *Journal of Economic Perspectives* 28 (2): 3–28.

Comment Andrea Prat

One of the key activities of organizations is to collect, process, combine, and utilize information (Arrow 1974). A modern corporation exploits the vast amounts of data that it accumulates from marketing, operations, human resources, finance, and other functions to grow faster and be more

Andrea Prat is the Richard Paul Richman Professor of Business at Columbia Business School and professor of economics at Columbia University.

For acknowledgments, sources of research support, and disclosure of the author's material financial relationships, if any, please see http://www.nber.org/chapters/c14022.ack.

productive. This exploitation process depends on the kind of information technology (IT) that is available to the firm. If IT undergoes a revolution, we should expect deep structural changes in the way firms are organized (Milgrom and Roberts 1990).

Agrawal, Gans, and Goldfarb explore the effects that an IT revolution centered on artificial intelligence could have on organizations. Their analysis highlights an insightful distinction between *prediction*, the process of forecasting a state of the world θ given observable information, and *judgment*, the assessment of the effects of the state of the world and the possible action x the organization can take in response to it, namely, the value of the payoff function $u(\theta,x)$.

This is an important point of departure from existing work. Almost all economists—as well as computer scientists and decision scientists—assume that the payoff function $u(\theta,x)$ is known: the decision maker is presumed to have a good sense of how actions and states combine to create outcomes. This assumption, however, is highly unrealistic. The credit card fraud example supplied by the authors is convincing. What is the long-term cost to a bank of approving a fraudulent transaction or labeling a legitimate transaction a suspected fraud?

Organizations can spend resources to improve both their prediction precision and their judgment quality. Agrawal, Gans, and Goldfarb characterize the solution to this optimization problem. Their main result is that, under reasonable assumption, investment in prediction and investment in judgment are complementary (Proposition 2). Investing in prediction makes investment in judgment more beneficial in expected value.

This complementarity suggests that moving from a situation where prediction is prohibitively expensive to one where it is economical should increase the returns to judgment. In this perspective, the AI revolution will lead to an increase in the demand for judgment. However, judgment is an intrinsically different problem—one that cannot be solved through the analysis of big data.

Let me suggest an example. Admissions offices of many universities are turning to AI to choose which applicants to make offers to. Algorithms can be trained on past admissions data. We observe the characteristics of applicants and the grades of past and present students. Leaving aside the censored observations problem arising from the fact that we only see the grades of successful applicants who decide to enroll, we can hope that AI can provide a fairly accurate prediction of an applicant's future grades given his or her observable characteristics. The obvious problem is that we do not know how admitting someone who is likely to get high grades is going to affect the long-term payoff of our university. The latter is a highly complex object that depends on whether our alums become the kind of inspiring, successful, and ethical people that will add to the academic reputation and financial sustainability of our university. There is likely to be a connection

between grades and this long-term goal, but we are not sure what it is. In this setting, Agrawal, Gans, and Goldfarb teach us an important lesson. Progress in AI should induce our university leaders to ask deeper questions about the relationship between student quality and the long-term goals of our higher-learning institutions. These questions cannot be answered within AI, but rather with more theory-driven retrospective approaches or perhaps more qualitative methodologies.

As an organizational economist, I am particularly interested in the implications of Agrawal, Gans, and Goldfarb's model for the study of organizations. First, this chapter highlights the importance of the dynamics of decision-making—a seriously underresearched topic. In a complex world, organizations are not going to immediately collect all the information they could possibly need about all possible contingencies they may face. Bolton and Faure-Grimaud (2009), a source of inspiration for Agrawal, Gans, and Goldfarb, model a decision maker who can "think ahead" about future states of the world in yet unrealized states of nature. They show that the typical decision maker does not want to think through a complete action plan, but rather focus on key short- and medium-term decisions. Agrawal, Gans, and Goldfarb show that Bolton and Faure-Grimaud's ideas are highly relevant for understanding how organizations are likely to respond to changes in information technology.

Second, Agrawal, Gans, and Goldfarb also speak to the organizational economics literature on mission. Dewatripont, Jewitt, and Tirole (1999) develop a model where organizational leaders are agents whose type is unknown, as in Holmstrom's (1999) career concerns paradigm. Each agent is assigned a mission, a set of measured variables that are used to evaluate and reward the agent. Dewatripont, Jewitt, and Tirole identify a tension between selecting a simple one-dimensional mission that will provide the agent with a strong incentive to perform well or a "fuzzy" multidimensional mission that will dampen the agent's incentive to work hard but will more closely mirror the true objective of the organization.

This tension is also present in Agrawal, Gans, and Goldfarb's world. Should we give the organization a mission that is close to a pure prediction problem, like admitting students who will get high grades? The pro is that it will be relatively easy to assess the leader's performance. The con is that the outcome may be weakly related to the organization's ultimate objective. Or should we give the organization a mission that also comprises the judgment problem, like furthering the long-term academic reputation of our university? This mission would be more representative of the organization's ultimate objective, but may make it hard to assess our leaders and give them a weak incentive to adopt new prediction technologies. One possible lesson from Agrawal, Gans, and Goldfarb is that, as the cost of adopting AI goes down, the moral hazard problem connected with judgment becomes rela-

tively more important, thus militating in favor of incentive schemes that reward judgment rather than prediction.

Third, Agrawal, Gans, and Goldfarb's section on reliability touches on an important topic. Is it better to have a technology that returns accurate predictions with a low probability or less accurate predictions with a higher probability? The answer to this question depends on the available judgment technology. Better judgment technology increases the marginal benefit of prediction accuracy rather than prediction frequency. More broadly, this type of analysis can guide the design of AI algorithms. Given the mapping between states, actions, and outcomes, and given the cost of various prediction technologies, what prediction technology should the organization select? A general analysis of this question may require using information theoretical concepts, introduced to economics by Sims (2003).

Fourth, Agrawal, Gans, and Goldfarb show that economic theory can make important contributions to the debate over how AI will affect optimal organization. There is a related area where the interaction between economists and computer scientists can be beneficial. Artificial intelligence typically assumes a stable flow of instances. When a bank develops an AI-based system to detect fraud, it assumes that the available data, which is used to build and test the detection algorithm, comes from the same data-generating process as future data on which the algorithm will be applied. However, the underlying data-generating process is not an exogenously given natural phenomenon: it is the output of a set of human beings who are pursuing their own goals, like maximizing the chance of getting their nonfraudulent application accepted or maximizing their chance of defrauding the bank. These sentient creatures will in the long term respond to the fraud-detection algorithm by modifying their application strategy, for instance, by providing different information or by exerting effort to modify the reported variables. This means that the data-generating process will be subject to a structural change and that this change will be endogenous to the fraud-detection algorithm chosen by the bank. A similar phenomenon occurs in the university admission example discussed above: a whole consulting industry is devoted to understanding admissions criteria and advising applicants on how to maximize their success chances. A change in admissions practices is likely to be reflected in the choices that high school students make.

If the data-generating process is endogenous and depends on the prediction technology adopted by the organization, the judgment problem identified by Agrawal, Gans, and Goldfarb becomes even more complex. The organization must evaluate how other agents will respond to changes in the prediction technology. As, by definition, no data is available about not yet realized data-generating processes, the only way to approach this problem is by estimating a structural model that allows other agents to respond to changes in our prediction technology.

In conclusion, Agrawal, Gans, and Goldfarb make a convincing case that the AI revolution should increase the benefit of improving our judgment ability. They also provide us with a tractable yet powerful framework to understand the interaction between prediction and judgment. Future research should focus on further understanding the implications of improvements in prediction technology on the optimal structure of organizations.

References

Arrow, Kenneth. J. 1974. *The Limits of Organization*. New York: W. W. Norton.

Bolton, P., and A. Faure-Grimaud. 2009. "Thinking Ahead: The Decision Problem." *Review of Economic Studies* 76: 1205–38.

Dewatripont, Mathias, Ian Jewitt, and Jean Tirole. 1999. "The Economics of Career Concerns, Part II: Application to Missions and Accountability of Government Agencies." *Review of Economic Studies* 66 (1): 199–21.

Holmstrom, Bengt. 1999. "Managerial Incentive Problems: A Dynamic Perspective." *Review of Economic Studies* 66 (1): 169–82.

Milgrom, Paul, and John Roberts. 1990. "The Economics of Modern Manufacturing: Technology, Strategy, and Organization." *American Economic Review* June: 511–28.

Sims, Christopher. 2003. "Implications of Rational Inattention." *Journal of Monetary Economics* 50 (3): 665–90.