Exploring the Impact of Artificial Intelligence: Prediction versus Judgment*

Ajay Agrawal, Joshua S. Gans and Avi Goldfarb
University of Toronto and NBER

Draft: 7th July 2017

VERY PRELIMINARY

We interpret recent developments in the field of artificial intelligence (AI) as improvements in prediction technology. In this paper, we explore the consequences of improved prediction in decision-making. To do so, we adapt existing models of decision-making under uncertainty to account for the process of determining payoffs. We label this process of determining the payoffs 'judgment.' There is a risky action, whose payoff depends on the state, and a safe action with the same payoff in every state. Judgment is costly; for each potential state, it requires thought on what the payoff might be. Prediction and judgment are complements as long as the expected payoffs in the two states (before judgment is applied) are not too different. We next consider a tradeoff between prediction frequency and accuracy. We show that as judgment improves, accuracy becomes more important relative to frequency. Finally, we explore the process of gaining experience over time, and show that a seller of predictions cannot extract the full value of the predictions from a buyer.

* Our thanks to Scott Stern and participants at the AEA (Chicago), Harvard Business School, MIT and University of Toronto for helpful comments. Responsibility for all errors remains our own. The latest version of this paper is available at joshuagans.com.

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 labour productivity in those tasks.

The computer revolution has blurred the line between physical and mental tasks. For instance, the invention of the spreadsheet in the late 1970s fundamentally changed the role of book-keepers. 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 labour but the mental productivity of labour 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 (2016) observe that capital substitutes for labour in certain tasks while at the same time technological progress creates new tasks. They make what they call a "natural assumption" that only labour can perform the new tasks as they are more complex than previous ones.[1] Benzell, LaGarda, Kotlikoff, and Sachs (2015) consider the impact of software more explicitly. Their environment has two types of labour – 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

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

be substituted directly for capital but also on the inability of workers themselves to substitute between classes.

In this paper, 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 artificial intelligence (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. 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. Thusfar, 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 paper, 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 take.

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. That 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:

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

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 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 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 transactions. In our model, judgment is the process of determining these payoffs. It is a costly activity.

As the new developments regarding AI all involve making prediction more readily available, we ask, how does judgment and its endogenous application 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? When judgment is a factor, how does this impact on the pricing of AI? Does judgment play a role in the way in which machines learn to predict? And do the answers to these questions change if judgment is an on-going activity versus something that can be gained with experience and become long-lived?

We proceed by first providing supportive evidence for our assumption that recent developments in AI overwhelmingly impact the costs of prediction. 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 the ex ante payoffs to the risky action (before judgment is applied) are expected to be similar. In other words, if the payoffs in the two states are not anticipated to be different before judgment is applied, then prediction and judgment are complements. They are substitutes if and only if anticipated differences in payoffs are high before the decision-maker invests in judging the specific payoffs. After these basic results are established, we show that there is no monotonic relationship between improvements in prediction and the value of judgment. We then separate prediction quality into prediction frequency and prediction accuracy. As judgment improves, accuracy becomes more important relative to frequency. Finally, we allow the decision maker to gain experience over time and learn the payoffs given knowledge of the state, without further need to apply judgment. In this dynamic model, we show that a seller of predictions (i.e. an AI service provider) cannot extract the full value of the predictions from the buyer.

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: Regression trees, neural networks, hidden Markov models, and reinforcement learning. Hastie, Tibshirani, and Friedman's (2009) *The Elements of Statistical Learning* covers similar topics. The 2014 *Journal of Economic Perspectives* symposium on big data covered several of these less

[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.

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, p. 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 non-artificial settings. This led to an "AI winter" in which research funding artificial intelligence projects largely dried up (Markov 2015).

Over the past 10 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 pre-determine 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 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 unknown payoff, 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.[3] 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 state θ_i is v_i. Thus, if before applying judgment, the decision maker does not have any knowledge about the difference in payoffs between the states, then $v_1 = v_2 = v$. To make things more concrete, we assume that there are only two possible payoffs from the risky action, R and r, where $R > S > r$.[4] In this case, we assume that v_i is the probability in state θ_i 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 only. Then the safe action will be chosen if:

---

[3] 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.

[4] 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.

$$\mu(v_1 R + (1-v_1)r) + (1-\mu)(v_2 R + (1-v_2)r) \leq S$$

If $v_i = v$ then this becomes $vR + (1-v)r \leq S$. So that the payoff is: $V_0 = \max\{\mu(v_1 R + (1-v_1)r) + (1-\mu)(v_2 R + (1-v_2)r), S\}$. 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) $\mu(v_1 R + (1-v_1)r) + (1-\mu)(v_2 R + (1-v_2)r) \leq S$

This assumption is made for simplicity only and will not change the qualitative conclusions below.[5] Under A1, in the absence of knowledge of the payoff function or a signal of the state, the decision-maker would choose *S*.

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 (as it would be assumed to do under the usual assumptions of economic rationality). Then the risky action will be chosen (1) if it is the preferred action in both states (which arises with probability $v_1 v_2$); (2) if it is the preferred action in θ_1 but not θ_2 and $\mu R + (1-\mu)r > S$ (with probability $v_1(1-v_2)$); or (3) if it is the preferred action in θ_2 but not θ_1 and $\mu r + (1-\mu)R > S$ (with probability $v_2(1-v_1)$). Thus, the expected payoff is:

$$v_1 v_2 R + v_1(1-v_2)\max\{\mu R + (1-\mu)r, S\} + v_2(1-v_1)\max\{\mu r + (1-\mu)R, S\} + (1-v_1)(1-v_2)S$$

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.

[5] 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.

Result 1. *If $\mu R + (1-\mu)r > S$, 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:

$$v_1 v_2 R + (1 - v_1 v_2)S$$

Note that, absent any cost, full judgment improves the decision-maker's 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 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*. 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. 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. 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 and to the other state or whether to simply choose an action. If an action is chosen, uncertainty is resolved and payoffs are realized and we move to a new decision stage.
4. If judgment is chosen, with probability, $1 - \lambda_{-i}$, they do not find out the payoffs for the risky action in that state, while with probability λ_{-i}, the decision-maker gains this knowledge. The decision-maker then chooses an action, uncertainty is resolved and payoffs are realized and we move to a new decision stage.

Below, when prediction is available, it will become available prior to the beginning of a decision stage. The various parameters are listed in Table 1.

Table 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_i	Prior probability that the payoff in θ_i is R
λ_i	Probability that decision-maker learns the payoff to the risky action θ_i if judgment is applied for one period
δ	Discount factor

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

$$\lambda_i(v_iR+(1-v_i)S)+(1-\lambda_i)\delta\lambda_i(v_iR+(1-v_i)S)+\sum_{t=2}^{\infty}(1-\lambda_i)^t\delta^t\lambda_i(v_iR+(1-v_i)S)$$

$$=\frac{\lambda_i}{1-(1-\lambda_i)\delta}(v_iR+(1-v_i)S)$$

The decision-maker eventually is expected to learn what to do and will earn a higher payoff then 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:

$$\frac{\lambda_1}{1-(1-\lambda_1)\delta}\big(\max\{v_1\big(\mu R+(1-\mu)(v_2R+(1-v_2)r)\big)$$
$$+(1-v_1)\big(\mu r+(1-\mu)(v_2R+(1-v_2)r)\big),S\}\big)$$
$$=\frac{\lambda_1}{1-(1-\lambda_1)\delta}\big(\max\{v_1\big(\mu R+(1-\mu)(v_2R+(1-v_2)r)\big),S\}+(1-v_1)S\big)$$

where the last step follows from (A1). To make exposition simpler, we suppose that $\lambda_1=\lambda_2=\lambda$. In addition, let $\hat{\lambda}=\frac{\lambda}{1-(1-\lambda)\delta}$. $\hat{\lambda}$ can be given a similar interpretation to λ, the quality of judgment. We, therefore, work with $\hat{\lambda}$ until we introduce a dynamic model with experience in Section 6.

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:

$$\hat{\lambda}^2(v_1v_2R+v_1(1-v_2)\max\{\mu R+(1-\mu)r,S\}+v_2(1-v_1)\max\{\mu r+(1-\mu)R,S\}$$
$$+(1-v_1)(1-v_2)S)=\hat{\lambda}^2(v_1v_2R+(1-v_1v_2)S)$$

where the last step follows from (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. To explore this, we assume that $v_1 \geq v_2$. 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:

$$\hat{\lambda}\big(v_1\hat{\lambda}(v_2R + (1 - v_2)\max\{\mu R + (1 - \mu)r, S\}) + (1 - v_1)\max\{\mu r + (1 - \mu)(v_2R + (1 - v_2)r), S\}\big) = \hat{\lambda}\big(v_1\hat{\lambda}(v_2R + (1 - v_2)S) + (1 - v_1)S\big)$$

It is easy to check that it is optimal to search θ_1 first rather than θ_2 if $v_1 \geq v_2$. 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:

$$\hat{\lambda}\big(v_1 \max\{\mu R + (1 - \mu)(v_2R + (1 - v_2)r), S\} + (1 - v_1)\hat{\lambda}(v_2\max\{\mu r + (1 - \mu)R, S\} + (1 - v_2)S)\big) = \hat{\lambda}\big(v_1 \max\{\mu R + (1 - \mu)(v_2R + (1 - v_2)r), S\} + (1 - v_1)\hat{\lambda}S\big)$$

Note, that if $\mu R + (1 - \mu)(v_2R + (1 - v_2)r) \leq S$, then this is option is dominated by not applying judgment at all. Thus, in what follows, whenever we evaluate this option we will assume that $\mu R + (1 - \mu)(v_2R + (1 - v_2)r) > S$. Also note that in this case, it is optimal to search θ_1 first if $v_1 \leq v_2$. Thus, if, as we assume, $v_1 \geq v_2$, for this strategy, it is optimal to search θ_2 first and the resulting payoff is:

$$\hat{\lambda}\big(v_2 \max\{(1 - \mu)R + \mu(v_1R + (1 - v_1)r), S\} + (1 - v_2)\hat{\lambda}S\big)$$

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:

$$\hat{\lambda} > \frac{S}{v_2 \max\{\mu r + (1 - \mu)R, S\} + (1 - v_2)S}$$

$$\hat{\lambda} > \frac{(1 - \mu)R + \mu(v_1R + (1 - v_1)r)}{v_2R + (1 - v_2)\max\{\mu R + (1 - \mu)r, S\}}$$

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 risky 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:

$$\hat{\lambda}\big(v_2 \max\{\mu R + (1-\mu)(v_1 R + (1-v_1)r), S\} + (1-v_2)\hat{\lambda}S\big)$$
$$> \hat{\lambda}(v_2 \max\{(1-\mu)R + \mu(v_1 R + (1-v_1)r), S\} + (1-v_2)S)$$

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

The intuition is similar to Propositions 1 and 2 in Bolton and Faure-Grimaud (2009). The intuition behind this proposition is that 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 other strategy (we will term here J2) is where judgment is applied to one state and if the risk 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:

$$\hat{\lambda}\big(v_1\hat{\lambda}(v_2 R + (1-v_2)S) + (1-v_1)S\big)$$
$$> \hat{\lambda}(v_2 \max\{(1-\mu)R + \mu(v_1 R + (1-v_1)r), S\} + (1-v_2)S)$$
$$\Rightarrow \hat{\lambda}v_1(v_2 R + (1-v_2)S)$$
$$> v_2 \max\{(1-\mu)R + \mu(v_1 R + (1-v_1)r), S\} + (v_1 - v_2)S$$
$$\Rightarrow \hat{\lambda} > \frac{v_2 \max\{(1-\mu)R + \mu(v_1 R + (1-v_1)r), S\} + (v_1 - v_2)S}{v_1(v_2 R + (1-v_2)S)}$$

This is intuitive. Basically, it is only when the efficiency of judgment is sufficiently high, that more judgment is applied.

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:

$$v_i R + (1 - v_i) r > S$$

Thus, the expected payoff is:

$$V_P = \mu \max\{v_1 R + (1 - v_1) r, S\} + (1 - \mu) \max\{v_2 R + (1 - v_2) r, S\}$$

Note that $V_P \geq V_0$ as prediction allows a state specific response to be chosen. If $v_i = v$ then this becomes:

$$V_P = \max\{vR + (1 - v) r, S\}$$

which is the same outcome if there is no judgment or prediction. This generates the following result:

Result 2. *Prediction (in the absence of judgment) is valuable only if the probability that the risky action is optimal differs between states.*

Note that, by (A1), if $v_1 R + (1 - v_1) r > S$ then $v_2 R + (1 - v_2) r < S$. Thus, a necessary condition for $V_P > V_0$ is that $v_1 R + (1 - v_1) r > S$ in which case, $V_P = \mu \max\{v_1 R + (1 - v_1) r, S\} + (1 - \mu) S$.

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:

$$V_{PJ} = \mu\max\{\hat{\lambda}(v_1R + (1 - v_1)S), S\} + (1 - \mu)\max\{\hat{\lambda}(v_2R + (1 - v_2)S), S\}$$

This represents the highest expected payoff possible (net of the costs of judgment).

We are now in a position to prove the following.

Proposition 2. *Under A1 and A2, prediction and judgment are complements (substitutes) if* $v_1 - v_2$ *is small (large). Specifically, the returns to prediction are increasing in λ if:*

$$\hat{\lambda} < \frac{v_2((1-\mu)R + \mu(v_1R + (1 - v_1)r)) + (v_1 - v_2)S}{v_1(v_2R + (1 - v_2)S)}$$

If $v_1R + (1 - v_1)r > S$ *this condition becomes necessary and sufficient.*

PROOF: Prediction and judgment are complements if:

$$\begin{aligned} &V_{PJ} - V_J > V_P - V_0 \\ &\quad\Rightarrow \mu\max\{\hat{\lambda}(v_1R + (1 - v_1)S), S\} + (1 - \mu)\max\{\hat{\lambda}(v_2R + (1 - v_2)S), S\} \\ &\quad - \max\{\hat{\lambda}(v_1\hat{\lambda}(v_2R + (1 - v_2)S) \\ &\quad + (1 - v_1)S), \hat{\lambda}(v_2\max\{(1 - \mu)R + \mu(v_1R + (1 - v_1)r), S\} \\ &\quad + (1 - v_2)S)\} > \mu\max\{v_1R + (1 - v_1)r, S\} + (1 - \mu)S - S \end{aligned}$$

There are numerous cases to consider:

(1) $v_1R + (1 - v_1)r < S$: In this case, the RHS of the inequality is equal to 0 and the remaining terms become:

$$\begin{aligned} &\mu\max\{\hat{\lambda}(v_1R + (1 - v_1)S), S\} + (1 - \mu)\max\{\hat{\lambda}(v_2R + (1 - v_2)S), S\} \\ &\quad > \max\{\hat{\lambda}(v_1\hat{\lambda}(v_2R + (1 - v_2)S) \\ &\quad + (1 - v_1)S), \hat{\lambda}(v_2\max\{(1 - \mu)R + \mu(v_1R + (1 - v_1)r), S\} \\ &\quad + (1 - v_2)S)\} \end{aligned}$$

which always holds. Note that this case is the only one if $v_1 = v_2 = v$, proving the complementarity result.

(2) $v_1R + (1 - v_1)r > S$ and $v_1\hat{\lambda}(v_2R + (1 - v_2)S) > v_2\max\{(1 - \mu)R + \mu(v_1R + (1 - v_1)r), S\}$

$$\hat{\lambda}(\mu(v_1R+(1-v_1)S)+(1-\mu)(v_2R+(1-v_2)S)) \\ -\hat{\lambda}\big(v_1\hat{\lambda}(v_2R+(1-v_2)S)+(1-v_1)S\big) \\ >\mu(v_1R+(1-v_1)r)+(1-\mu)S-S \\ \Rightarrow \hat{\lambda}\Big(\big(\mu v_1+(1-\mu)v_2-v_1v_2\hat{\lambda}\big)(R-S)+v_1(1-\hat{\lambda})S\Big) \\ >\mu(v_1R+(1-v_1)r-S)$$

As $v_1 \to 1$ and $v_2 \to 0$, this becomes

$$\hat{\lambda}S > \mu(R-S)$$

Recall that $v_1\hat{\lambda}(v_2R+(1-v_2)S) > v_2 \max\{(1-\mu)R+\mu(v_1R+(1-v_1)r), S\}$ if:

$$\hat{\lambda} > \frac{v_2 \max\{(1-\mu)R+\mu(v_1R+(1-v_1)r), S\}+(v_1-v_2)S}{v_1(v_2R+(1-v_2)S)}$$

which becomes, as $v_1 \to 1$ and $v_2 \to 0$,

$$\hat{\lambda} > 1$$

Thus, this case is ruled out.

(3) $v_1R+(1-v_1)r > S \Rightarrow v_2R+(1-v_2)r < S$; $\hat{\lambda}\big(v_1\hat{\lambda}(v_2R+(1-v_2)S)+(1-v_1)S\big) < \hat{\lambda}(v_2((1-\mu)R+\mu(v_1R+(1-v_1)r))+(1-v_2)S)$ which implies that $(1-\mu)R+\mu(v_2R+(1-v_2)r) > S$.

$$\hat{\lambda}\big(\mu(v_1R+(1-v_1)S)+(1-\mu)(v_2R+(1-v_2)S)\big) \\ -\hat{\lambda}\big(v_2\big((1-\mu)R+\mu(v_1R+(1-v_1)r)\big)+(1-v_2)S\big) \\ >\mu(v_1R+(1-v_1)r)+(1-\mu)S-S$$

$$\Rightarrow \hat{\lambda}(v_1(1-v_2)R-(v_1-v_2)S-v_2(1-v_1)r) > v_1R+(1-v_1)r-S$$

As $v_1 \to 1$, this becomes:

$$\hat{\lambda}(1-v_2)(R-S) > R-S$$

which cannot hold implying that prediction and judgment are substitutes.

The returns to judgment are increasing in prediction if:

(1) $v_1R+(1-v_1)r < S$ and $v_1\hat{\lambda}(v_2R+(1-v_2)S) > v_2 \max\{(1-\mu)R+\mu(v_1R+(1-v_1)r), S\}$:

Note that the payoff for prediction and judgment will exceed S if $\hat{\lambda} > \frac{S}{v_2R+(1-v_2)S}$.

The condition of the proposition says that in the absence of prediction, the payoff from judgment is: $\hat{\lambda}(v_2 \max\{(1-\mu)R+\mu(v_1R+(1-v_1)r), S\}+(1-v_2)S)$. Note that for this to be feasible, it must be the case that: $\hat{\lambda}(v_2 \max\{(1-\mu)R+\mu(v_1R+(1-v_1)r), S\}+(1-v_2)S) > S$ or $\hat{\lambda} > \frac{S}{v_2 \max\{(1-\mu)R+\mu(v_1R+(1-v_1)r), S\}+(1-v_2)S}$. This requires $(1-\mu)R+\mu(v_1R+(1-v_1)r) > S$ so we can refine this condition to: $\hat{\lambda} > \frac{S}{v_2((1-\mu)R+\mu(v_1R+(1-v_1)r), S)+(1-v_2)S}$. It is easy to see that the RHS of this expression exceeds $\frac{S}{v_2R+(1-v_2)S}$. Thus, whenever judgment alone is feasible, prediction and judgment together is feasible.

Next note that, for $\hat{\lambda} > \frac{S}{v_2((1-\mu)R+\mu(v_1R+(1-v_1)r),S)+(1-v_2)S}$:

$$\frac{\partial(V_{PJ}-V_J)}{\partial\hat{\lambda}} = \mu(v_1R+(1-v_1)S)+(1-\mu)(v_2R+(1-v_2)S) - (v_2(1-\mu)R+\mu(v_1R+(1-v_1)r))-(1-v_2)S)$$

$$v_1(1-v_2)R-(v_1-v_2)S-v_2(1-v_1)r > 0$$

Note that for $v_1 \to v_2$, this is positive and that as v_1 the LHS of this inequality decreases. Finally, as $v_1 \to 1$, this becomes: $(1-v_2)(R-S) > 0$.

(2) $v_1R+(1-v_1)r > S \Longrightarrow v_2R+(1-v_2)r < S$; $\hat{\lambda}\big(v_1\hat{\lambda}(v_2R+(1-v_2)S)+(1-v_1)S\big) < \hat{\lambda}(v_2((1-\mu)R+\mu(v_1R+(1-v_1)r))+(1-v_2)S)$

$$v_1(1-v_2)R-(v_1-v_2)S-v_2(1-v_1)r > 0$$

which is the same condition as (1) above. Thus, we can conclude that in this case, the returns to judgment are increasing in prediction.

Now assume that $v_1R+(1-v_1)r > S$ and $v_1\hat{\lambda}(v_2R+(1-v_2)S) > v_2\max\{(1-\mu)R+\mu(v_1R+(1-v_1)r),S\}$. In this case,

$$\frac{\partial(V_{PJ}-V_J)}{\partial\hat{\lambda}} = \big(\mu v_1+(1-\mu)v_2-2v_1v_2\hat{\lambda}\big)(R-S)+v_1\big(1-2\hat{\lambda}\big)S > 0$$

$$\Rightarrow \frac{(\mu v_1+(1-\mu)v_2)(R-S)+v_1S}{2v_1(v_2(R-S)+S)} > \hat{\lambda}$$

Note, however, if $\hat{\lambda}$ is high enough for $v_1\hat{\lambda}(v_2R+(1-v_2)S) > v_2\max\{(1-\mu)R+\mu(v_1R+(1-v_1)r),S\}$, then it is also high enough that this condition would not hold and the returns to judgment are decreasing in prediction.

The intuition is straightforward. When the prior probability that the risky action is optimal is similar across states, then using prediction to reveal the state does not assist the decision-maker much in the absence of judgment that tells the decision-maker what to do should a given state arise. On the other hand, if it is known that the risky action is optimal if θ_1 should arise, then even if judgment is not costly (i.e., $\hat{\lambda}$ is close to 1), prediction is just as valuable as the decision-maker knows what to do if the state is identified (that is, take the risky action in θ_1 and the safe action in θ_2). Similarly, it is only when judgment is particularly costly (i.e., that it is optimal, in the absence of prediction, to apply judgment to a single state), that prediction can increase the returns to judgment by ensuring that judgment need only be applied to the state that actually arises.

This result yields an important corollary:

Corollary 1. *Under A1 and A2, prediction and judgment are complements if* $v_1 = v_2 = v$.

In other words, if the decision maker does not have strong information about differences in payoffs between the states before exercising judgment, then prediction and judgment are always complements.

4 How does the value of prediction vary with the quality of judgment?

Consider a firm that is trying to work out where to apply prediction machines first. Should it target tasks where judgment is weak (low λ) or strong (high λ)? Specifically, on the continuum of possible qualities of judgment, at what point is the value of prediction (that is, prediction above and beyond what judgment can achieve on its own) highest?

To keep things simple, we assume here that $v_i = v$. From Proposition 1, in the absence of prediction, there are only two *judgment strategies* (that is, ways of applying judgment) that can be optimal: J1 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 and J2 where judgment is applied to one state and if the risk action is optimal, then judgment is applied to the next state; otherwise the safe default is taken immediately. Earlier we determined that, because it involves more application of judgment, J2 will be optimal only if $\hat{\lambda}$ exceeds some threshold which we here label, $\hat{\lambda}_{J2}$:

$$\hat{\lambda} > \hat{\lambda}_{J2}$$

$$\equiv \max\{\frac{\max\{\mu R + (1-\mu)(vR + (1-v)r), S\}}{vR + (1-v)S}, \frac{\sqrt{S\big(4v^2R + S(1 + 2v - 3v^2)\big)} - (1-v)S}{2v(vR + (1-v)S)}\}$$

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 $\mu \rightarrow \frac{S-r}{R-r}$ (its highest possible level consistent with A1 and A2), then $\hat{\lambda}_{J2} \rightarrow 1$. On the other hand, J1 can only arise if its payoff exceeds the default without judgment, S; that is, if $\hat{\lambda}$ exceeds a threshold, $\hat{\lambda}_{J1}$:

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

Note that if $\mu R + (1-\mu)(vR + (1-v)r \leq S$, J1 is never optimal as $\hat{\lambda}_{J1} = 1$. If $\mu R + (1-\mu)(vR + (1-v)r) > S$, note that:

$$\hat{\lambda}_{J2} > \hat{\lambda}_{J1} \Rightarrow \frac{\mu R + (1-\mu)(vR + (1-v)r)}{vR + (1-v)S} > \frac{S}{v\big(\mu R + (1-\mu)(vR + (1-v)r)\big) + (1-v)S}$$
$$\Rightarrow (1-v)S(\mu R + (1-\mu)(vR + (1-v)r) - S)$$
$$> v(RS - \big(\mu R + (1-\mu)(vR + (1-v)r)\big)^2)$$

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 allocations of labour to tasks. If $\hat{\lambda}_{J2} > \hat{\lambda}_{J1}$, then low ordered tasks are allocated to people who use J2, the next tranche of tasks are allocated to those who use J1 while the remainder do not exercise judgment at all. On the other hand, if $\hat{\lambda}_{J2} < \hat{\lambda}_{J1}$, then the low order tasks are allocated to people using J2 while the remainder are allocated to people not using judgment at all.

Having established how tasks are allocated in the absence of prediction, we now turn to consider what the value of prediction is conditional on the difficulty ($\hat{\lambda}$) of a task. In particular, while prediction (if it were free) would enhance the value of all tasks and even for tasks with a low $\hat{\lambda}$, may enable judgment – thereby, expanding the range of tasks where judgment is applied – if it is costly on what sort of tasks will prediction be applied first. That is, how does the incremental value of prediction ($V_{PJ} - V_{J2}$, $V_{PJ} - V_{J1}$ or $V_{PJ} - V_0$) change with $\hat{\lambda}$?

To examine this, note the following:

$$\frac{\partial(V_{PJ} - V_{J2})}{\partial\hat{\lambda}} = v(R - 2\hat{\lambda}(vR + (1-v)S))$$

$$\frac{\partial(V_{PJ} - V_{J1})}{\partial\hat{\lambda}} = v(1-\mu)(1-v)(R-r)$$

$$\frac{\partial(V_{PJ} - V_0)}{\partial\hat{\lambda}} = vR + (1-v)S$$

Note that:

$$\frac{\partial(V_{PJ} - V_0)}{\partial\hat{\lambda}} > \frac{\partial(V_{PJ} - V_{J1})}{\partial\hat{\lambda}} > \frac{\partial(V_{PJ} - V_{J2})}{\partial\hat{\lambda}}$$

where the final inequality can be shown be noting that:

$$\frac{\partial(V_{PJ}-V_{J2})}{\partial\hat{\lambda}}\bigg|_{\hat{\lambda}=\hat{\lambda}_{J2}} = R-2\hat{\lambda}_{J2}(vR+(1-v)S) < (1-\mu)(1-v)(R-r) = \frac{\partial(V_{PJ}-V_{J1})}{\partial\hat{\lambda}}$$

$$\Rightarrow -R(\mu+v-v\mu) < (1-\mu)(1-v)r$$

In other words, as we reduce λ, the returns to prediction rise. The issue is whether the incremental returns to prediction (over judgment alone) are increasing in $\hat{\lambda}_i$. We know that they increase up until $\hat{\lambda}_{J2}$ but at a diminishing rate. In the regime where J2 is optimal, they increase up until the point where:

$$v\left(R-2\hat{\lambda}^*(vR+(1-v)S)\right)=0 \Rightarrow \hat{\lambda} < \hat{\lambda}^* = \frac{R}{2(vR+(1-v)S)}$$

Thus, if $R \geq 2(vR+(1-v)S)$, then the highest incremental value for prediction is where $i=0$ (that is, the least costly tasks in terms of applying judgment). On the other hand, if

$$\hat{\lambda}_{J2} < \hat{\lambda}^* \Rightarrow \frac{R}{2(vR+(1-v)S)}$$

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

then the highest incremental value for prediction is where $\hat{\lambda}=\hat{\lambda}^*$. Otherwise, the highest incremental value for prediction is where $\hat{\lambda}_i=\hat{\lambda}_{J2}$.

Thus, generally the highest value for prediction resides where judgment takes on an intermediate value. If judgment is non-existent, it is that way because it is difficult and, while prediction can enhance value, it is constrained by that difficulty. By contrast where judgment is relatively easy, prediction plays a diminished role in increasing the effectiveness of judgment. That said, there are conditions under which the highest value for prediction resides at an extreme of difficult or easy judgment. These conditions are summarized in the following proposition.

Proposition 3. *The highest incremental value of prediction is where* $\lambda=1$ *if* $(\frac{1}{2}-v)R > (1-v)S$ *and it is where no judgment would otherwise be applied only if J1 is never optimal and* $vR+(1-v)S < \sqrt{S\left(4v^2R+S(1+2v-3v^2)\right)}$.

What this suggests is that the mapping from quality of judgment to the value of prediction is not straightforward. It depends on the relative values of v, S, R, and μ in complicated and nonlinear ways.

5 Prediction Reliability and Judgment

Up until this point we have treated prediction as something that just becomes available. In reality, prediction is something that is available but improves over time. In this regard, designers of AI have an important decision. Do they emphasize the extensive margin (returning a prediction) versus the intensive margin (the reliability of predictions)? In particular, we assume that with probability e, an AI yields a prediction while otherwise, the decision must be made in its absence. We also assume that if an AI predicts a state θ_i, there is a probability (a), independent of the probability a state arises, that it is reliable. Otherwise, with probability $1 - a$, the state predicted is not the state that will actually arise.

We suppose it is the case that if you want a prediction to be reported more often (a higher e), then that only comes about with a sacrifice in reliability (a). That is, you can design the machine prediction technology to be more optimistic (that is, reporting a prediction that the state is positive more often) but at the expense of that prediction being true less often. By contrast, a cautious prediction would be one that it was reported more sparingly but that was more likely to be true when reported. An alternative interpretation of this trade-off is to consider e as not simply a prediction but the ability of a human to parse the prediction (that is, to understand it). In this interpretation, the more a prediction can be explained, the less reliable it becomes. Regardless of interpretation, what interests us here are situations where there is a technical constraint that relates the reliability of prediction to its availability.

To consider this, assume that the technical relationship between e and a is described by $e(a)$; a decreasing, quasi-concave function. What we are interested in is how the effectiveness of human judgment changes the type of prediction technology chosen.

Proposition 4. *As λ increases, the optimal value of e decreases while the optimal value of a increases.*

PROOF: The fact that the AI can be an imperfect predictor impacts upon V_{PJ} which now becomes:

$$V_{PJ} = e\max\{\hat{\lambda}(v(aR + (1-a)(vR + (1-v)r)) + (1-v)S), S\} + (1-e)\max\{\hat{\lambda}(v\hat{\lambda}(vR + (1-v)S) + (1-v)S), \hat{\lambda}(v\max\{(1-\mu)R + \mu(vR + (1-v)r), S\} + (1-v)S)\}$$

For l low, it is not optimal to use judgment as all. But as l rises, it is optimal to use it in conjunction with prediction. When $\hat{\lambda} < \hat{\lambda}_{J2}$, J1 is the optimal (no prediction) strategy but it can readily be seen that in this case,

$$\frac{\partial V_{PJ}/\partial e}{\partial V_{PJ}/\partial a} = \frac{(v(aR + (1-a)(vR + (1-v)r))) - \hat{\lambda}(v\max\{(1-\mu)R + \mu(vR + (1-v)r), S\})\}}{e(v(R - (vR + (1-v)r)))}$$

is independent of $\hat{\lambda}$. However, for $\hat{\lambda} \geq \hat{\lambda}_{J2}$, J2 is optimal so that:

$$\frac{\partial V_{PJ}/\partial e}{\partial V_{PJ}/\partial a} = \frac{(v(aR + (1-a)(vR + (1-v)r))) - (v\hat{\lambda}(vR + (1-v)S))\}}{e(v(R - (vR + (1-v)r)))}$$

which is decreasing in $\hat{\lambda}$. This proves to comparative static in the proposition.

Intuitively, an improvement in judgment (λ) impacts on the expected payoff from prediction and judgment and judgment alone in the same manner. However, it has a larger impact on J2 as judgment is applied potentially to both states prior to a decision being made. This strategy is only likely to be used if no prediction is realized. Thus, the greater is judgment, the lower is the return to having a prediction per se and thus, it falls relative to the return to more reliable prediction. Hence, improved judgment (or easier decisions) will be associated with an AI design that favours reliability over supplying a prediction. As judgment improves, the decision maker needs better predictions but less often.

It is useful to note that the returns to improving *a* by an AI provider are subject to increasing returns. Below some threshold, the decision-maker chooses not to apply judgment and takes the safe action:

$$a \leq \frac{\frac{1}{\hat{\lambda}}S - vR}{v(1-v)(R-r)} - \frac{S+r}{v(R-r)}$$

Above this threshold, the expected discounted payoff to the decision-maker increases linearly in a. This arises because of the complementarity between prediction and judgment. We observe this here because it is something that may impact on the strategic value of AI.

6 Experience and Judgment

The analysis to date takes a static approach to the interdependencies between prediction and judgment. This is because both are required every time a new decision is made. However, while predictions must be provided every period in order to identify the state, judgment is a learning process and it can easily be imagined that once an aspect of the payoff function is known, that knowledge can be applied in future periods. In other words, judgment allows us to give a formal language to the degree of experience a decision-maker has acquired.

Thusfar, we have not distinguished between judgment as acquired in the absence of prediction versus judgment acquired after knowing which state has arisen. This is perhaps a natural starting point when judgment is really the application of thought but when it is gained with experience, it is easy to imagine situations in which judgment applied when the agent has knowledge of the state is different from judgment acquired through a hypothetical thought process (i.e., thinking what you should do if a state could arise). Here we sidestep this issue by assuming that judgment can only be applied when a state arises. Interestingly, this means that experience is tied to having encountered particular states – a somewhat natural interpretation of the word.

In order to gain experience that translates into knowledge of the payoff function, we assume here that the agent must actually experience the state they are applying judgment to and know that they are doing so. This means that without prediction, judgment cannot be applied. Now that the dynamics of the model matter, we return to focusing on λ as the parameter for quality of judgment.

A key difference between the static game and the new dynamic game here is that the decision state and period of time are now one in the same. The new timing of the game is as follows:

0. The AI tells the decision-maker the state that applies that period.
1. The decision-maker chooses whether to apply judgment and to what state or whether to simply choose an action. If an action is chosen, uncertainty is resolved and payoffs are realized and we move to the next time period.
2. If judgment is chosen, with probability, $1 - \lambda_i$, they do not find out the payoffs for the risky action in that state, the decision-maker then takes an action, uncertainty is resolved and payoffs are realized and we move to the next time period. With probability λ_i, the decision-maker gains this knowledge. The decision-maker then

takes an action, uncertainty is resolved and payoffs are realized and we move to the next time period and the decision-maker retains the knowledge of the payoffs.

Let π_i denote the expected present discounted value if the agent already knows what the optimal action is in θ_i. Then:

$$\pi_1 = \mu(vR + (1-v)S + \delta\pi_1) + (1-\mu)\left((1-\lambda)(S + \delta\pi_1) + \lambda\frac{(vR+(1-v)S)}{1-\delta}\right) \Rightarrow \pi_1 = \frac{(\mu + \frac{(1-\mu)\lambda}{1-\delta})(vR + (1-v)S) + (1-\mu)(1-\lambda)S}{1-(1-(1-\mu)\lambda)\delta}$$

$$\pi_2 = (1-\mu)(vR + (1-v)S + \delta\pi_2) + \mu\left((1-\lambda)(S + \delta\pi_2) + \lambda\frac{(vR+(1-v)S)}{1-\delta}\right) \Rightarrow \pi_2 = \frac{(1-\mu + \frac{\mu\lambda}{1-\delta})(vR + (1-v)S) + \mu(1-\lambda)S}{1-(1-\mu\lambda)\delta}$$

Thus, the expected present discount payoff prior to any experience is:

$$V_{PJ} = \mu(\lambda(vR + (1-v)S + \delta\pi_1) + (1-\mu)\big(\lambda(vR + (1-v)S + \delta\pi_2)) + (1-\lambda)(S + \delta V_{PJ})\big)$$

$$\Rightarrow V_{PJ} = \frac{1}{1-(1-\lambda)\delta}\left(\lambda(vR + (1-v)S) + (1-\lambda)S + \delta\mu\frac{(\mu + \frac{(1-\mu)\lambda}{1-\delta})(vR + (1-v)S) + (1-\mu)(1-\lambda)S}{1-(1-(1-\mu)\lambda)\delta} + \delta(1-\mu)\frac{(1-\mu + \frac{\mu\lambda}{1-\delta})(vR + (1-v)S) + \mu(1-\lambda)S}{1-(1-\mu\lambda)\delta}\right)$$

Thus, there is a learning period of uncertain length followed by a period whereby the agent can apply full experience to decisions into the future earning $vR + (1-v)S$ on average.

Pricing AI as a service

Without any judgment or experience, the net present discounted value earned by the agent would be $\frac{1}{1-\delta}S$. Without initial access to an AI, the agent cannot apply judgment and gain experience to improve upon this. This suggests that a monopolist provider of AI could charge a fixed sum of $V_{PJ} - \frac{1}{1-\delta}S$. Moreover, as V_{PJ} is increasing in λ, that provider would want to target agents with judgment ability (or ease) as high as possible first before moving on to harder decisions.

It is easy to imagine that selling an AI which has an ongoing service in terms of providing prediction each period when the environment may be changing, cannot be credibly conducted through an upfront payment. What if, instead, an agent could subscribe to AI as a service paying a price, p, each period? If the AI provider does not have knowledge of the experience level – and indeed, the experience – of each agent, this is a non-trivial pricing problem.

To see this, let us consider the purchase decisions of fully experienced agents who know their payoff function. For some of these agents, they would have found that neither the safe nor risky action is dominated and their per period expected payoff is $\mu R + (1-\mu)S$ or $(1-\mu)R + \mu S$ as the case may be. They can realize these payoffs with prediction but in the absence of prediction, they earn S per period (by A1). Thus, their willingness to pay for prediction is $\mu(R - S)$ or $(1-\mu)(R-S)$. For other agents, their experience has shown them that one of the actions is dominated. Those agents either earn R or S per period but do not need prediction to do so. What this means is that the long-term market for prediction is at most a share $2v(1-v)$ of the original market; that is, prediction is only valuable to those who have found neither action to be dominated.

To keep things simple, in this section we assume that $v = \mu = \frac{1}{2}$. In this case, a fully experienced agent will continue to purchase AI if $\frac{1}{2}(R-S) \geq p$. What about an agent who has learned the payoff for one state but not the other? If they have learned that the risky action is optimal in that state, their expected discounted payoff is π_R where:

$$\pi_R = \frac{1}{2}(R + \delta\pi_R) + \frac{1}{2}\left((1-\lambda)(S + \delta\pi_R) + \lambda\frac{(3R+S-2p))}{4(1-\delta)}\right) - p \Rightarrow \pi_R$$
$$= \frac{R + \frac{\lambda}{4(1-\delta)}(3R + S - 2p) + (1-\lambda)S - 2p}{2\left(1 - \left(1 - \frac{1}{2}\lambda\right)\delta\right)}$$

If this agent did not have access to an AI after this point, their expected discounted payoff would be: $\frac{1}{1-\delta}\max\{\frac{1}{4}(3R + r), S\}$. On the other hand, if they learned the safe action was optimal, their expected discounted payoff is π_S where:

$$\pi_S = \frac{1}{2}(S + \delta\pi_S) + \frac{1}{2}\left((1-\lambda)(S + \delta\pi_S) + \lambda\frac{(R+3S-2p)}{4(1-\delta)}\right) - p \Rightarrow \pi_S$$
$$= \frac{S + \frac{\lambda}{4(1-\delta)}(R + 3S - 2p) + (1-\lambda)S - 2p}{2\left(1 - \left(1 - \frac{1}{2}\lambda\right)\delta\right)}$$

If this agent did not have access to an AI after this point, their expected discounted payoff would be: $\frac{1}{1-\delta}S$. These two options differ both in terms of the payoffs they generate while learning as well as what the potential upside is from moving to full experience. If the agent has learned that the risky action is optimal, this upside is $\frac{3}{4}R + \frac{1}{2}S - p$ while otherwise it is $\frac{1}{2}R + \frac{3}{4}S - p$. Thus, $\pi_R > \pi_S$.

This leads to a pricing dilemma on the part of an AI provider. They have two pricing options: they can set *p* so that $\min\{\pi_R - \frac{1}{1-\delta}\max\left\{\frac{1}{4}(3R + r), S\right\}, \pi_S - \frac{1}{1-\delta}S\} \geq 0$ selling to the entire market or price above this level so that either $\pi_R \geq \frac{1}{1-\delta}\max\left\{\frac{1}{4}(3R + r), S\right\}$ or $\pi_S \geq \frac{1}{1-\delta}S$ and sell to half of the market. The following proposition demonstrates, however, that for a far sighted AI provider, servicing the entire market is the more profitable approach; however, the AI provider does not extract the full value of the prediction despite having perfect knowledge of the state.

Proposition 5. *For δ sufficiently high, the AI provider will set a price equal to:*

$$p = min\,\{\pi_R - \frac{1}{1-\delta} max\left\{\frac{1}{4}(3R + r), S\right\}, \pi_S - \frac{1}{1-\delta}S\}$$

and cover the entire market.

PROOF: Let's first assume that they price to include at this stage. Then, working backwards and taking into account *p*:

$$V_{PJ} = \lambda\frac{1}{2}(R + \delta\pi_R + S + \delta\pi_S) + (1-\lambda)\big(S + \delta V_{PJ}\big) - p \Longrightarrow V_{PJ} = \frac{\lambda\frac{1}{2}(R + \delta\pi_R + S + \delta\pi_S) + (1-\lambda)S - p}{1-(1-\lambda)\delta}$$

The issue is whether an AI provider can charge a price that extracts the maximal rents at this stage. If this could be done, *p* will be:

$$p = \lambda\left(\frac{1}{2}(R + \delta\pi_R + S + \delta\pi_S) + \frac{1}{1-\delta}S\right)$$

Substituting and solving for *p* we have:

$$p = \lambda\frac{(2-\delta)(1-(1-\lambda)\delta)}{4-\delta\big(8-\lambda(6+\lambda)-\delta(4-6\lambda)\big)}(R - S)$$

However, it easy to check that at this price $\pi_S < \frac{1}{1-\delta}S$, so this would not result in full inclusion. Thus, the maximum price would be that $\min\{\pi_R - \frac{1}{1-\delta}\max\left\{\frac{1}{4}(3R + r), S\right\}, \pi_S - \frac{1}{1-\delta}S\} \geq 0$. (a) If $\pi_R - \frac{1}{1-\delta}\max\left\{\frac{1}{4}(3R + r), S\right\} \geq \pi_S - \frac{1}{1-\delta}S$, this price is found by setting $\pi_S = \frac{1}{1-\delta}S$ which yields:

$$p = \frac{\lambda}{2\lambda + 8(1-\delta)}(R - S)$$

This price is less than $\frac{1}{2}(R - S)$; otherwise there will be no demand beyond this stage. (b) If $\pi_R - \frac{1}{1-\delta}\max\left\{\frac{1}{4}(3R + r), S\right\} < \pi_S - \frac{1}{1-\delta}S$, the price is found by setting $\pi_R = \frac{1}{4(1-\delta)}(3R + r)$ which yields:

$$p = \frac{1}{2\lambda + 8(1-\delta)}(\lambda(3(1-\delta)(R - S) + (S - r)\delta) - 2(r + R - 2S)(1-\delta))$$

By assumption this price must also be less than $\frac{1}{2}(R - S)$ as it is lower than the price in the previous case.

By contrast, the exclusion strategy would price so that $\pi_R = \frac{1}{1-\delta}\max\left\{\frac{1}{4}(3R + r), S\right\}$. (a) Consider first the price where $\pi_R = \frac{1}{1-\delta}S$ and note that:

$$\begin{aligned} p = \tfrac{\lambda}{1-\delta}\left(\tfrac{1}{2}\left(R + \delta\pi_R + S + \delta\tfrac{1}{1-\delta}S\right) + \tfrac{1}{1-\delta}S\right) \\ > \tfrac{1}{1-(1-\lambda)\delta}\lambda\left(\tfrac{1}{2}(R + S) + \tfrac{1+\delta}{1-\delta}S\right) + \tfrac{1}{2(1-\delta)}\lambda\left(\tfrac{1}{2}(R + S) + \tfrac{1+\delta}{1-\delta}S\right) \\ \Longrightarrow (1 - (1-\lambda)\delta - \tfrac{1}{2}(3(1-\delta) + \lambda\delta))(R + S) + (1 - (1-\lambda)\delta)\delta\pi_R \\ > ((3(1-\delta) + \lambda\delta)(1+\delta) - (1 - (1-\lambda)\delta)(2-\delta))S \end{aligned}$$

which always holds for δ sufficiently high.

Intuitively, when some initial judgment is complete, there is either good news (in that the risky strategy is optimal) or bad news (in which it is not). An inclusion strategy requires price to be low enough that following bad news, learning still occurs. However, while the upside potential for the user following good news is higher than that following bad news, the value of prediction after full experience is gained is the same. Thus, the AI provider has no mechanism by which they can share in the upside. In the absence of that mechanism, they choose to price low and not exclude any users at this stage. Half of the users eventually opt out when they find that either the safe or risky action is dominant. What this means is that an AI provider who cannot implement upfront pricing is restricted in the value they can appropriate. While learning can yield good or bad news to the decision-maker, good news may cause prediction to lose its value as the decision-maker discovers the risky action is dominant. Thus, the AI provider must sacrifice rents in order to ensure that they can capture some rents as the decision-maker gains experience.

Can versioning – selling an AI product which has lower performance – improve this outcome for the AI provider? The intuition would be that until they are fully experienced, users will purchase the lower performing product allowing the AI provider to charge more in the long-term.

The downside is a lower performing product may slow the gathering of experience and push that long-term out further. This is something we might be able to work out but leave for future work at the moment.

Experience through experimentation

Using an experience frame to judgment suggests an alternative way of 'learning' the reward function: experimentation. In particular, when coupled with prediction, a decision-maker could, by choosing the risky action, evaluate whether that is the right action for that state. The expected cost would be $S - vR - (1-v)r$. In this conception, we have the following:

$$\pi_1 = \mu(vR + (1-v)S + \delta\pi_1) + (1-\mu)\left(vR + (1-v)r + \delta\frac{(vR+(1-v)S)}{1-\delta}\right) \Rightarrow \pi_1 = \frac{\left(\mu + \frac{(1-\mu)\delta}{1-\delta}\right)(vR + (1-v)S) + (1-\mu)(vR + (1-v)r)}{1-\mu\delta}$$

$$\pi_2 = (1-\mu)(vR + (1-v)S + \delta\pi_2) + \mu\left(vR + (1-v)r + \delta\frac{(vR+(1-v)S)}{1-\delta}\right) \Rightarrow \pi_2 = \frac{(1-\mu + \frac{\mu\delta}{1-\delta})(vR + (1-v)S) + \mu(vR + (1-v)r)}{1-(1-\mu)\delta}$$

Thus, the expected present discount payoff prior to any experience is:

$$V_{PJ} = \mu(vR + (1-v)r + \delta\pi_1) + (1-\mu)(vR + (1-v)r + \delta\pi_2))$$

$$\Rightarrow V_{PJ} = vR + (1-v)r + \delta(\mu\pi_1 + (1-\mu)\pi_2)$$

The convenient property of this frame is that it relates the cost of judgment explicitly to the expected cost of experimentation. In particular, as r decreases, experimentation becomes more costly. This may be a way of an AI to learn the reward function independent of a human. It could engage in experimentation and then learn what to do should a particular state arise. Other than that, at this stage, we have no further insights from this approach.

7 Complexity and Automation (STILL INCOMPLETE)

The literature on automation is sometimes synonymous with AI. However, there are key differences. In our conception, AI is all about prediction. Such prediction can enable the full automation of tasks but it is not a given that it will. In order for a task to be fully automated, judgment needs to be codified. That is, you need to engineer the reward function into capital that can operate without an labour input. This marks our key departure from the literature on automation. In general, automation is seen as a capital investment that is either available or is not. In Acemoglu and Restrepo (2017), it can also emerge as the result of innovation that is produced by specialist scientists. But here judgment can only be codified if a human worker has mapped the objective function. This comes through experience and critically, the efficiency by which that experience is generated is impacted on by AIs through prediction.

We now consider a situation in which there are $N > 2$ states in order to explore the impact of complexity on prediction and judgment. This has the advantage of allowing us to consider complexity explicitly – this is useful as it is often assumed that more ‘complex’ tasks can only be done by humans and cannot be automated (for example, in Acemoglu and Restrepo, 2017). The question we ask is whether more complex tasks will be undertaken by humans and, if so, why?

In order *N*-state model, the probability of state i is μ_i. To keep things simple, we assume that $v_i = v$. In this situation, the expected present discounted value when both prediction and judgment are available is:

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

Similarly, it is easy to see that $V_P = S = V_0$ as $vR + (1 - v)r \leq S$.

For the experience model, for one task, with $\mu_i = 1/N$, then

$$V_{PJ} = \frac{(N - (N - 1)\delta)\lambda(vR + (1 - v)S) + N(1 - \delta)(1 - \lambda)S}{(1 - \delta)(N - (N - \lambda)\delta)}$$

TBC

8 Case: Radiology

In 2016, Geoff Hinton – one of the pioneers of deep learning neutral 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[6]) 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 amongst radiologists and pathologists (Jha and Topol, 2016). What does the model in this paper 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% chance of being benign, 33.3% chance of being malignant, and a 0.1% of not being real.*[7]

The action is whether some intervention is needed. For instance, if a potential tumor is identified in a non-invasive 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 non-invasive 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.

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

[6] "You did not go to medical school to measure lung nodules." http://www.medscape.com/viewarticle/863127#vp_2
[7] http://www.medscape.com/viewarticle/863127#vp_3

involve new job classes that merge diagnostic information such as a combined radiologist/pathologist (Jha and Topol, 2016).

9 Conclusions

In this paper, 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, unless there isn't that much useful to learn about payoffs. Thus, predictions are more valuable when the payoffs are understood, and understanding payoffs is more valuable when the state is known. At the same time, as the quality of judgment varies, the incremental value of predictions to the decision maker is non-linear. 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 a dynamic model in which judgment is accumulated with experience.

Overall, this paper emphasized that AI is a prediction technology, and predictions cannot be valued in the absence of knowing how payoffs arise. Of course, there may be many other consequences of improved prediction, and we leave those for future work.

10 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, MIT.

Alpaydin, Ethem. 2010. *Introduction to Machine Learning*. Second Edition. MIT Press, Cambridge MA.

Aghion, Philippe, and Jean Tirole. 1997. Formal and real authority in organizations. *Journal of political economy*, 1-29.

Athey, S., J.S. Gans, S. Schaefer and S. Stern. 1994. The Allocation of Decisions in Organizations. *unpublished manuscript*, Stanford.

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.

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, LaGarda, Kotlikoff, and Sachs. 2015. Robots Are Us: Some Economics of Human Replacement. NBER Working Paper No. 20941.

Bernanke, Ben S. 1983. Irreversibility, Uncertainty, and Cyclical Investment. *The Quarterly Journal of Economics* 98(1): 85-106.

Bolton, P. and A. Faure-Grimaud (2009), "Thinking Ahead: The Decision Problem," *Review of Economic Studies*, 76: 1205-1238.

Brynjolfsson, Erik, and Andrew McAfee. 2014. *The Second Machine Age*. W.W. Norton, New York.

Domingos, Pedro. 2015. *The Master Algorithm.* Basic Books, New York.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data mining, Inference, and Prediciton*. Second Edition. Springer, New York.

Hawkins, Jeff. 2004. *On Intelligence*. Times Books, New York.

Jha, S. 2016. Will computers replace radiologists? *Medscape*. 30 December 2016; http://www.medscape.com/viewarticle/863127#vp_1

Jha, S. and E.J. Topol. 2016. Adapting to Artificial Intelligence: Radiologists and Pathologists as Information Specialists. *JAMA*. 316(22):2353-2354.

Kaplan, Jerry. 2015. *Humans Need Not Apply.* Yale University Press, New Haven.

Lusted, L.B. 1960. Logical analysis in roentgen diagnosis. *Radiology*, 74: 178-193.

Markov, John. 2015. *Machines of Loving Grace*. HarperCollins Publishers, New York.

Nordhaus, William. 2016. Are We Approaching an Economic Singularity. Working paper, Yale University.

Ng, Andrew. 2016. What Artificial Intelligence Can and Can't Do Right Now. *Harvard Business Review Online*. https://hbr.org/2016/11/what-artificial-intelligence-can-and-cant-do-right-now. Accessed December 8 2016.

Varian, Hal R. 2014. Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, 28(2): 3-28.