In the preceding chapter we considered some of the purposes to which machine learning might be put—for example, to predict the topic of a court case given words that the judge used in the court’s written judgment—and we described pattern finding as the method behind prediction. More important than the method however is the goal, in this example “to predict the topic,” and in particular the keyword predict. We introduced that word in the preceding chapter to begin to draw attention to how computer scientists use it when they engineer machine learning systems. In machine learning systems, predictive accuracy is the be-all and end-all—the way to formulate questions, the basis of learning algorithms, and the metric by which the systems are judged. In this chapter we consider prediction, both in Holmes’s view of law and in the machine learning approach to computing.

In Holmes’s view of law, prediction is central. His answer to the question, What constitutes the law? has become one of the most famous epigrams in all of law:

The prophecies of what the courts will do in fact, and nothing more pretentious, are what I mean by the law.¹

Holmes’s interest in the logic and philosophy of probability and statistics has come more to light thanks to recent scholarship²; he immersed
himself early in his career in those subjects. Holmes’s use of the word “prophecy” was deliberate. It accorded with his overall view of law by getting away from the scientific and rational overtones of “prediction,” even as he used that word too. Arguably, given how elusive explanations have been of how machine learning systems arrive at the predictions they make, “prophecy” is a good term in that context too.

We expand in this chapter on Holmes’s idea that prophecies constitute the law, and then we return to prediction in machine learning.

### 5.1 Prophecies Are What Law Is

Holmes’s famous epigram has been widely repeated, but it is not widely understood. Taken in isolation from *The Path of the Law*, where Holmes set it down, and in isolation from Holmes’s development as a thinker, it might sound like no more than a piece of pragmatic advice to a practicing lawyer: don’t get carried away by the cleverness of your syllogisms; ask yourself, instead, what the judge is going to do in your client’s case. If that is all it meant, then it would be good advice, but it would not be a concept of law. Holmes had in mind a concept of law. The epigram needs to be read in context:

> The confusion with which I am dealing besets confessedly legal conceptions. Take the fundamental question, What constitutes the law? You will find some text writers telling you that it is something different from what is decided by the courts of Massachusetts or England, that it is a system of reason, that it is a deduction from principles of ethics or admitted axioms or what not, which may or may not coincide with the decisions. But if we take the view of our friend the bad man we shall find that he does not care two straws for the axioms or deductions, but that he does want to know what the Massachusetts or English courts are likely to do in fact. I am much of this mind. The prophecies of what the courts will do in fact, and nothing more pretentious, are what I mean by the law.³

Holmes was contrasting “law as prophecy” to “law as axioms and deductions.” He saw an inductive approach to law—the pattern finding approach that starts with data or experience—not just to improve upon or augment legal formalism. He saw it as a corrective. The declaration in his dissent in the *Lochner* case a few years after *The Path of the Law* that “[g]eneral propositions do not decide concrete cases”⁴ was not just to
say that the formal, deductive approach is insufficient; it was to say that
formalism gets in the way.

The centrality of concern for Holmes was the reality of decision, the
output that a court might produce. The realism or positivism in this
understanding of law contrasted with the formalist school that had long
prevailed. To shift the concern of lawyers in this way was to lessen the
role of doctrine, of formal rules, and to open a vista of social and histor-
ical considerations heretofore not part of the law school curriculum and
ignored, or at any rate not publicly acknowledged, by lawyers or judges.
Jurists have been divided ever since as to whether the shift of conception
was for better or worse. Whatever one’s assessment of it, the concept of
law as Holmes expressed it continues to influence the law.

There is more still to Holmes’s epigram about prophecies. True,
the contrast it entails between the inductive method and the deductive
method alone has revolutionary implications. But Holmes was not merely
concerned with what method “our friend the bad man” (or indeed the
bad man’s lawyer) should employ to predict the outcome of a case. He
wasn’t writing a law practice handbook. He was interested in individual
encounters with the law to be sure, but this was because he sought to
reach a general understanding of law as a system. Holmes’s invocation of
prophecies, like his use of terms from logic and mathematics, was memo-
rable use of language, but it was more than rhetoric: it was at the core of
Holmes’s definition of law. He referred to the law as “systematized predic-
tion.”6 This was to apply the term “prediction” broadly—indeed across
the legal system as a whole. Holmes was not sparing in his use of the word
“prophecy” when defining the law. The word “prophesy” or its derivates
appear nine times in The Path of the Law.7 He used it in the same sense
when writing for the Supreme Court.8 Holmes’s concern with prediction
is traceable in his other writings too.9 The heart of Holmes’s insight, and
what has so affected jurisprudence since, is that the law is prediction.10
Prophecy does not refer solely to the method for predicting what courts
will do. Prophecy is what constitutes the law.

Prophecy of what, by whom, and on the basis of which input data?

Holmes gave several illustrations. For example, he famously described
the law of contract as revolving around prediction: “The duty to keep a
contract at common law means a prediction that you must pay damages
if you do not keep it, and nothing else.”11 He stated his main thesis in
similar terms: “a legal duty so called is nothing but a prediction that if a
man does or omits certain things he will be made to suffer in this or that
way by judgment of the court.” This is a statement about “legal duty” irrespective of the content of the duty. It thus just as well describes any duty that exists in the legal system.

We think that Holmes’s concept of law as prediction indeed is comprehensive. Many jurists don’t see it that way. Considering how Holmes understood law to relate to decisions taken by courts, one sees why his concept of law-as-prophecy often has received a more limited interpretation.

Holmes wrote that “the object of [studying law], then, is prediction, the prediction of the incidence of the public force through the instrumentality of the courts.” Making the equation directly, he wrote, “Law is a statement of the circumstances, in which the public force will be brought to bear upon men through the courts…”; a “word commonly confined to such prophecies… addressed to persons living within the power of the courts.” It is often assumed that Holmes’s description here does not account for the decisions of the highest courts in a jurisdiction, courts whose decisions are final. After all, in a system of hierarchy, the organ at the apex expects its commands to be obeyed. To call decisions that emanate from such quarters “predictions” seems to ignore the reality of how a court system works. In a well-functioning legal system, a judgment by a final court of appeal saying, for example, that the police are to release such and such a prisoner, should lead almost certainly to that outcome. The court commands it; the prisoner is released.

In two respects, however, one perhaps trivial but the other assuredly significant, the highest court’s statements, too, belong to the concept of law as prophecy.

First, even in a well-functioning legal system, the court’s decision is still only a prediction. As outlandish as the situation would be in which the police ignored the highest court, it is a physical possibility. A statistician might say that the probability is very high (say, 99.9999%) that the highest court’s judgment that the law requires the prisoner to be released will in fact result in an exercise of public power in accordance with that judgment. We will say more below about the relation between probability and prediction. Leaving that relation aside for the moment, a judgment even of the highest court is a prediction in Holmes’s sense. It is a prediction in this way: the implementation of a judgment by agents of public power is an act of translation, and in that act the possibility exists for greater or lesser divergence from the best understanding of what the judge commanded. So the definition of law as prophecy is instanced in
the chance that the “public force” will not properly implement the judicial decision. In a well-functioning legal system, the chance is remote. In legal systems that don’t function well, the predictive character of final judgments is more immediate, because the risk in those systems is greater that the public force will not properly implement the courts’ commands. “Finality” in some judicial systems is more formal than real. 16

The further respect in which the concept of law as prophecy is comprehensive comes to view when we consider how judges decide cases and how advocates argue them. In deciding a case, a judge will have in mind how that decision is likely to be interpreted, relied upon, or rejected, by future courts and academics and public opinion, as well as by the instruments of public force. The barrister, for her part, in deciding what line of argument to pursue, will have in mind how the judge might be influenced, and this in turn requires consideration of the judge’s predictions about posterity. Holmes thus described a case after the highest court had decided it as still in the “early stages of law.” 17 As Kellogg puts it, Holmes situated a case “not according to its place on the docket but rather in the continuum of inquiry into a broader problem.” 18

The law is a self-referential system, whose rules and norms are consequences of predictions of what those rules and norms might be. 19 Some people participate in the system in a basic and episodic way, for example the “bad man” who simply wants advice from a solicitor about likely outcomes in respect of his situation. Some people participate in a formative way. The apex example is the judge of the final court of appeal whose predictions about the future of the legal system are embodied in a judgment which she expects as a consequence of her authority in the legal system to be a perfect prediction of the exercise of public power in respect of the case. But her outlook, indeed her self-regard as a judge, entails more than that; a judge does more than participate in disconnected episodes of judging: she hopes that any judgment she gives in a case, because she strives for judgments that withstand the test of time, will be a more or less accurate prediction of how a future case bearing more or less likeness to the case will be decided. The judge describes her judgment as command, not as prophecy; but the process leading to it, and the process as the judge hopes it will unfold in the future, is predictive. Law-as-prophecy, understood this way, has no gap.

Holmes’s claim, as we understand it, holds law to be predictive through and through. Prophecy is what law is made of. The predictive character of law, in this constitutive sense, is visible in the process of judicial decision,
regardless what level of the judiciary is deciding; and it is visible in all other forms of legal assertion as well. Prophecy embraces all parts of legal process.

So everyone who touches the law is making predictions, from the self-interested “bad man” to the judge in the highest court of the land, and they make predictions about the full range of possible outcomes. The experience that influences their predictions, as we saw in Chapter 3, Holmes understood to be wide, and the new situations that they make predictions about are unlimited. People on Holmes’s path of the law thus engage in tasks much broader than standard machine learning tasks. As we also discussed in Chapter 3, machine inputs, while they consist in very large datasets (“Big Data”), are limited to inputs that have been imparted a considerable degree of structure—a degree of structure almost certainly lacking in the wider (and wilder) environment from which experience might be drawn. Machine outputs are correspondingly limited as well. Let us now further explore the machine learning side of the analogy—and its limits.

5.2 Prediction Is What Machine Learning Output Is

Holmes, writing in 1897, obviously did not have machine learning in mind. Nevertheless, his idea that prophecy constitutes the law has remarkable resonance with machine learning, a mechanism of computing that, like law as Holmes understood it, is best understood as constituted by prediction.

The word *prediction* is a term of art in machine learning. It is used like this:

In a typical scenario, we have an outcome measurement, usually quantitative (such as a stock price) or categorical (such as heart attack/no heart attack), that we wish to predict based on a set of features (such as diet and clinical measurements). We have a training set of data, in which we observe the outcome and feature measurements for a set of objects (such as people). Using this data we build a prediction model, or learner, which will enable us to predict the outcome for new unseen objects. A good learner is one that accurately predicts such an outcome.
Though in common parlance the term “prediction” means forecasts—that is to say, statements about future events—in machine learning the term has a wider meaning. We have touched on the wider meaning in Chapter 4 and at the opening of the present chapter. Let us delve a little more into that wider meaning now.

It is true that some machine learning outputs are “prediction” in the sense in which laypersons typically speak: “Storm Oliver will make landfall in North Carolina” or “the stock price will rise 10% within six months.” Other outputs are not predictions in the layperson’s sense. Indeed, the main purposes for which machine learning is used do not involve predictions of that kind—purposes like classifying court cases by topic or controlling an autonomous vehicle. Whatever the purposes for which it is used, machine learning involves “prediction” of the more general kind computer scientists denote with that term.

The essential feature of prediction in machine learning is that it should concern “the outcome for new unseen objects,” i.e. for objects not in the training set. Thus, for example, if the training set consists of labelled photographs, and if we treat the pixels of the photograph as features and the label as the outcome, then it is prediction when the machine learning system is given a new photograph as input data and it outputs the label “kitten.” In machine learning prediction, “pre-” simply refers to before the true outcome measurement has been revealed to the machine learning system. The sense of “pre-” in “prediction” holds even though other parties might well already know the outcome. For example, the computer scientist might well already know that the new photograph is of a tiger, not a kitten. That assignment of label-to-picture has already happened, but the machine learning system has not been told about it at the point in time when the system is asked to predict. Philosophers of science use the terms “postdiction” or “retrodiction” to refer to predicting things that have already happened. These words are not used in the machine learning community, but the concept behind them is much what that community has in mind when it talks about prediction.

A significant part of the craft of machine learning is to formulate a task as a prediction problem. We have already described how labelling a photograph can be described as prediction. A great many other examples may be given. Translation can be cast as prediction: “predict the French version of a sentence, given the English text,” where the training set is a human-translated corpus of sentences. Handwriting synthesis can as well. Given a dataset of handwritten text, recorded as the movements of a pen...
nib, and given the same text transcribed into text in a word processor, the task of handwriting synthesis can be cast as prediction: “predict the movements of a pen nib, given text from a word processor.” As Judea Pearl observed in the interview with which we opened Chapter 4, it is truly remarkable how many tasks can be formulated this way. In the social sciences, it is “a rather new epistemological approach […] and research agendas based on predictive inference are just starting to emerge.” A theory of law based on predictive inference, however, emerged over a century ago: Holmes theorized law to be constituted by prophecy. So too might we say that machine learning is constituted by prediction.

Moreover, prediction is not just the way that machine learning tasks are formulated. It is also the benchmark by which we train and evaluate machine learning systems in the performance of their tasks. The goal of training is to produce a “good learner,” i.e. a system that makes accurate predictions. Training is achieved by measuring the difference between the machine’s predictions (or postdictions, as the philosophers would say) and the actual outcomes in the training dataset; and iteratively tweaking the machine’s parameter values so as to minimize the difference. The machine that reliably labels tigers as “tigers” has learned well and, at least for that modest task, needs no more tweaking. The machine that labels a tiger as a “kitten” needs tweaking. The one that labels a tiger as “the forests of the night,” though laudable if its task had been to predict settings in which tigers are found in the poetry of William Blake, needs some further tweaking still to perform the task of labeling animals. This process of iterative tweaking, as we noted in Chapter 2, is what is known as gradient descent, the backbone of modern machine learning. Thus, a mechanism of induction, not algorithmic logic, is at the heart of machine learning, much as Holmes’s “inductive turn” is at the heart of his revolutionary idea of law.

It is not machine learning’s fundamental characteristic that it can be used to forecast future events—when will the next hurricane occur, where will it make landfall? One doesn’t need machine learning to make forecasts. One can make forecasts about hurricanes and the like with dice or by sacrificing a sheep (or by consulting a flock of dyspeptic parrots). One can also make such forecasts with classic algorithms, by simulating dynamical systems derived from atmospheric science. This sort of prediction is not the fundamental characteristic of machine learning.
The fundamental characteristic of machine learning is that the system is trained using a dataset consisting of examples of input features and outcome measurements; until, through the process of gradient descent, the machine’s parameter values are so refined that the machine’s predictions, when we give it further inputs, differ only minimally from the actual outcomes in the training dataset. Judges, litigants, and their lawyers certainly try to align their predictive statements of law with what they discern to be the relevant pattern in law’s input data, that is to say in the collective experience that shapes the law. It is equally the case, in Holmes’s understanding of the law, that we do not test court judgments by comparing against stipulated “correct” labels the way our spam email or tiger detector was tested. Judgments are, however, tested against future judgments. This is to the point we made earlier about the judge’s aim that her judgments withstand the test of time. The test is whether future judgments show her judgment to have been an accurate prediction, or at least not so far off as to be set aside and forgotten.

A machine learning system must be trained on a dataset of input features and outcome measurements. This is in contrast to the classic algorithmic approach, which starts instead from rules. For example the classic approach to forecasting the weather works by solving equations that describe how the atmosphere and oceans behave; it is based on scientific laws (which are presumably the result of codifying data from earlier experiments and observation). Just as machine learning rejects rules and starts instead with training data, Holmes rejected the idea that law is deriving outcomes based on general principles, and he cast it instead as a prediction problem—prophesying what a court will do—to be performed on the basis of experience.

5.3 Limits of the Analogy

As we noted in Chapter 3, the predictions made by a machine learning system must have the same form as the outcomes in the training dataset, and the input data for the object to be predicted must have the same form as objects already seen. In earlier applications of machine learning, “same form” was very narrowly construed: for example, the training set for the ImageNet challenge consists of images paired with labels; the machine learning task is to predict which one of these previously seen labels is the best fit for a new image, and the new image is required to be the same dimensions as all the examples in the training set. Human ability to
make predictions about novel situations is far ahead of that of machines. A human lawyer can extrapolate from experience and make predictions about new cases that don’t conform to a narrow definition of “cases similar to those already seen.” The distance is closing, however, as researchers develop techniques to broaden the meaning of “same form.” For example, an image captioning system is now able to generate descriptions of images, rather than just repeat labels it has already seen. Thus, it is well within their grasp for machines to label an image as “tiger on fire in a forest,” but they are still a long way, probably, from describing, as the poet did, the tiger’s “fearful symmetry.”

There is a more significant difference between predictions in machine learning and in law. In machine learning, the paradigm is that there is something for the learning agent—i.e., the machine—to learn. The thing to learn is data, something that is given, not a changing environment affected by numerous factors—including by the learning agent. A machine for translating English to French can be trained using a human-translated corpus of texts, and its translations can be evaluated by how well they match the human translation. Whatever translations the machine comes up with they do not alter the English nor French languages. In law, by contrast, the judgment in a case becomes part of the body of experience to be used in subsequent cases. Herein, we think, Holmes’s concept of law as a system constituted from prediction may hold lessons for machine learning. In Chapters 6–8, we will consider some challenges that machine learning faces, and possible lessons from Holmes, as we discuss “explainability” of machine learning outputs and outputs that may have invidious effects because they reflect patterns that emerge from the data (such as patterns of racial or gender discrimination). In Chapter 9, we will suggest that Holmes, because he understood law to be a self-referential process in which each new prediction shapes future predictions, might point the way for future advances in machine learning.

Before we get to the challenges of machine learning and possible lessons for the future from Holmes, we will briefly consider a question that prediction raises: does prediction, whether as the constitutive element of law or as the output of machine learning, necessarily involve the assessment of probabilities?
5.4 Probabilistic Reasoning and Prediction

“For the rational study of the law the blackletter man may be the man of the present, but the man of the future is the man of statistics,” said Holmes. It is not certain that Holmes thought that the predictive character of law necessarily entails a probabilistic character for law. He was certainly interested in probability. In the time after his Civil War service, a period that Frederic Kellogg closely examined in Oliver Wendell Holmes, Jr. and Legal Logic, Holmes studied theories of probability and was much engaged in discussions about the phenomenon, including how it relates to logic and syllogism. Later, as a judge, he recognized the part played by probability in commercial life, for example in the functioning of the futures markets. In personal correspondence, Holmes said that early in his life he had learned “that I must not say necessary about the universe, that we don’t know whether anything is necessary or not. So that I describe myself as a bettabilitarian. I believe that we can bet on the behavior of the universe...” Holmes would have been comfortable with the idea that law, in its character as prediction, concerned probability as well. Some jurists indeed have discerned in Holmes’s idea of law-as-prophecy just such a link.

Predictions made by machine learning are not inherently probabilistic. For example, the “k nearest neighbors” machine learning algorithm is simply “To predict the outcome for a new case, find the k most similar cases in the dataset, find their average outcome, and report this as the prediction.” The system predicts a value, which may or may not turn out to be correct. Modern machine learning systems such as neural networks, however, are typically designed to generate predictions using the language of probability, for example “the probability that this given input image depicts a kitten is 93%.”

Separately, we can classify machine learning systems by whether or not they employ probabilistic reasoning to generate their predictions:

[One type of] Machine Learning seeks to learn [probabilistic] models of data: define a space of possible models, learn the parameters and structure of the models from data; make predictions and decisions. [The other type of] Machine Learning is a toolbox of methods for processing data: feed the data into one of many possible methods; choose methods that have good theoretical or empirical performance; make predictions and decisions.
Are legal predictions expressed in the language of probability? Lawyers serving clients do not always give probability assessments when they give predictions, but sometimes they do. Some clients need such an assessment for purposes of internal controls, financial reporting, and the like. Others ask for it for help in strategizing around legal risk. Modern empirical turns in law scholarship, it may be added, are much concerned with statistics. Attaching a probability to a prediction of a legal outcome is an inexact exercise, but it is not unfamiliar to lawyers.

Holmes, when he referred to the prophecies of what courts will do, is often read to mean that the law should be made readily predictable. Though we don’t doubt he preferred stable judges to erratic ones, we don’t see that that was Holmes’s point. Courts whose decisions are hard to predict are no less sources of legal decision. Even when the lawyer has the privilege to argue in front of a “good” judge, whom for present purposes we define as a judge whose decisions are easy to predict, the closer the legal question, the harder it is to predict the answer. It is inherent that lawyers will be more confident in some of their predictions than in others.

Judges, practically by definition of their role as legal authorities, do not proffer a view as to the chances that their judgments are correct. It is hard to see how the process of judgment would keep the confidence of society, if every judgment were issued with a \( p \)-value. Yet reading judgments through a realist’s glasses, one may discern indicia of how likely it is that the judgment will be understood in the future to have stated the law. Judges do not shy from describing some cases as clear ones; others as close ones. They don’t call it hedging, but that’s very much what it’s like. When a judge refers to how finely balanced such and such a question was, it has the effect of qualifying the judgment. It thus may be that one can infer from a judgment’s text how much confidence one should have in the judgment as a prediction of future results. The text, even where it does not express anything in terms about the closeness of a case, still may give clues. The structure of the reasoning may be a clue: the more complex and particularistic a judge’s reasoning, the more the judgment might be questioned, or at least limited in its future application. Textual clues permit an inference as to how confident one should be that the judgment accurately reflects a pattern in the experience that was the input behind it.

Does the law use probabilistic reasoning to arrive at a prediction? In other words, once a judgment has been made and it becomes part of the
body of legal experience, do lawyers and judges reason about their *level of confidence* that an earlier judgment is relevant for their predictions about a current case? Ex post, every judgment is in fact, to a greater or lesser extent, questioned or rejected or ignored—or affirmed or relied upon. Nullification, reversal, striking down—by whatever term the legal system refers to the process, a rejection of a judgment by a controlling authority is a formal expression that the judge got it wrong. Endorsement, too, is sometimes formal and explicit, the archetype being a decision on appeal that affirms the judgment. Formal and explicit signals, whether of rejection or of reliance, entail a significant adjustment in how much confidence we should have in a judgment as a prediction of a future case.

It is not just in appeals that we look for signals as to how confident we should be in a given judgment as a prediction of future cases. Rejection or endorsement might occur in a different case on different facts (i.e., not on appeal in the same case) and in that situation is therefore only an approximation: ignoring, rejecting, or “distinguishing” a past judgment; or invoking it with approval, a judge in a different case says or implies that the judge in the past judgment had the law wrong or he had it right, but indirect treatment in the new judgment, whether expressed or implied, says only so much about the past one. A jurist, considering such indirect treatment, would struggle to arrive at a numerical value to adjust how much confidence to place in the past judgment. In evidence about judgments—evidence inferable from the words of the judgments themselves and evidence contained in their reception—one nevertheless discerns at least rough markers of the probability that they will be followed in the future.

There is no received view as to what Holmes thought the function of probability is in prediction. As is the case with machine learning, jurists make probabilistic as well as non-probabilistic predictions. You can state the law—i.e., give a prediction about the future exercise of public power—without giving an assessment of your confidence that your prediction is right. Jurists also use both probabilistic and non-probabilistic reasoning. Holmes, when referring to prophecies, was not however telling courts how to reason (or for that matter, legislatures or juries; we will return to juries in Chapter 7). His concern was to state what it is that constitutes the law. True, we don’t call wobbly or inarticulate judges good judges. But Holmes was explicitly not concerned with the behavior of the “good” litigant; and, in his thinking about the legal system as a whole, his concern was not limited to the behavior of the “good” judge.
Notes

1. Holmes, *The Path of the Law*, 10 Harv. L. Rev. 457, 461 (1896–1897).
2. See Kellogg (2018) op. cit.
3. 10 Harv. L. Rev. at 460–61 (1896–1897).
4. *Lochner*, 198 U.S. 45, 76 (1905) (Holmes, J., dissenting).
5. Though legal advice was not what Holmes was giving, his writings supply ample material for that purpose, and so have judges sometimes read him: see, e.g., Parker v. Citimortgage, Inc. et al., 987 F.Supp.2d 1224, 1232 n 19 (2013, Jenkins, SDJ).
6. 10 Harv. L. Rev. at 458 (emphasis added).
7. The text, which runs to 21 pages (less than 10,000 words), contains the word “prophecy,” “prophecies,” or the verb “to prophecy” on five pages: 10 Harv. L. Rev. at 457, 458, 461, 463, and 475.
8. American Banana Company v. United Fruit Company, 213 U.S. 347, 357, 29 S.Ct. 511, 513 (Holmes, J., 1909).
9. Moskowitz emphasized this line of Holmes’s thought in *The Prediction Theory of Law*, 39 Temp. L.Q. 413, 413–16 (1965–1966).
10. 10 Harv. L. Rev. at 462.
11. Id. at 462.
12. Id.
13. Id. at 457.
14. American Banana Company, 213 U.S. at 357, 29 S.Ct. at 513.
15. This chapter, pp. 59–61.
16. See for example White, *Putting Aside the Rule of Law Myth: Corruption and the Case for Juries in Emerging Democracies*, 43 Corn. Int’l L.J. 307, 321 n. 118 (2010) (reporting doubt whether Mongolia’s other branches of government submit to judicial decisions, notwithstanding the Constitutional Court’s formal power of review). The predictive character of judgments of interstate tribunals, in this sense, is pronounced. Judgments of the International Court of Justice, to take the main example, under Article 60 of the Court’s Statute are “final and without appeal,” but no executive apparatus is generally available for their enforcement, and the Court has no ancillary enforcement jurisdiction. Even in the best-functioning legal system, high courts may have an uneasy relation to the executive apparatus whose conduct they sit in judgment upon. Recall Justice Frankfurter’s concurrence in Korematsu v. United States where, agreeing not to overturn wartime measures against persons of Japanese, German, and Italian ancestry, he declared “[t]hat is their [the Government’s] business, not ours”: 323 U.S. 214, 225, 65 S.Ct. 193, 198 (1944) (Frankfurter, J., concurring).
17. Vegelahn v. Guntner & others, 167 Mass. 92, 106 (1896) (Field, C.J. & Holmes, J., dissenting).
18. Kellogg (2018) at 82.
19. See also Kellogg at 92: “Prediction had a broader and longer-term reference for [Holmes] than immediate judicial conduct, and was connected with his conception of legal ‘growth’.” See further Chapter 9.
20. Chapter 3, pp. 34–35.
21. Chapter 3, p. 38.
22. Hastie et al. (2009) 1–2.
23. It is the use of machine learning to make “predictions” in this sense (“predictive analytics”) on which legal writers addressing the topic to date largely have focused. See, e.g., Berman, 98 B.U. L. Rev. 1277 (2018).
24. The terms “postdiction” and “retrodiction” are sometimes used in legal scholarship too, though writers who use them are more likely to do so in connection with other disciplines. See, e.g., Guttel & Harel, Uncertainty Revisited: Legal Prediction and Legal Postdiction, 107 Mich. L. Rev. 467–99 (2008), who considered findings from psychology that people are less confident about their postdictions (e.g., what was the outcome of the dice roll that I just performed?) than their predictions (e.g., what will be the outcome of the dice roll that I am about to perform?) id. 471–79.
25. See Chapter 4, p. 41.
26. Dumas & Frankenreiter, Text as Observational Data, in Livermore & Rockmore (eds.) (2019) at 63–64.
27. As to gradient descent, see Chapter 2, p. 23.

Gradient descent is often coupled with another technique called cross validation, also based on prediction. The term derives from the idea of a “validation dataset.” When training a machine learning system, it is not possible to measure prediction accuracy by testing predictions on the same dataset as was used to train the machine. (This can be shown mathematically.) Therefore, the training dataset is split into two: one part for training parameter values, the other part for measuring prediction accuracy. This latter part is called the “validation dataset.” Cross validation is totemic in machine learning: stats.stackexchange.com, a popular Internet Q&A site for machine learning, calls itself CrossValidated. It is also technically subtle. See Hastie, Tibshirani & Friedman (2009) in §7.10, for a formal description

28. Chapter 3, p. 38.
29. Russakovsky, Deng et al., op. cit. (2015).
30. Hossain, Sohel, Shiratuddin & Laga, ACM CSUR 51 (2019).
31. Chapter 6, p. 70.
32. Chapter 7, pp. 81–88; and Chapter 8, pp. 89–100.
33. Chapter 9, pp. 103–111.
34. 10 HARV. L. REV. at 469.
35. Kellogg (2018) 36–53.
36. Board of Trade of the City of Chicago v. Christie Grain & Stock Company et al., 198 U.S. 236, 247, 25 S.Ct. 637, 638 (Holmes, J., 1905). See also Ithaca Trust Co. v. United States, 279 U.S. 151, 155, 49 S.Ct. 291, 292 (Holmes, J., 1929) (mortality tables employed to calculate for purposes of tax liability the value of a life bequest as of the date it was made).
37. Letter from Holmes to Pollock (Aug. 30, 1929), reprinted De Wolfe Howe (ed.) (1942) vol. 2, p. 252 (emphasis original). Kellogg quotes this passage: Kellogg (2018) at 52.
38. See for example Coastal Orthopaedic Institute, P.C. v. Bongiorno & anthr, 807 N.E.2d 187, 191 (2004, Appeals Court of Massachusetts, Bristol, Berry J.). Cf. observing a relation between uncertainty and the predictive character of law, Swanson et al. v. Powers et al., 937 F.2d 965, 968 (1991, 4th Cir. Wilkinson, CJ):

The dockets of courts are testimonies... to the many questions that remain reasonably debatable. Holmes touched on this uncertain process when he defined ‘the law’ as ‘[t]he prophecies of what the courts will do’.

39. Hastie et al., op. cit. (n. 27) § 2.3.2. This simple description of the $k$ nearest neighbor algorithm does not reflect the real cleverness, which consists in inventing a useful similarity metric such that the simple algorithm produces good predictions. Even cleverer is to use a neural network to learn a useful similarity metric from patterns in the data.
40. Id., § 4.1.
41. Lecture given by Zoubin Ghahramani at MIT, 2012. http://mlg.eng.cam.ac.uk/zoubin/talks/mit12csail.pdf.
42. Lawyers indeed give probability assessments to their clients often enough that behavioral decision theorists have studied the factors that influence lawyers’ views as to the chances of winning or losing in court. See, e.g., Craig R. Fox & Richard Birke, Forecasting Trial Outcomes: Lawyers Assign Higher Probability to Possibilities That Are Described in Greater Detail, 26(2) LAW HUM. BEHAV. 159–73 (2002).
43. As to empiricism in legal scholarship generally, see Epstein, Friedman & Stone, Foreword: Testing the Constitution, 90 N.Y.U. L. REV. 1001 and works cited id. at 1003 nn. 4, 5 and 1004 n. 6 (2015); in one of law’s subdisciplines, Shaffer & Ginsburg, The Empirical Turn in International Legal Scholarship, 106 AJIL 1 (2012).
44. See for example Kern v. Levolor Lorentzen, Inc., 899 F.2d 772, 781–82 (1989, 9th Cir., Kozinski, C.J., dissenting).
45. I.e., a numerical value representing the probability that a future court will not treat the judgment as a correct statement of law. A “p-value” is a term familiar to courts, but not one they use to describe their own judgments. See Matrixx Initiatives, Inc. v. Siracusano, 563 U.S. 27, 39, 131 S Ct. 1309, 1319 n. 6 (Sotomayor, J., 2011):

‘A study that is statistically significant has results that are unlikely to be the result of random error….’ To test for significance, a researcher develops a ‘null hypothesis’—e.g., the assertion that there is no relationship between Zicam use and anosmia… The researcher then calculates the probability of obtaining the observed data (or more extreme data) if the null hypothesis is true (called the p-value)... Small p-values are evidence that the null hypothesis is incorrect. (citations omitted)

See also In re Abilify (Aripiprazole) Products Liability Litigation, 299 F. Supp. 3d 1291, 1314–15; Abdul-Baaqiy v. Federal National Mortgage Association (Sept. 27, 2018) p. 7.

46. On a court consisting of more than one judge and on which it is open to members of the court to adopt separate or dissenting opinions, the existence and content of such opinions are another source of evidence as to how much confidence one might place in the result. The possibility of assigning confidence intervals to judgments on such evidence is suggested here: Posner & Vermeule, The Votes of Other Judges, 105 Geo. L.J. 159, 177–82 (2016). Regarding the influence of concurring opinions on future judgments, see Bennett, Friedman, Martin & Navarro Smelcer, Divide & Concur: Separate Opinions & Legal Change, 103 Corn. L. Rev. 817 (2018), and in particular the data presented id. at 854 and passim. Cf. Eber, Comment, When the Dissent Creates the Law: Cross-Cutting Majorities and the Prediction Model of Precedent, 58 Emory L.J. 207 (2008); Williams, Questioning Marks: Plurality Decisions and Precedential Constraint, 69 Stan. L. Rev. 795 (2017); Plurality Decisions—The Marks Rule—Fourth Circuit Declines to Apply Justice White’s Concurrency in Powell v. Texas as Binding Precedent—Manning v. Caldwell, 132 Harv. L. Rev. 1089 (2019).

47. Or got something in the judgment wrong while having gotten other things right. We speak above, for sake of economy of expression, about a judgment struck down in toto.

48. A recent study, though for different purposes, makes an observation apt to our point: “It is one thing to say that the standards of juridical proof are to be explicated in probabilistic terms, it is another to provide such an explication.” Urbaniak (2018) 345 (emphasis added).
Open Access  This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.