Chapter XYZ: Weight of Evidence

1 Introduction

The aim of legal-fact finding is, quite literally, to find out the facts. To find out the facts—for example, whether or not the defendant embezzled a large sum of money as the prosecution claims—it is necessary to gather, scrutinize and assess evidence from a variety of sources. One side will present evidence against the defendant, and the other side will present evidence in favor of the defendant. But this adversarial process must come to an end at some point. A decision about the facts—did the defendant commit the wrongdoing they are accused of?—must be made based on the evidence presented up to that point. This decision is preceded by another, namely a decision about whether the process of presenting, scrutinizing and assessing evidence should come to an end.

In principle, more evidence could always be presented about a disputed factual issue: witnesses, forensic experts, telephone conversations, camera recordings, and so on. But, at some point, gathering more evidence will no longer bring additional value because the costs of seeking more evidence may become too great. In this chapter, we aim to provide an account of the value of evidence. This account can then be used to formulate an account of when new evidence should no longer be sought.

1.1 Plan for the chapter

Section 2 articulates the question of the value of evidence we seek to address in this chapter. Section 3 outlines a first, probability-based attempt at modeling the informational value of the evidence. We argue that this model is inadequate. Section 4 turns to our proposal that is based on higher-order probabilities. Section 5 and the following sections apply our formal model to legal cases.

2 The value of evidence

What does it mean for new evidence to be valuable? Some terminology can help here.

2.1 Epistemic or informational value

New evidence can be *epistemically* or *informationally* valuable. It provides novel information that existing, already available evidence does not provide, for example, a new witness testimony reveals additional details about the defendant's whereabouts. This novel information can be consistent or inconsistent with what the existing evidence suggests to be the case. New evidence can be informationally valuable even when it simply agrees with other evidence, for example, when two witnesses report having seen the same thing. So, new evidence is epistemically valuable if it adds new information or strengthens the credibility of existing evidence.

2.2 Decisional value

The epistemic value of new evidence need not be the same as its *decisional* value. A new piece of evidence is decisionally valuable if, when added to the existing body of evidence, it would change the trial decision, given whatever decision criterion is used. A new piece of evidence can be epistemically valuable without being decisionally valuable, but not the other way around. Perhaps a new piece of evidence can add novel information or strengthen the credibility of other evidence, but these changes might not be significant enough to change the decision.

2.3 When to stop seeking more evidence?

If new evidence is regarded as valuable in the sense specified, should the trial decision be postponed to seek it? Not necessarily. We should not assume that evidence that is informationally and decisionally valuable should be *ipso facto* sought by investigators or that trial proceedings should be prolonged in search of this evidence. The evidence could be too expensive or even impossible to obtain. For example, having a detailed camera recording of what happened has great informational and decisional value. Still, a trial judgment should not be postponed until the recording is obtained if the costs of obtaining it are too great.

In general, the question of when to stop seeking more evidence is a practical question, moral or political. It requires balancing the costs of seeking more evidence and the informational and decisional value of the new evidence. These costs can be monetary, but also in terms of time and delayed trial decisions. Before performing this balancing act, however, we need an account of when additional evidence is epistemically and decisionally valuable.

In this chapter, we aim to provide an account of the informational value of new evidence. This account then can be used to formulate an account of the decisional value of new evidence.

2.4 Prospective and retrospective questions, and remedies

We identified three different types of questions:

- What is the epistemic or informational value of an item of evidence?
- What is its decisional value?
- Should the trial decision be postponed to seek more evidence that is valuable?

These three questions are typically viewed prospectively but they can also be formulated retrospectively. Viewed prospectively, the first two questions about the epistemic and decisional value of evidence are about evidence that has not yet been obtained. In this prospective sense, the first two questions inform the third question about what investigators, prosecutors or defense lawyers should do as they seek more evidence.

Viewed retrospectively, the first two questions concern evidence that was not obtained or evidence that is missing and could no longer be obtained. Now, even when the decision about seeking more evidence is foreclosed, one can still ask retrospectively what the investigators should have done. When a piece of evidence is missing—perhaps because it was destroyed and is no longer available—one of the two parties may benefit from this absence. So the question of possible remedies then arises, for example, whether the defendant should be acquitted because of missing evidence that the investigators should have gathered.

2.5 Stylized legal case

A stylized legal case will help fix ideas. A more detailed discussion of realistic legal cases is to be found at the end of this chapter.

The defendant confessed to having killed the victim when he burglarized the victim's apartment and stole cash and other valuables kept in a tin box. The confession was the prosecution's key incriminating evidence, but at trial, the defense raised doubts that the defendant was covering up for his son. These doubts were corroborated by the fact that the defendant was rather short and it would take a tall person to kill the victim. And the defendant's son was much taller. On the other hand, the defendant never recanted his confession. He provided abundant details all consistent with the findings of the investigators.

But one piece of evidence is missing: because the defendant confessed right away, the investigators did not examine the tin box for fingerprints. The box was later wiped clean. Had they found the son's fingerprints on it, that would be evidence of the defendant's innocence.

This case raises several questions:

- First, questions on the epistemic or informational value of the evidence. Retrospectively—knowing what we know now—what is the value of the missing fingerprint evidence? Prospectively—knowing what the investigators knew back them—what was the value of the fingerprint evidence?
- Second, questions about the decision. Retrospectively, would the fingerprint evidence—if it had been obtained during the investigation—have changed the trial decision?
- Third, questions about what to do. What should the investigators have done? This question is relevant to questions about remedies. Should the defendant be acquitted because of the missing fingerprint evidence? Should the jurors be instructed to make an adverse inference against the prosecutor for failing to present the fingerprint evidence?

2.6 Weight v. balance of evidence

Another bit of terminology is the distinction between 'weight of evidence' and 'balance of evidence'. This distinction goes back to Keynes. The balance tracks whether the evidence presented tips for or against a hypothesis of interest. Weight, for Keynes, tracks the quantity of the evidence. Here both balance and weight are properties of the *total* body of evidence.

Instead, the notion of epistemic or informational value of evidence we are after is a property of a piece of evidence relative to an existing body of other evidence. We will set aside the distinction balance/weight for now. We aim to provide an account of the informational value of evidence, viewed prospectively or retrospectively. We will return to balance/weight distinction at the end.

3 First pass

3.1 Absolute probability difference

Let's start with the simplest account of the value of new evidence, mostly in terms of a probability change. Formally, the epistemic value of new evidence E_n, against the background of other evidence E_b, can be measured by the probability change that the new evidence brings about relative to a hypothesis H of interest. So the epistemic value of E_n given background evidence E_b relative to hypothesis H is the absolute difference

$$|P(H|E_b) - P(H|E_b \wedge E_n)|.$$

The greater this difference, positive or negative, the greater the epistemic or informational value of the new evidence. The smaller the difference, the lesser the value of the new evidence.

Dahlman's account of informativeness closely follows this definition. Nance offers a similar account. For him, the informational value (or weight) of the evidence is the absolute value of the log likelihood ratio

$$log(\frac{P(E_n|H \wedge E_b)}{P(E_n|\neg H \wedge E_b)}).$$

Likelihood ratios mirror changes in probabilities: the greater the likelihood ratio, the greater the change upwards; and the lower the likelihood ratio, the greater the change downwards.

3.2 Objection

This account, however, is not satisfactory, as it may count additional evidence as worthless, while it is instead quite valuable. Suppose two witnesses testify one in favor and another against the defendant. They were both near the crime scene at the relevant time, and one identified the defendant as the perpetrator and the other did not. Absent any reason to believe one witness more than the other, assume the chances are 50/50, that is $P(G|W1 \wedge W2) = .5$

Next, we ask two other witnesses, and again we find one favoring the defendant and the other incriminating the defendant. So, again, $P(G|W1 \wedge W2 \wedge W3 \wedge W4) = .5$. The probabilities have not changed, but the testimonies of the two additional witnesses do not seem worthless. If, at first, one might be somewhat doubtful whether or not the defendant was the perpetrator, now this doubt has been entrenched. It is a stronger doubt, so to speak.

3.3 Weighted probability difference

A defender of the standard account might insist that the input of the two additional witnesses, after listening to them, is actually worthless. But *before* listening to them, we did not know what they would have said. They could have spoken both against (or in favor) of the defendant, and in that case, the probability of G would have changed quite dramatically, either upwards or downwards. So, there are three possible scenarios: (S_1) they both speak against the defendant; (S_2) they both speak in favor; and (S_3) they speak one in favor and the other against. If each scenario is 1/3 likely, there is a 2/3 chance the new evidence would be epistemically valuable and only a 1/3 chance it would not be. So, overall, the new evidence is valuable.

The proposal would be this. The epistemic value of new evidence E_n given E_b and relative to H is the weighted sum of the difference $P(H|E_b) - P(H|E_b \wedge E_n^i)$, where the weights are the probabilities of the scenarios $P(S_i)$ that could materialize and $i \in \{1, 2, 3, ...k\}$:

$$\sum_{1}^{k} P(S_i) \times |P(H|E_b) - P(H|E_b \wedge E_n^i)|$$

3.4 Another objection (or recall higher-order probabilities)

The distinction between the value of evidence that has fully formed (say a witness has testified) and evidence that has not yet fully formed (say a witness has yet to testify) is worth keeping in mind. But the two new witnesses should be regarded as valuable both before and after they have testified. True, the evidence is in equipoise before and after they testified, but after they did, the equipoise is better justified.

The general point here stems from the discussion about higher-order probability in the previous chapter. In it, we argued that precise, first-order probabilities are not enough because they fail to capture higher-order uncertainty. If precise probabilities are not enough to model uncertainty, then an account of the epistemic or informational value of new evidence that is based on precise probabilities will fall short for similar reasons.

Here is an example. Suppose an expert who listened to a telephone conversation testifies that the voice of the caller matches the defendant's voice. To quantify the uncertainty associated with the voice match, the expert introduces as is customary the random match probability, a number that expresses the likelihood that a random person would still be a match, simply by coincidence. Say this number is 1/1,000,000. Compare the voice expert with a genetic expert who testifies that the defendant genetically matches the crime traces. Suppose the genetic expert and the voice expert both testify, in different trials, that the defendant is a match, and they both give roughly similar random match probabilities, 1/1,000,000. Should we say that both forms of evidence have the same value? After all, they change the guilt probability to the same extent. But the genetic match is better evidence than the voice match: better sample, better models, etc. The random match probability is based on a probability model and sample data, so it is itself subject to uncertainty.[1] Sticking to first-order probabilities neglects this additional layer of uncertainty and the difference between genetic match and voice match. The next section outlines an account of the epistemic or informational value of the evidence that incorporates higher-order uncertainty.

4 Our proposal

The standard approach models the informational value of new evidence E_n as the difference between posterior and prior (first-order) probabilities, that is, the difference between $P(H|E_n)$ and $P(H)$. Our approach models informational value as the difference between the informativeness of the posterior and prior (second-order) distributions of (first-order) probabilities.

[1] This uncertainty is higher-order and is best modeled by a higher probability, a probability of a probability. What is the probability that the random match probability is what the expert says it is? Since the random match probability can take any value between 0 and 1, each of these values will have a higher-order probability, or in other words, there will be a distribution over the first-order probabilities (in this case, a distribution over the possible random match probabilities).

Say you have gathered some fallible evidence about hypothesis H. Your uncertainty about H up to that point can be modeled by the precise, prior probability $P(H)$. But since this first-order probability can itself be subject to uncertainty, higher-order probabilities offer a more adequate model of your overall uncertainty about H (see discussion in Chapter ZZZ). Let $PRIOR$ be a probability distribution defined over the possible values of the first-order, prior probability $P(H)$. This is your prior state of uncertainty about H given the evidence you have up to that point. Now you want to know the informational value of new evidence E_n. As you take this evidence into account, your state of uncertainty will change accordingly. Let $POST_{E_n}$ be the probability distribution defined over the possible values of the first-order, posterior probability $P(H|E_n)$.

4.1 Difference between prior and posterior higher-order distributions

Intuitively, the change that new evidence brings about to your state of uncertainty—that is, the change from prior to posterior higher-order distributions—is the informational value of the evidence. The higher-order prior and posterior distributions can have different shapes: they can have a peak and be more or less concentrated around it, or they can have multiple peaks and be more or less concentrated around them. In general, the more spread out a distribution, the less informative it will be. So the change in shape from prior to posterior distributions brought about by the new evidence will correspond to its informational value. To make this point more precise, we need a couple of notions: first, a measure of the informativeness of a distribution; and second, a measure of the change in informativeness across distributions.

Let X be a (discrete) random variable X which takes values x and is distributed according to P. The informativeness of X with its underlying distribution P is defined using standard information theory as $1-H(X,\mathsf{P})$, where H is the entropy of the distribution, typically defined as:

$$H(X,\mathsf{P}) = \sum_x \mathsf{P}(x)\log_2\frac{1}{\mathsf{P}(x)} = -\sum \mathsf{P}(x)\log_2\mathsf{P}(x),$$

Note that x is a possible value of X, for example, a first-order probability that the distributions $PRIOR$ or $POST$ could take. Crucially, entropy is not a measure of the amount of information contained in a distribution. It is rather the opposite: the expected amount of information you receive once you learn the value of X. The less informative a distribution, the more you expect to learn when you find out the value of X, the higher the entropy. The least informative distribution, the one with the highest entropy, is the uniform distribution.

The informational value of new evidence E_n is modeled by the difference between the informativeness of prior and posterior distributions. This difference boils down to the difference

between the entropies of the distribution $H(PRIOR) - H(POST_{E_n})$.[2] If this difference is positive, the entropy has decreased (and thus informativeness has increased). Conversely, if the difference is negative, the entropy has increased (and thus informativeness has decreased). This difference can be normalized by dividing by the entropy of the uniform distribution which is by definition maximal. So our account of the value VAL of new evidence E_n is as follows:

$$VAL[E_n] = \frac{H(PRIOR) - H(POST_{E_n})}{H(UNIFORM)}$$

4.2 Illustration: equipoise based on two or four witnesses

Recall the stylized example we introduced earlier as a challenge to the standard account of the value of new evidence. Initially, two witnesses testify, one in favor and another against the defendant. Next, two other witnesses testify, again one in favor and another against the defendant. The prior probability of guilt is .5, given the testimony of two witnesses, and the posterior probability is again .5, given the testimony of four witnesses. So, the two new witnesses seem to have contributed nothing. This is the verdict from the standard account. We noted that this verdict was problematic.

How does our proposal model this case? Initially, after two witnesses have testified, the prior distribution is centered at .5 and spread around .5 to some extent, where .5 is the value of the first-order probability of guilt. After two more witnesses have testified, the posterior distribution is again centered around .5. The state of equipoise persists. But the posterior distribution should now be less spread out than the prior distribution. The difference in the spreads of the two high-order distributions, prior and posterior, corresponds to the informational value of the two additional witnesses that were interrogated.

4.3 Expected difference

As noted earlier, new evidence may favor one side or the other. A witness who has not yet been interrogated may speak in favor or against the defendant. So, we should make a distinction between the value of evidence that has already been formed and evidence that has not yet formed, as in the case of a witness who has not yet given their testimony or a DNA test whose result is not yet known. The expected value of evidence can be defined as follows:

$$EXPVAL[E] = \sum_{1}^{k} P(S_i) \times VAL[E_i],$$

[2]Here we are abusing notation a bit. As noted earlier, $PRIOR$ and $POST_{E_n}$ refer to distributions, but can also be viewed as random variables that can take as values different first-order probabilities.

where S_i with $i \in \{1, 2, 3, ..., k\}$ are the different scenarios that can materialize once the evidence becomes fully formed.

5 Tin box

5.1 What happened

Let's now revisit in more detail the tin box case we mentioned earlier. The defendant confessed and turned himself to the authorities. The police found the victim dead and began the investigation immediately after. While confessions can be weak incriminating evidence when they are coerced, uncoerced confessions seem highly probative. The many details contained in the defendant's confession convinced the investigators they had a very strong case. No other investigation seemed necessary. This decision turned out to be premature, as know now.

At trial, the defense put forward an alternative hypothesis: the defendant was covering up for his son. The missing evidence—the perpetrator's fingerprints on the tin box in which the money and valuables were kept—now becomes an essential piece of evidence. Too bad the tin box was wiped clean and the fingerprints on it are lost forever.

It is clear, even without any formal modeling, that the fingerprints on the tin box would be of great value. If they match the son's fingerprints, they would be strong evidence against the son. If they match the defendant's fingerprint, they would be equally strong evidence against the defendant. It is as though the fingerprints alone could decide the case. Too bad they are gone forever. And the Swedish Court who decided the case concluded that the missing fingerprints were a good enough reason to acquit the defendant.

It is unclear whether the decision was motivated by epistemic concerns—the evidence was not good enough; why blindly trust the defendant's confession?—or policy concerns—the police did a sloppy job which should not be encouraged. Perhaps it was a combination of both.

5.2 Expected value of fingerprint evidence, before trial

We now are attempting to give a formal analysis of the case, focusing on the informational value of the fingerprint evidence. What information would the fingerprints add to the case? To answer this question, we need to answer another. What is the state of uncertainty about the defendant's guilt in light of other evidence in the case? The other items of evidence are key: (a) the confession; (b) the angle at which the victim was stabbed. How should this evidence be assessed?

Initially, the likelihood ratio associated with the confession and relative to the competing hypotheses would seem extremely high:

$$\frac{P(\text{defendant confessed}|\text{defendand did it})}{P(\text{defendant confessed}|\text{some else did it})} >>>>> 1$$

So, the posterior probability that the defendant who confessed—without being coerced—did it would seem extremely high, as well. The fact that the angle of the stabbing did not perfectly match the defendant's height seems, at this point at least, relatively unimportant.

What is the *expected* value of the fingerprint evidence from this initial point of view, that is, the point of view of the investigators who have not yet considered the possibility that the confession could be a coverup? A fingerprint analysis of the tin box could either show a match with the guy who confessed or a non-match. Clearly, given the state of the evidence up to this point, a match is much more likely than a non-match. While the expected value of the fingerprint evidence is still positive, it is not necessarily very significant (see our calculations below). This fact, together with time constraints and possible budgetary constraints, might be a good reason why the investigators decided not to collect fingerprints from the tin box.

TO ADD: for Nikodem or Rafal: *using our proposed model of the informational value of evidence, add expected value of evidence calculations relative to what the investigators knew initially. Perhaps we can also just use the precise probability model, no higher order stuff for now.*

5.3 Expected value of fingerprint evidence, after trial

At trial, things took an unexpected turn. The defense raised the possibility that the defendant who confessed might be covering up for his son. The defendant never said that and never recanted his confession. And yet, the evidence that the angle of the stabbing did not quite match the defendant's height suddenly became quite significant. In addition, if the two possible suspects are narrowed down to the defendant and his son, the likelihood ratio should go down to about one. The confession, then, becomes worthless evidence:

$$\frac{P(\text{defendant confessed}|\text{defendand did it})}{P(\text{defendant confessed}|\text{defendant's son did it})} = 1$$

That the defendant did it or that the defendant was covering up for his son would seem equally good explanations for why the defendant confessed. We can now repeat the calculations of the expected value of the fingerprint evidence. Given the change in probabilities, the fingerprint evidence now has a much greater value than it had before.

TO ADD: for Nikodem or Rafal: *using the proposed model of value of evidence, add expected value of evidence calculations relative to what we know at trial (and after narrowing hypotheses down to two suspects)*

5.4 More fine-grained model

Earlier we assumed that the hypotheses 'the son did it (and the father confessed to cover up for him)' and 'the father did it 9and confessed)' were equally good explanations of the evidence. But are they? One might argue that, oddly, the defense waited until the trial to bring up the son as a possible perpetrator. To confess first and then to raise a reasonable doubt that cannot be settled could just be a trick. So, let's refine the model by making the following assumptions. (a) The confession is more likely under the hypothesis that the defendant did it than the son did not, but assume there is significant higher-order uncertainty about this assessment of relative probabilities. (b) The angle of the stabbing is more likely under the hypothesis that the son did it than the defendant did it, but assume there is less higher-order uncertainty about this relative probability assessment.

Since the two items of evidence—essentially—cancel each other out, the state of uncertainty would seem in equipoise. But this is not quite right. Since there is greater higher-order uncertainty in the assessment of the evidence against the father (the confession) than the evidence against the son (the angle of the stabbing), the son is more likely the perpetrator given these assumptions.

TO ADD: for Nikodem or Rafal: *Run model here with these assumptions and starting from uniform guilt probability, .5 father and .5 son. Add guilt son probability distribution graph and guilt father probability distribution graph*

The final step is to add the fingerprint evidence under two possible scenarios scenarios, that the son is a match or that the father is a match. As we can see, depending on which scenario materializes, the guilt probabilities for the father or the son will change dramatically.

TO ADD: for Nikodem or Rafal: *add guilt son probability distribution graph (given fingerprints) and guilt father probability distribution graph (given fingerprints)*

TO ADD: for Nikodem or Rafal: *using the results of the above calculations, add expected value of evidence calculations relative to what we know at trial (and after narrowing hypotheses down to two suspects and using assumption (a) and (b))*

5.5 What have we learned?

As expected, the missing fingerprints have a great informational value. Admittedly, we did not need a formal model to know that. The decision of the Swedish Court to acquit the defendant could be motivated by the observation that the police should have done a better job collecting evidence. But, as we noted earlier, it is not at all clear the police should have done a better job given what they knew then. Viewed from the perspective of the investigators, the fingerprint evidence had some value but its value need not outweigh all the costs associated with collecting the additional evidence.

Suppose that the police had done a very careful job. They found that the tin box had been wiped clean by the perpetrator himself. Then what? Here the question of whether to convict or acquit the defendant becomes more strictly epistemic. The police can no longer be blamed for a sloppy job. So, is the evidence strong enough to warrant conviction? If the question becomes more strictly epistemic then, relying on the after-trial perspective seems justified. And, given the information known at the time of the trial along with the doubts raised by the defense, the evidence is overall quite weak.

Marcello comment: *What do we learn by using the formal model here? Not sure we learn anything extremely interesting by using the formal model or am I wrong?*

6 Missing fingers