

STRAINED INTERPRETATIONS

Adam Perry

Note to readers. This is a draft, and some things are missing, including some citations. Criticisms and suggestions, no matter how small, are very welcome at adam.perry@law.ox.ac.uk. For comments on earlier drafts I thank Tara Alberts, Hasan Dindjer, James Edwards, Arie Rosen, and Jan van Zyl Smit.

Abstract. Judges often give statutes strained interpretations. What makes an interpretation strained? What, if anything, justifies a strained interpretation? Is there always a point past which an interpretation would be *too* strained, or are there some interpretations which judges are entitled to adopt, whatever a statute says? I claim to answer these questions. Using the formula about conditional probability known as Bayes' theorem, I offer a precise definition of a strained interpretation, and identify two reasons for strained interpretations. I claim that the case for strained interpretations is often stronger than is realised, and illustrate that claim with a novel defence of the House of Lords' decision in *Anisminic Ltd. v Foreign Compensation Commission*. Finally, I suggest that even if a legislature is sovereign, there may be meanings which a legislature is unable to enact into law, and thus there may be limits on even a sovereign legislature's powers.

1. Ordinary mysteries	2
2. Interpretation and evidence	3
3. Bayes' theorem.....	5
4. A process for statutory interpretation	8
5. What is a strained interpretation?	9
6. Reasons for strained interpretations	10
7. Low prior odds	11
8. High posterior odds.....	15
9. Very strained interpretations	20
10. <i>Anisminic Ltd. v Foreign Compensation Commission</i>	21
11. Impossible interpretations?	25
12. Summary and conclusion.....	27

1. Ordinary mysteries

Imagine an ordinary case, about an ordinary statute, heard by an ordinary judge. The statute could conceivably be given two interpretations, *A* or *B*, and the issue is which interpretation to give it. Our judge accepts that insofar as possible she ought to give effect to the legislature's intention. If she knows that the legislature intended to enact *A* or *B*, then that is the meaning she ought to give the statute. The statutory text and context is stronger evidence that the legislature intends *A* than *B*, and the judge accepts as much. Despite her commitments and the evidence, the judge gives the statute meaning *B*. She therefore gives the statute a strained interpretation. Like the case, statute, and judge, the result is nothing special: judges give statutes strained interpretations all the time, in everyday cases and in leading cases.

Even though all this is ordinary, it's also mysterious, at least to me, at least initially. Why should judges adopt strained interpretations, if indeed they ever should? If a judge is committed to giving effect to legislative intention, and she thinks that the text and context is strong evidence of an intention to enact a specific meaning, isn't her work done, and if not, why not? This *justificatory question* is related to two others. One is the *limit question*: is there always a point past which an interpretation would be *too* strained, or are judges sometimes justified in giving a statute a certain meaning no matter what it says? The other is the *conceptual question*: what makes an interpretation strained, or conversely, what makes an interpretation natural? My aim is to answer these questions, and so to improve our understanding of a common interpretive issue.

My aim is ambitious, but I plan to make use of a powerful tool: Bayes' theorem. Bayes' theorem is a mathematical formula for updating the prior odds of an event given new evidence. In the interpretive context, Bayes' theorem provides a means to update the prior odds that Parliament intended a certain meaning (e.g., *A*) given evidence in the form of the statutory text, context, and so forth. Bayes' theorem has, to my knowledge, never been deployed in relation to statutory interpretation.¹ However Bayes' theorem is commonly discussed in the law of evidence. That discrepancy – absence in statutory interpretation, prominence in evidence law – is difficult to justify given that evidential matters closely bear on the legal meaning of a statute. The tools we use to help answer evidential questions should be relevant for interpretation. I can't conclusively argue for that general claim here. However – and this is my second aim – I hope to motivate it by demonstrating the usefulness of Bayes' theorem for interpretation.

Here's how I'll proceed. In Section 2, I set out some preliminary points about the nature of interpretation and evidence. In Section 3, I introduce Bayes' theorem. Section 4 is a sketch of a Bayesian process for statutory interpretation. The

¹ Or, with one exception, in the legal interpretive context generally. The (estimable) exception is Y. Listokin, 'Bayesian Contractual Interpretation' (2010) 39 Journal of Legal Studies 359.

theoretical heart of the article is sections 5-8. There I define a “strained” interpretation, identify two reasons for strained interpretations, and propose a partial typology of interpretive presumptions. In sections 9 and 10, I explain why a very strained interpretation is sometimes justified, and apply that lesson to *Anisminic Ltd v Foreign Compensation Commission*². Section 11 is speculative but far-reaching: if my arguments are correct, then there may be meanings which even a sovereign legislature cannot enact into law, and hence limits on a sovereign legislature’s powers. I conclude in Section 12.

2. Interpretation and evidence

Statutory interpretation is a large topic, and I need to start somewhere. I’ll treat as true three propositions that courts and commentators often endorse:

- that statutory interpretation is an exercise in determining or fixing the meaning of a statute;
- that the meaning of a statute is, usually, the meaning it’s reasonable to infer that the legislature intended it to have; and
- that what it’s reasonable to infer as to the legislature’s intention depends on the evidence of its intention.³

Evidence of legislative intent is found in the statutory text, the statutory context (including related statutes), relevant rules of syntax and grammar, and so on. We piece together this evidence using background assumptions, including assumptions about what Parliament is likely to have intended.

When we talk of “evidence”, we could equally talk of “proof”. So, it would be accurate if a little unusual to say that statutory interpretation is largely about proof of legislative intent. Lawyers are, of course, used to talking and thinking about matters of proof. In non-interpretive contexts, the law has an elaborate set of rules which govern proof of fact. At any trial, there will be rules addressing at least these questions (where *p* is a relevant proposition):

Admissibility What evidence may be used to prove or disprove *p*?

² *Anisminic Ltd v Foreign Compensation Commission* [1969] 2 AC 147, [1969] 2 WLR 163.

³ See e.g. *Regina v Secretary of State for the Environment, Transport and the Regions, ex parte Spath Holme Ltd.* [2001] 2 A.C. 349, 396, per Lord Nicholls (“the task of the court is . . . to ascertain the intention of Parliament expressed in the language under consideration”); O. Jones and F. Bennion, *Bennion on Statutory Interpretation* (LexisNexis 2013) 441-2. The second point commits me to an “objective” understanding of statutory meaning, in line with J. Goldsworthy, *Parliamentary Sovereignty: Contemporary Debates* (CUP 2010) 248. Those who adopt a “subjective” view of statutory meaning – according to which a statute’s meaning is the meaning the legislature intended to convey – will still agree that ‘the interpreter’s best inference about what was in fact intended is decisive ... even if this departs from what is in fact the legislative intent’: R. Ekins, *The Nature of Legislative Intent* (OUP 2012) 247.

<i>Weight</i>	How much weight or significance should the trier of fact give to a piece of evidence?
<i>Standards</i>	To what standard must p be proved for the trier of fact to proceed on the basis that p ?
<i>Burden</i>	On what basis should the trier of fact proceed when neither p nor not- p has been proven?
<i>Closure</i>	How much evidence must be considered before the trier of fact may conclude that p has been proved, disproved, or neither proved nor disproved? ⁴

The answers to these questions depend on the context (e.g., the standard and burden of proof is different in a criminal trial than in a civil trial⁵).

These questions are relevant to proof of fact, so they should be relevant to proof of legislative intent, because this is a factual matter.⁶ In turn, they should be relevant to statutory meaning. As Gary Lawson says in his book *Evidence of the Law*:

If consideration of any factual claim requires consideration of admissibility, significance, standards of proof, burdens of proof, and closure, then (virtually) any claim about the state of the law or legal meaning requires consideration, in some fashion, of those five elements, because (virtually) any claim about the state of the law or legal meaning is a factual claim.⁷

Lawson takes himself to be making an uncontroversial claim. I agree; on reflection, it seems obvious that we should be asking the same questions about proof of legislative intent as we do about proof of other factual propositions. Yet Lawson's book has an exasperated air, and it's easy to see why: it's rare, to say the least, for judges or lawyers (or academics) to talk about proof of legislative intent in terms of admissibility, weight, standards, burdens, or closure.

There are honourable exceptions, of course. *Pepper v Hart*, for example, is about whether statements in Hansard can be used as an aid in statutory interpretation, and thus about the admissibility of evidence in interpretation. The case has its defenders and its detractors, with each side endorsing different

⁴ G. Lawson, *Evidence of the Law* (University of Chicago Press 2016) 19-28.

⁵ In some jurisdictions, including England and Wales, there is also significant variation within these standards: T. Anderson, D. Schum, and W. Twining, *Analysis of Evidence* (2nd edn, CUP 2005) 243-4.

⁶ To be clear, though what a legislature intends is a factual issue, what a statute means is a matter of what it's reasonable to infer the legislature intends, where the reasonableness of an inference is a normative issue. The answers to the five questions listed are addressed to that normative issue. Whether normative issues are factual issues is beyond the scope of this article. I thank Hasan Dindjer (**) for encouraging me to clarify this point.

⁷ Lawson (n 4) 35.

‘principles of admissibility’⁸ for proof of legislative intent.⁹ Later, I’ll argue that at least many interpretive presumptions can be understood as rules which establish the standard or burden of proof for certain interpretive claims. But these are, to repeat, exceptions. There is little attempt to relate ideas and insights about admissibility, weight, standards, burdens, and closure in the law of evidence to proof of legislative intent.

This failure to think systematically about proof and statutory interpretation leads to at least two problems. First, what thinking has been done about proof and statutory interpretation tends to be partial and inchoate. Second, there are ideas and tools in use in the law of evidence and allied areas which haven’t been deployed in the interpretive context. My aim isn’t to develop a general theory of proof and interpretation, and I won’t say anything about admissibility, weight, or closure rules. But I do want to contribute to our understanding of standards and burdens of proof in interpretation. And I want to show the relevance of Bayes’ theorem, commonly discussed in evidence law, to interpretation. In these ways I hope to go a short way towards addressing the problems I’ve just mentioned.

3. Bayes’ theorem

Suppose that Anders and Britta are suspects in Claus’ murder. Knowing everything we do about Anders and Britta and the circumstances, we estimate that Anders is three times as likely to be the murderer as Britta. But then we discover a new piece of evidence: Claus was poisoned. It’s twice as likely that Britta would use poison to kill Claus as Anders would. Given that Claus was poisoned, who should we think is the more probable killer, Anders or Britta?¹⁰

We answer this sort of question with Bayes’ theorem. Bayes’ theorem is a mathematical formula for updating the prior odds of two competing hypotheses conditional on (given, assuming) new evidence. I’ll use H and $notH$ to represent mutually exclusive and exhaustive hypotheses, E to represent the evidence, and P to represent a probability. Following standard conditional probability notation, I’ll use the vertical bar “|” to indicate a conditional probability. Bayes’ theorem tells us that:

Bayes’ theorem

$$\frac{P(H)}{P(notH)} \times \frac{P(E|H)}{P(E|notH)} = \frac{P(H|E)}{P(notH|E)}$$

The first term, $P(H)/P(notH)$, represents the “prior odds” of H versus $notH$ (“prior” meaning the odds of H versus $notH$ without taking into account the

⁸ The phrase is Lawson’s: (n 4) 21.

⁹ See, e.g., A. Kavanagh, ‘*Pepper v Hart* and Matters of Constitutional Principle’ (2005) 121 LQR 98.

¹⁰ The example is adapted from ‘Bayes’ Rule: Odds Form’ at <https://arbital.com/p/bayes_rule_odds/>, last accessed 10 April 2018.

evidence). The second term, $P(E|H)/P(E|notH)$, is known as the “likelihood ratio”, and is the ratio of the probability of evidence E conditional on H versus the probability of E conditional on $notH$. (More on the likelihood ratio below.) Multiplying the prior odds by the likelihood ratio gives us the answer we were looking for, in the form of the “posterior odds” of H versus $notH$, that is, $P(H|E)/P(notH|E)$.¹¹ So, in words, this is Bayes’ theorem:

$$\text{prior odds} \times \text{likelihood ratio} = \text{posterior odds}$$

It’s possible to use a version of Bayes’ theorem to determine the posterior odds when there are more than two hypotheses; but, for simplicity’s sake, I’ll focus on cases in which there are only two hypotheses.

Returning to our example, we can use Bayes’ theorem to calculate the odds that Anders versus Britta is the murderer. Using *Anders* to represent the hypothesis that Anders is the murderer, *Britta* to represent the hypothesis that Britta is the murderer, and *poison* to represent the evidence, Bayes’ theorem says:

$$\frac{P(\text{Anders})}{P(\text{Britta})} \times \frac{P(\text{poison}|\text{Anders})}{P(\text{poison}|\text{Britta})} = \frac{P(\text{Anders}|\text{poison})}{P(\text{Britta}|\text{poison})}$$

Before taking into account the evidence we said that Anders is three times as likely to be the murderer. Thus the prior odds are:

$$\frac{P(\text{Anders})}{P(\text{Britta})} = \frac{3}{1}$$

We also said that it’s half as likely that poison was used assuming Anders killed Claus than that Britta killed Claus. Thus the likelihood ratio is:

$$\frac{P(\text{Poison}|\text{Anders})}{P(\text{Poison}|\text{Britta})} = \frac{1}{2}$$

The posterior odds are easy to calculate:

$$\frac{P(\text{Anders}|\text{Poison})}{P(\text{Britta}|\text{Poison})} = \frac{3}{1} \times \frac{1}{2} = \frac{3}{2}$$

Between Anders and Britta, the odds are 3:2 that Anders is the murderer. Anders is more likely the killer, notwithstanding that Claus was killed with poison.

When the hypotheses are mutually exclusive and exhaustive, the odds imply the probability of each hypothesis. Thus, if either Anders or Britta murdered Claus, the *prior* probability that Anders murdered Claus is $3/(1 + 3) = 3/4 = 0.75$, and the *prior* probability that Britta murdered Claus is $1/(3 + 1) = 1/4 = 0.25$. Meanwhile, the *posterior* probability that Anders murdered Claus is $3/(2 + 3) = 3/5 = 0.6$, and the *posterior* probability that Britta murdered Claus is

¹¹ For an introduction to Bayes’ theorem, and proofs of the theorem in both odds and probability form, see N. Fenton and M. Neil, *Risk Assessment and Decision Analysis with Bayesian Networks* (CRC Press 2013) ch 5.

$2/(3 + 2) = 2/5 = 0.4$. In other words, given that Claus was poisoned, there's a 60% chance that Anders murdered Claus and a 40% chance that Britta did. Note that even though Britta is much more likely than Anders to use poison, taking into account this evidence only decreased the probability that Anders is the murderer by $0.75 - 0.6 = 0.15$. That shows the "stickiness" of priors – it takes quite strong evidence to overcome the initial improbability of some hypothesis. Indeed, Britta is probably the murderer only if she is more than *six* times as likely as Anders to use poison.

If you have only a passing acquaintance with Bayes' theorem, you may not have seen it stated as above, which is its "odds form". Bayes' theorem is more commonly stated in "probability form"¹². I'll tend to use odds rather than probabilities in what follows. For one thing, thinking in terms of odds is more natural for many people than thinking in terms of probabilities. Some people find Bayes' theorem counter-intuitive, and it may be no coincidence that Bayes' theorem is often stated in probability rather than odds form. Also, I'll be making use of the likelihood ratio in what follows, which is a term in the odds form of Bayes' theorem but not the probability form. To be clear, nothing of substance turns on the form of the theorem. I'll always assume that the hypotheses are mutually exclusive and exhaustive, so it's always possible to convert odds into implied probabilities, as indeed I sometimes will.

Let me end with two clarifications. The first is about what Bayes' theorem is and isn't useful for. Bayes' theorem is useful for identifying probable causes from effects, e.g., who the killer is (cause) given that poison was used (effect). But the theorem has no normative, action-guiding significance in itself. Anders is more likely than Britta to have killed Claus, but it doesn't follow that the police ought to arrest Anders, that a jury ought to convict him, that his friends ought to shun him, etc. The odds may favour some hypothesis, but it's a separate question what we should do about it, if anything.

The second clarification is about the scope of my argument. I'm not trying to provide a from-scratch argument for the relevance of Bayes' theorem to matters of legal proof. There's a large and robust scholarly literature devoted to that task, and generally to defending the importance of probabilistic analyses to factual questions in law.¹³ I have nothing helpful to add to what's already been said on the general issue. In what follows, I'll take it as given that Bayes' theorem provides a useful framework in which to evaluate factual claims, and aim instead to extend that framework to the interpretive context.

¹² That is:

$$\frac{P(H) \times P(E|H)}{P(H) \times P(E|H) + P(\text{not}H) \times P(E|\text{not}H)} = P(H|E)$$

¹³ For a small sample of the literature on the uses of Bayes' theorem, see B. Robertson and G.A. Vignaux, 'Probability – The Logic of the Law' (1993) 13 OJLS 457; J Koehler, 'When Do Courts Think Base Rates Statistics Are Relevant?' (2002) 42 Jurimetrics 373; AP Dawid, 'Bayes's Theorem and Weighing Evidence by Juries' in R Swinburne (ed), *Bayes's Theorem* (British Academy 2005); N Fenton, M Neil, and D Berger, 'Bayes and the Law' (2016) 3 Annu. Rev. Stat. Appl. 51.

4. A process for statutory interpretation

How would a proof-minded, Bayes-friendly court go about interpreting a statute?

It would start by identifying the possible alternative intended meanings, say, *A* and *B*. It would identify the relevant evidence, consisting at least of the statutory text and context, which I'll represent as *E*. The court would then ask itself at least these questions:

Q1. What is the likelihood ratio? The court would estimate the probability of seeing *E* given that the legislature intends *A*. It would estimate the probability of seeing *E* given that the legislature intends *B*. The ratio of these probabilities is the likelihood ratio.

Q2. What are the prior odds? Drawing on what it knows about the legislature's capacities, past actions, goals, principles, and values, the court would estimate the prior probability of the legislature intending *A*. Drawing on the same sources, it would estimate the probability of the legislature intending *B*. The first probability divided by the second is the prior odds.

A word about the reference in Q2 to 'goals, principles, and values'. What the legislature's goals, principles, and values *are* is one thing. What its goals, principles, and values *ought to be* is another. Q2 is only about the former. Suppose that a legislature is committed to gender inequality. Even though this goal is blameworthy, it's still relevant to the prior odds that the legislature has certain intentions on matters of gender. So, to answer Q2, the court will need to inquire into the legislature's moral views. But the court won't need to engage in moral theorising itself.¹⁴

Given its answers to Q1 and Q2, the court would use Bayes' theorem to calculate the posterior odds, i.e., the odds of *A* given *E* versus *B* given *E*. In light of those odds, the court would go on to ask:

Q3. What is the standard of proof? The default standard of proof is the balance of probabilities, or in terms of odds, better than evens (1:1). But the standard of proof can be higher or lower, and it can be different for different meanings (more on this below).

Q4. Does either meaning satisfy the relevant standard of proof?

If a meaning is proved to the relevant standard – i.e., if the answer to Q4 is "yes" – then the court would adopt that meaning. If neither meaning has been proved, the court still needs to decide the case. It must somehow 'break the logjam'¹⁵. And so it must ask:

Q5. Who bears the burden of proof?

The court would decide the interpretive issue against whomever bore the burden.

¹⁴ I thank Arie Rosen (**) for pointing out the importance of the difference to me.

¹⁵ P. Roberts and A. Zuckerman, *Criminal Evidence* (2nd edn, OUP 2010) 224.

To emphasise, I'm not suggesting that real courts are openly or self-consciously Bayesian, or that they follow these steps, described in these ways. Of course they aren't, and don't.¹⁶ But what real courts do when interpreting statutes often approximates what a court following these steps would do. And, as a result, we can sometimes use the analytical process sketched here to rationalise the interpretations real courts reach, including their adoption of strained interpretations.

5. What is a strained interpretation?

To define a strained interpretation, I need to say a little more about the likelihood ratio. To recall, where H and $notH$ are the hypotheses, and E is the evidence, then:

Likelihood ratio

$$\frac{P(E|H)}{P(E|notH)}$$

If $P(E|H)$ is larger than $P(E|notH)$, then the posterior odds of H versus $notH$ are higher than the prior odds. That is, if:

$$\frac{P(E|H)}{P(E|notH)} > 1$$

then:

$$\frac{P(H|E)}{P(notH|E)} > \frac{P(H)}{P(notH)}$$

Put another way, when the likelihood ratio is > 1 , taking into account the evidence increases the odds of H . The larger the likelihood ratio, the greater the increase. Conversely, if the likelihood ratio is < 1 , taking into account the evidence decreases the odds of H . Thus, the likelihood ratio provides all the information (relevant to the hypotheses) given by the new evidence. That is, from a Bayesian point of view, the likelihood ratio sums up the evidential meaning of new information.¹⁷

In the example of the murder investigation, the likelihood ratio – the ratio of the probability that poison is used if Anders killed Claus versus the probability that poison is used if Britta killed Claus – is 1:2, or 0.33. Because the likelihood ratio is < 1 , the posterior odds that Anders killed Claus are less than the prior odds he killed Claus. Conversely, the posterior odds that Britta killed Claus are greater

¹⁶ Real-life courts have sometimes rejected Bayesian methods, most notably the Court of Appeal for England and Wales in *R v T*, [2011] 1 Cr. App. R. 9, [2010] EWCA Crim 2439. For vigorous criticism of *R v T*, see B. Robertson, G.A. Vignaux, and C. Berger, 'Extending the Confusion About Bayes' (2011) 74 MLR 444; G.S. Morrison, 'The Likelihood-Ratio Framework and Forensic Evidence in Court: A Response to *R v T*' (2012) 16 International Journal of Evidence & Proof 1.

¹⁷ See the helpful explanation of likelihood ratios in I. Hacking, *An Introduction to Probability and Inductive Logic* (CUP 2001) 180-181.

than the prior odds she did so. That Claus was poisoned is therefore evidence that Anders is not the murderer, and that Britta is the murderer. So the likelihood ratio tells us what a piece of evidence is evidence *for*. It also tells us how *strong* the evidence is.¹⁸ Were the likelihood ratio 0.5, say, instead of 0.33, the poison would be weaker evidence that Anders is not the murderer and that Britta is the murderer. Were it 0.25 instead of 0.33, the poison would be stronger evidence that Anders is not murderer and that Britta is.

How does this tell us what a strained interpretation is? It's plausible that one interpretation is strained relative to another interpretation because it accounts for the text, context, and other evidence of legislative intent less well than the competing interpretation.¹⁹ Put another way, one interpretation is strained compared to another because it's less probable that we'd see the statutory text, context, and other evidence if the legislature intended the first interpretation over the second. We can define a strained interpretation accordingly:

Strained interpretation

Where A and B are exclusive and exhaustive interpretations of a statute, and E is the evidence of legislative intent, B is a strained interpretation if and only if $P(E|A)/P(E|B) > 1$.

The larger the gap between $P(E|A)$ and $P(E|B)$, the more strained an interpretation B is. I'll call an interpretation which is not strained is "unstrained" or "natural".

As I said, I think this definition makes precise a plausible understanding of a strained interpretation. That's one point in its favour. Another point is that it yields intuitive results in particular cases, which I show through examples in sections 7, 8, and 10.

6. Reasons for strained interpretations

What justifies a strained interpretation? It'll take me a while to answer this question, so let me indicate where I'm heading. The analytical process outlined in section 4 tells us that there are two disjunctive conditions under which a strained interpretation is justified. The first is that the strained meaning has been proved to the relevant standard. For this condition to be satisfied, the prior odds of the natural meaning must be low. How low? It depends on how strained the interpretation is, but in any case, $\ll 1$. The second condition is that no meaning has been proved and the burden of proof favours the strained interpretation. For this second condition to be satisfied (when the first is not), the posterior odds needed to adopt the natural meaning must be high, and in any case, $\gg 1$. The next section

¹⁸ Robertson and Vignuax (n 13) 466.

¹⁹ See Jones and Bennion (n 3) 429 for a similar definition. Whereas *Bennion's* definition is about departures from the 'clear grammatical meaning', my definition is about departures from the meaning favoured by the evidence generally.

is about the first reason for a strained interpretation. Section 8 is about the second reason. Section 9 discusses their combination.

7. Low prior odds

While on patrol one night, police constable Bowley tried to stop a fight outside of a club. Ms Wills intervened and said to him: “Why don’t you fuck off”. When the constable asked her to be quiet, she replied with a stream of other obscenities. And when Bowley and his fellow constables tried to arrest her, Wills kicked them in the legs and bit one of them on the arm. Section 28 of the Town Police Clauses Act 1847 says:

Every person who in any street, to the obstruction, annoyance, or danger of the residents or passengers, commits any of the following offences shall be liable to a penalty ... and any constable ... shall take into custody, without warrant, and forthwith convey before a justice, any person who within his view commits any such offence; ... [including any person who] ... uses any profane or obscene language.

In addition, s 51 of the Act makes it an offence to assault a police constable in the execution of his duty. Wills was charged under both sections: under s 28 with using obscene language to the annoyance of others, and under s 51 with assaulting Bowley in the execution of his duty. Wills was acquitted of the first charge because no one was actually annoyed by her obscene language. But she was convicted of assaulting Bowley.

On appeal in *Wills v Bowley*²⁰, it was argued on Wills’ behalf that the combination of verdicts didn’t make sense: her acquittal on the first charge necessitated her acquittal on the second charge. That would be right if s 28 meant:

Actual

A constable shall arrest only those who *actually* commit an offence.

Wills hadn’t committed an offence under s 28, hence Bowley hadn’t been authorised to arrest her, hence she hadn’t assaulted him in the execution of his duty. Her conviction under s 51 couldn’t stand. Or so it was argued. But there was a second way to interpret s 28:

Belief

A constable shall arrest anyone he *reasonably believes* commits an offence.

Because Bowley reasonably believed that Wills had used obscene language to the annoyance of others, he was executing his duty by arresting her. Her conviction under s 51 should stand – or so the prosecution said.

If Parliament had intended that a constable may arrest a person only if he or she has actually committed an offence, then it’s probable we’d see language like

²⁰ *Wills v Bowley* (1982) 75 Cr. App. R. 164.

we do in s 28. After all, that section does say: 'Every person who ... commits any of the following offences ...'. Isn't that exactly the kind of thing Parliament would have said had it meant to authorise arrests of only those who commit the listed offences? Supposing it's certain that we'd see text like in s 28 given *actual*, and using *text* to represent the statutory text, then:

$$P(\text{text}|\text{actual}) = 1$$

Of course it's possible that Parliament intended a broad power of arrest and simply forgot to include certain words in s 28. The Town Police Clauses Act 1847 dates from a time when drafting standards were lower. And maybe Parliament thought it so obvious it didn't need saying that a constable should be able to arrest those they believe to have committed an offence. These things are possible, but improbable. Suppose we estimate the following:

$$P(\text{text}|\text{belief}) = 0.1$$

The likelihood ratio is therefore $P(\text{text}|\text{actual})/(P(\text{text}|\text{belief}) = 1/0.1 = 10$. It's ten times more likely that we'd see the language in s 28 assuming Parliament intends *actual* versus *belief*. For Wills to win, all she needs is for the court to give the statute its natural, unstrained interpretation. She'll lose if the court gives it a strained interpretation. I take this to be an intuitive description of the question in the case. It is also how commentators have viewed it.²¹

Even though it was strained, a majority of the House of Lords adopted *belief*. Why? Lord Bridge (with whom Lord Wilberforce and Lord Russell agreed) thought that, setting aside the statutory text, it was deeply implausible that Parliament intended to empower constables to arrest only those who had actually committed offences. For one thing, constables can't know for sure whether a judge or jury will decide that a person has committed an offence. If constables guess wrong when they make arrests, they'll be liable in tort. Of course constables won't want to risk that, so they'll arrest very few people – which Parliament didn't want. Moreover, given that s 28 says that constables 'shall' arrest anyone who commits a relevant offence, *actual* would put constables in a terrible bind. It would be 'quite ridiculous' to interpret s 28 to 'force upon the constable a choice between the risk of making an unlawful arrest and the risk of committing a criminal neglect of duty'²². It would 'impale him on the horns of an impossible dilemma'²³.

Suppose that we estimate the prior probability of Parliament's intention to enact *actual* and put constables in this 'impossible' dilemma is just 0.05. Because probabilities sum to 1, the prior probability of *belief* is therefore $1 - 0.05 = 0.95$. So:

²¹ Oliver and Bennion describe *belief* as a 'strained construction': (n 3) 872. Lord Lowry describes *actual* as the 'natural and literal construction': *Wills v Bowley* (n 20) 182.

²² *Wills v Bowley* (n 20) 197.

²³ *Wills v Bowley* (n 20) 197-198.

$$P(actual) = 0.05$$

$$P(belief) = 0.95$$

Bayes' theorem tells us that:

$$\frac{P(actual)}{P(belief)} \times \frac{P(text|actual)}{P(text|belief)} = \frac{P(actual|text)}{P(belief|text)}$$

Plugging in the numbers gives us:

$$\frac{0.05}{0.95} \times \frac{1}{0.1} = \frac{0.05}{0.095} = \frac{10}{19}$$

The odds are 10:19 that Parliament intended *actual*. In terms of implied probability, there's a $10/(19 + 10) = 10/29 = 0.35$ or 35% probability that Parliament intended *actual*. Hence there's a $1 - 0.35 = 0.65$ or 65% chance that Parliament intended *belief*. Applying a civil standard of proof, a Bayesian court would interpret the statute to mean *belief*—just as the House of Lords did.

A visualisation can help us see what might be going on in *Wills v Bowley*. Suppose that Parliament enacts a statute like the one at issue here 200 times. The chart below shows the distribution of those 200 statutes:²⁴

Table 1. *Wills v Bowley*

<i>Text</i> \ <i>Intention</i>	No reference to "belief"	Reference to "belief"	Total
Actual	10	0	10
Belief	19	171	190
Total	29	171	200

Of the 200 statutes, Parliament intends *actual* in $200 \times 0.05 = 10$ statutes. In none of those statutes will we see a reference to "belief". Of the remaining 190 statutes, in which Parliament intends *belief*, we'll see no reference to "belief" in $190 \times 0.1 = 19$ statutes, but will see a reference to "belief" in $190 \times 0.9 = 171$ statutes. Overall, there's no reference to "belief" in $10 + 19 = 29$ statutes. The key point is that, even when there's no reference to "belief", a majority of the time – $19/29 = 0.65$ – Parliament intends *belief*. Even when Parliament doesn't say "belief", it probably intends *belief*. That perhaps counter-intuitive result is why *belief* is proved on a civil standard, despite it being strained.

²⁴ I've adapted Table 1 from Listokin (n 1) 366.

Now that we know that it's possible for low prior odds to justify strained interpretations, I want to state the justification in general terms. As above I'll represent the two interpretations as *A* and *B* and the evidence as *E*. Even when *B* is strained, the posterior odds favour *B* if:

$$\frac{P(E|A)}{P(E|B)} < \frac{P(B)}{P(A)}$$

Assuming a civil standard of proof, a Bayesian court would consider *B* proved, and hence a justified interpretation, when this condition is met. We can check this claim against *Wills v Bowley*. The strained interpretation, *belief*, is justified if:

$$\frac{P(\text{text}|\text{actual})}{P(\text{text}|\text{belief})} < \frac{P(\text{belief})}{P(\text{actual})}$$

And we know this condition is satisfied, because:

$$\frac{1}{0.1} < \frac{0.95}{0.05}$$

The court should therefore consider *belief* proved, as it did. More simply, the idea is that a strained interpretation is justified if the evidence in favour of the natural interpretation doesn't make up for its intrinsic improbability.

Wills v Bowley is an example of reliance on the presumption against absurdity. Lord Bridge thought it would be absurd to put police constables in an impossible situation. One way to think of the presumption against absurdity is as a direction to courts to set the prior odds of any absurd meaning low.²⁵ Indeed, many interpretive presumptions can be understood as directions as to prior odds. Consider the *ejusdem generis* rule, according to which a general term takes its meaning from the more specific items in an accompanying list. The rule is based on an unobjectionable generalisation about how language tends to be used. When they invoke the *ejusdem generis* rule, courts can be understood to invoke a justification for setting the prior probability in favour of an interpretation of a general term low. Much the same could be said of other so-called "linguistic canons", such as *noscitur a sociis* and *expressio unius est exclusio alterius*. The presumption against absurdity is a stronger presumption than the linguistic canons, meaning that it is a direction to set the odds of a certain interpretation lower. I'll call this type of presumption a "low prior odds rule" or "low prior rule" for short.

Why do we need low prior rules? Why not just rely on courts to estimate the prior odds accurately on their own? One reason is that courts might not have the knowledge or expertise to estimate them accurately. Rules like *ejusdem generis*

²⁵ The term "presumption" is sometimes used to refer to a certain attitude (e.g., "My presumption is that a child born in wedlock is legitimate"). As I use the term, it refers to a legal rule which requires that attitude. See E. Ullmann-Margalit, 'On Presumption' (1983) 80 *The Journal of Philosophy* 143.

may sound obvious or even trivial to us, but they reflect the accumulation of judicial thinking about language and legislative intent over centuries. A second reason is suggested by psychological research showing that people tend to neglect the significance of prior odds (or prior probabilities) when determining posterior odds (or posterior probabilities).²⁶ Because people tend to err in this way, they tend to overestimate the significance of new evidence, and to underestimate the significance of the prior odds or probability. This is the “base rate fallacy” (“base rate” being another term for prior probability). In the interpretive context, committing the base rate fallacy means overestimating the significance of text and context and underestimating the importance of the base rates of legislative intentions. By requiring courts to have regard to prior odds, and to establish those odds in a particular way, low prior rules may help us guard against this sort of error.

8. High posterior odds

The record unfortunately doesn’t say why Ann Harris bit off the end of a woman’s nose, but bite it off she did, and she was duly charged with a criminal offence under the Offences Against the Persons Act 1828. The statute referred to shooting at and drawing a trigger upon a person, and to attempting to discharge a loaded weapon at someone, and then added ‘or shall unlawfully and maliciously stab, cut, or wound any person’²⁷. The question for the court in *R v Ann Harris*²⁸ was whether biting off the end of a person’s nose was a ‘wound’ for the purpose of the statute. At first it might seem that the answer is obviously “yes”. After all, biting off the end of someone’s nose is a grievous wound in an ordinary sense. The argument on Harris’s behalf was that the statutory context suggested a narrower interpretation. The references to ‘shooting’, ‘drawing’, and discharging a weapon, as well as the words ‘stab’ and ‘cut’, suggested that the relevant category of wounds were those made by instruments, not by teeth or hands or the like.

We can reframe the central question in *Harris* this way. There were two possible interpretations of the statute:

Teeth

‘Wound’ includes wounds from teeth as well as instruments.

Instrument

²⁶ The classic article is D. Kahneman and A. Tversky, ‘On the Psychology of Prediction’ (1973) 80 Psychological Review 237.

²⁷ 9 Geo. IV, c. 31, s 12.

²⁸ *R v Ann Harris* (1836) 7 Carrington & Payne 446, 173 ER 198.

'Wound' only includes wounds from instruments (e.g., guns, knives).

It's surely more probable that we'd see the language we do if Parliament intends *teeth* than *instrument*. Suppose we estimate the following probabilities (with *text* representing the statutory text):

$$P(\text{text}|\text{teeth}) = 1$$

$$P(\text{text}|\text{instrument}) = 0.2$$

The likelihood ratio is $1/0.2 = 5$. Thus *teeth* is the natural interpretation, and *instrument* a strained interpretation – an intuitive result.

Earlier I said that low prior odds can sometimes justify a strained interpretation. But that won't work in *Harris*. Setting aside the statute, it's plausible that the odds are in favour of Parliament intending to prohibit wounding generally, rather than just wounding with instruments. *Harris* therefore isn't a low prior odds case, unlike *Wills v Bowley*. However, like in *Wills v Bowley*, the judge in *Harris* adopted a strained interpretation – that is, *instrument*. He directed the jury to acquit Harris and it did so. What, if anything, justifies the judge's choice to adopt *instrument*? We know the answer is not low prior odds; what remains?

My answer has three steps. The first step is about knowledge. If a court knows that Parliament intends a statute to have some meaning, then that is the meaning a court should give it.²⁹ When does a court "know" that Parliament intends some meaning? It may be tempting to say that a court knows that a meaning is intended when it knows that the odds are that Parliament intended that meaning. Thus, a court would know that Parliament intends *A* if the odds are that Parliament intends *A*, and it would know that Parliament intends *B* if the odds are that Parliament intends *B*. But many epistemologists would disagree. They say that what we claim to know, and what we're willing to say that others know, is sensitive to the practical stakes of acting or relying on some premise. Specifically, as the practical stakes of acting on some premise increase, our willingness to claim or attribute knowledge of it decreases, even holding evidence of it constant.³⁰

²⁹ The line of argument in this and the next four paragraphs is modeled on Ryan Doerfler's excellent article 'High Stakes Interpretation' (2018) 116 Michigan Law Review 523. What is mine is the situation of the argument in relation to odds and standards of proof.

³⁰ This basic claim is common ground amongst many competing theories. One is *epistemic contextualism*, which holds that what's expressed by an attribution of knowledge depends partly on the context of the attributor. For an overview of epistemic contextualism, see P. Rysiew, 'Epistemic Contextualism' in E.N. Zalta (ed), *The Stanford Encyclopedia of Philosophy* (Winter 2016), at <<https://goo.gl/y9ue5r>> accessed 18 April 2018. A competing theory is *interest-relative invariantism*, according to which the greater the practical investment you have in a belief, the stronger your evidence must be to know it. See Jason Stanley, *Knowledge and Practical Interests* (OUP 2005). For a helpful discussion of epistemic contextualism and interest-relative invariantism in relation to the standard of proof in legal contexts, see H.L. Ho, *A Philosophy of Evidence Law: Justice in the Search for Truth* (OUP) 2008 at 201-210. Doerfler discusses these and other possible theories in connection with statutory interpretation: (n 28) 544-551.

Philosophers demonstrate the point using carefully constructed hypotheticals³¹, but the basic point is a familiar one. You'll be less willing to say you "know" you turned the iron off if you're leaving for a week-long holiday than if you're going to the shop on the corner.³² You'll be less willing to say the waiter "knows" the dish is free of peanuts if you have a serious allergy than if you have a mere dislike of them. And we think that epistemic standards are rightly more stringent at a trial, where the stakes are high, than in most everyday contexts, where they're lower.³³ As Ryan Doerfler says, it appears to be a general principle 'that one needs greater epistemic justification to act on some premise the higher the practical stakes'.³⁴

The second step is to apply this general principle to interpretation. Suppose that the choice between interpretations *A* and *B* is a high stakes choice. Whereas *A* would have grave or troubling consequences, *B* wouldn't. The court should still be prepared to adopt *A* if it knows that's what Parliament intended. But, given the high stakes, the court won't know that *A* is the intended meaning unless there is strong epistemic justification – i.e., proof to a high standard – that *A* is the intended meaning. The higher the stakes, the higher the standard of proof should be. For example, the criminal standard of proof is often estimated at between 0.9 to 0.95 (in terms of odds, between 9:1 and 19:1).³⁵ In principle, a court could adopt a similarly high standard of proof in the interpretive context.

Imagine that the court does adopt a high standard of proof for *A*, such that *A* is proved if and only if $P(A|E)/P(B|E) \geq x$, where $x \gg 1$. The court adopts the default civil standard towards *B*. Now suppose that:

³¹ For instance:

Mary and John are at the L.A. airport contemplating taking a certain flight to New York. They want to know whether the flight has a layover in Chicago. They overhear someone ask a passenger Smith if he knows whether the flight stops in Chicago. Smith looks at the flight itinerary he got from the travel agent and respond, 'Yes I know—it does stop in Chicago.' It turns out that Mary and John have a very important business contact they have to make at the Chicago airport. Mary says, 'How reliable is that itinerary? It could contain a misprint. They could have changed the schedule at the last minute.' Mary and John agree that Smith doesn't really *know* that the plane will stop in Chicago. They decide to check with the airline agent.

S. Cohen, 'Contextualism, Skepticism, and the Structure of Reasons' (1999) 13 *Philosophical Perspectives* 57, 58-59. The example is said to show that attributions of knowledge are context-sensitive. See K. DeRose, 'Contextualism and Knowledge Attributions' (1992) 52 *Philosophy and Phenomenological Research* 913, 913-914 for a similar example.

³² The example is from Doerfler (n 28) 549.

³³ Ho (n 29) 211 ('The fact-finder must not reach factual conclusions as easily at a trial as she does in everyday affairs, where little is at stake in her beliefs and assertions').

³⁴ Doerfler (n 28) 528.

³⁵ See e.g. L. Laudan, 'Is Reasonable Doubt Reasonable?' (2003) 9 *Legal Theory* 295, 310, 312-313; A. Walen, 'Proof Beyond a Reasonable Doubt: A Balances Retributive Account' (2015) 76 *Louisiana Law Review* 355, 374.

$$x > \frac{P(A|E)}{P(B|E)} > 1$$

Because $x > P(A|E)/P(B|E)$, A hasn't been proved. Because $P(A|E)/P(B|E) > 1$, B hasn't been proved. Neither A nor B has been proved. The court doesn't know that Parliament intends A , and it doesn't know that Parliament intends B . The intended meaning is therefore indeterminate.³⁶ What is the court to do? It can't simply suspend judgment. The case must be decided; the dispute must be resolved. The question is how.

Now for the third step, about safe choices. If you don't know whether you turned the iron off and you're about to go on vacation, go and double check. If you don't know whether a dish contains peanuts and you have a serious peanut allergy, don't order the dish. If you don't know whether a defendant is guilty, acquit. In short: when the stakes are high, and you're uncertain, opt for the "safe" choice, if one is available.³⁷ When a court is faced with a choice between only two interpretations, the safe choice is to adopt the meaning without the grave or troubling consequences, even if it's strained. In my example this means adopting B . In general, a court should adopt a strained interpretation over a natural one if neither interpretation has been proved (despite an appropriate opportunity) and the strained interpretation is the safe choice. Put another way, the party seeking to convince the court to adopt a grave or troubling interpretation bears the burden of proof. When neither interpretation has been proved, the court should rely on the burden of proof to decide the case.

(I imagine someone might object at this point. Interpretation is about figuring out what the legislature intended, she'll say. When a court gives a statute a meaning on some other basis, like the burden of proof, it isn't interpreting it at all. So, this isn't an argument for a strained interpretation, but for something quite different. That's the objection. Here it's helpful to distinguish, as Jeffrey Goldsworthy does, between clarifying a pre-existing meaning of a legal text, on the one hand, and adding to a text a meaning it didn't already possess, on the other.³⁸ Some scholars reserve "interpretation" for the clarifying activity and prefer terms like "construction" for the creative activity.³⁹ Others use "interpretation" to cover both activities.⁴⁰ I think that the second, broader sense of "interpretation" more closely tracks everyday and legal usage, so that's the sense

³⁶ Lawson (n 4) 113 (when neither p nor not- p is proved to the relevant standard, p is indeterminate).

³⁷ Doerfler (n 28) 549-50.

³⁸ J. Goldsworthy, 'The Case for Originalism' in G. Huscroft and B.W. Miller, *The Challenge of Originalism: Theories of Constitutional Interpretation* (CUP 2011) 60.

³⁹ For a similar distinction between interpretation and construction, see L.B. Solum, 'The Interpretation-Construction Distinction' (2010) 27 *Constitutional Commentary* 95.

⁴⁰ e.g. Goldsworthy (n 38) 60; M.N. Berman, 'Reflective Equilibrium and Constitutional Method' in G. Huscroft and B.W. Miller, *The Challenge of Originalism: Theories of Constitutional Interpretation* (CUP 2011) 255-256.

in which I've been using the word here. But this is just a labeling issue.⁴¹ I'm arguing that we should resolve a certain type of indeterminacy in a certain way, which I think of as an act of creative interpretation. However, if you prefer a different word, I invite you to read this section as an argument for a kind of "strained construction".)

With the argument now set out, we can better sense of *Harris*. Bayes' theorem tells us that:

$$\frac{P(\text{teeth})}{P(\text{instrument})} \times \frac{P(\text{text}|\text{teeth})}{P(\text{text}|\text{instrument})} = \frac{P(\text{actual}|\text{teeth})}{P(\text{belief}|\text{teeth})}$$

As I suggested, it's plausible that the prior odds favour *teeth* over *instrument*. There's an obvious public benefit to prohibiting things like nose-biting. So let's estimate the prior odds at 3:1 in favour of *teeth*. And, as I said, the likelihood ratio is 5:1. Therefore the posterior odds are:

$$\frac{3}{1} \times \frac{5}{1} = \frac{15}{1}$$

Odds of 15:1 in favour of *teeth* are equivalent to a probability of 0.94. As a result, *teeth* would easily be proved on a civil standard. However, there's another factor to consider. *Harris* is a 19th century case. Criminal punishments then were much harsher than they are now. It may be reasonable to regard *Harris* as a high stakes case. It would make sense, as a result, for the judge to demand a high standard of proof to adopt *teeth*, and perhaps even proof to a standard greater than 0.94. When the evidence fails to meet this high standard, the judge opts for *instrument* instead. On my rational reconstruction, the judge doesn't adopt *instrument* because it's been proved; it hasn't been. He adopts *instrument* because he doesn't know what Parliament intended, and the safe choice is the lenient choice.

The judge in *Harris* would have been able to rely on an old presumption known as the "rule of lenity". In past centuries the rule was very strong. As the authors of *Cross on Statutory Interpretation*, if a statute was 'capable of two meanings, however unreasonable one of those meanings be, it was applied if favourable to the accused.'⁴² The presumption was so strong because criminal penalties were draconian. With few formal ways to avoid imposing very harsh penalties, judges resorted to other devices, like the presumption of lenity.⁴³ Nowadays, the criminal law is less punitive, and the rule of lenity is much weaker.

⁴¹ As noted by, e.g., R.E. Barnett, 'Interpretation and Construction' (2011) 34 Harvard Journal of Law and Public Policy 65, 65.

⁴² R. Cross, J. Bell, G. Engle, *Cross on Statutory Interpretation* (3rd edn., Butterworths 1995) 134.

⁴³ The early history of the rule of lenity is set out in L. Hall, 'Strict or Liberal Construction of Penal Statutes' (1935) 48 Harvard Law Review 748, 748-51. See also W. Blackstone, *Commentaries on the Laws of England* (Philadelphia: JB Lippincott Company, 1893) 88 (describing how a statute prohibiting 'stealing sheep, or other cattle' was taken to only prohibit stealing sheep).

It tells courts to opt for the lenient interpretation only when the choice between the two interpretations is evenly balanced.⁴⁴ In other words, the rule plays a role only if the posterior odds of the possible interpretations are 1:1. This bit of history confirms that the rule of lenity isn't a low prior rule. Were it a low prior rule, it would have been strongest when the prior odds of a punitive intention were *lowest*, that is, in modern times. But the opposite is true: the presumption was strongest when the prior odds of a punitive intention were *highest*, that is, in previous centuries.

I'll call presumption rules which tell courts to demand a high standard of proof "high stakes presumption rules" or "high stakes rules" for short.⁴⁵ The notion of a high stakes rule is similar to what American scholars call a 'substantive canon'.⁴⁶ The rule of lenity is a paradigmatic substantive canon; other examples include the presumption against retroactivity and the *Charming Betsy* canon, according to which the legislature is presumed to act consistently with international treaties. 'Most of the substantive canons are hard if not impossible to defend on ordinary-use-of-language or this-is-what-the-legislature-would-want grounds'⁴⁷. The better view is that these rules 'permit a court to forego a statute's most natural interpretation in favour of a less plausible one more protective of a particular value'⁴⁸. No wonder that Lawson says that substantive canons are one of the few occasions 'the law openly says something substantive about standards of proof for initial legal decisions', namely, that 'under certain conditions, propositions of law are to be accepted by decision makers only if they are strongly compelled'⁴⁹.

9. Very strained interpretations

If the prior odds of the natural interpretation are low, or the posterior odds needed to adopt that interpretation are high, then a strained interpretation may be justified. What if the prior odds are low *and* the posterior odds demanded are

⁴⁴ Cross, Bell, and Engle (n 42) 134.

⁴⁵ This is an adaptation of Doerfler's terminology: n 28.

⁴⁶ The other term for a substantive canon is a 'clear statement rule'. That label makes sense if we think of the rule as directing courts to demand a likelihood ratio above a certain high threshold. Why? It's plausible that the clarity of words is a function of their strength as evidence of one meaning over other meanings. If a piece of text is evidence of some meaning, it is more or less "clear" depending on how strong or weak evidence it is. As I said, the strength of evidence is measured by the likelihood ratio. When the prior odds of a meaning are low, or standard of proof is high, then a court is entitled to demand a high likelihood ratio before giving a statute that meaning. In other words, the court is entitled to demand clear words before adopting that meaning.

⁴⁷ W. Eskridge Jr, 'Norms, Empiricism, and Canons in Statutory Interpretation' (1999) 66 University of Chicago Law Review 671, 682.

⁴⁸ A.C. Barrett, 'Substantive Canons and Faithful Agency' (2010) 90 Boston University Law Review 109, 109-10. See also Solum (n 39) 110 (arguing that substantive canons are canons of construction rather than interpretation).

⁴⁹ Lawson (n 4) 59.

high? Then the path is open for a court to reach a *very* strained interpretation. I don't expect that to be surprising. What may be surprising is just how strained a very strained interpretation may be.

Suppose that the prior odds of *A* versus *B* are 1:3. The likelihood ratio needed to achieve even posterior odds (1:1) is therefore 3. Suppose instead that a court demands posterior odds of 3:1 in favour of *A* over *B* to adopt *A*. The likelihood ratio needed to reach this high standard, assuming even prior odds, is again 3. What happens if we combine these assumptions? That is, what happens if the prior odds are 1:3, and the posterior odds demanded are 3:1? What likelihood ratio is needed to adopt *A*? You might initially think that we'd need evidence which is 2x as strong as in either of the initial cases, that is, a likelihood ratio of 6. In fact, we need evidence which is 3x as strong, that is, a likelihood ratio of 9. We don't *add* the likelihood ratios; we *multiply* them. The lower the prior odds, and the higher the posterior odds demanded, the more dramatic the effect of their combination on the necessary likelihood ratio. For example, if the prior odds are 1:100, reaching even posterior odds takes a likelihood ratio of 100. If the prior odds are even, then reaching posterior odds of 100:1 again takes a likelihood ratio of 100. But, if the prior odds are 1:100, reaching posterior odds of 100:1 requires a likelihood ratio of 10,000 – that is, it takes evidence 100x stronger than is needed given either low prior odds or a high standard alone. If we simply added the effects of the low prior odds and the high standard, we would have thought we'd need a likelihood ratio of $100 + 100 = 200$, and thus have underestimated the strength of the evidence needed by a factor of $10,000/200 = 50$.

The points in the last paragraph have not to my knowledge been made before in the interpretive context. Courts intuitively know that low prior odds matter. They intuitively know that posterior odds matter. But without using Bayes' theorem, it's hard to see that combining low prior odds with a demand for high posterior odds can dramatically increase the strength of the evidence needed to adopt certain interpretations. Because it's hard to see, I suspect the case for strained interpretations is often stronger than is realised. I want to show this by considering what is probably the best known strained interpretation of a statute, in *Anisminic Ltd. v Foreign Compensation Commission*.

10. *Anisminic Ltd. v Foreign Compensation Commission*

Following the Suez incident, the property of Anisminic Ltd., a British company, was 'sequestered' by the Egyptian government. Later, Egypt paid millions of pounds to the British government to help compensate those whose property was sequestered. That sum was administered by the Foreign Compensation Commission (FCC) under the Foreign Compensation Act 1950. Anisminic applied to the FCC for compensation, but the FCC (purported to) determine that Anisminic was ineligible. When Anisminic applied for a declaration that the

determination was a nullity, the FCC argued that the court couldn't hear the claim by virtue of s 4(4) of the 1950 Act, which reads:

The determination by the commission of any application made to them under this Act shall not be called in question in any court of law.

The central issue in the case was the meaning of s 4(4). Did it 'oust' the courts' jurisdiction? Or, did s 4(4) have some other effect, such that a court could decide whether the FCC had acted lawfully?

The FCC argued that the meaning of s 4(4) was plain: if the FCC receives an application, and comes to a view on its eligibility, then its view can't be questioned in a court. Thus one possible interpretation of s 4(4) is:

Oust

The courts can't question the FCC's determinations, as ordinarily understood.

The alternative interpretation of s 4(4) is a little more complicated. The idea is that there is a difference between determinations, which are lawful, and mere purported determinations, which are unlawful and therefore have no legal existence. 'It is one thing to question a determination which does exist: it is quite another to say that there is nothing to be questioned'⁵⁰. Section 4(4) prohibits the calling into question of real or actual determinations, not purported determinations.

Preserve

The courts can't question the FCC's legally valid determinations.

The first interpretation would oust the courts; the second would preserve their jurisdiction.

Let's suppose that Parliament would be *certain* to use language like that found in the 1950 Act assuming it intends to oust the courts, such that:

$$P(\text{text}|\text{oust}) = 1$$

Let's also suppose that Parliament would only very rarely use language like that found in the 1950 Act assuming it intends to preserve the jurisdiction of the courts, such that:

$$P(\text{text}|\text{preserve}) = 0.01$$

Although there's a chance that Parliament would use language as in the 1950 Act when it wants to preserve the role of the courts, it's much more likely that it would use that language when it wants to oust the courts. The likelihood ratio is:

⁵⁰ *Anisminic* (n 2) 170 (per Lord Reid).

$$\frac{P(\text{text}|\text{oust})}{P(\text{text}|\text{preserve})} = \frac{1}{0.01} = 100$$

The likelihood ratio strongly favours *oust*, so *preserve* would be a very strained interpretation.

Once the likelihood ratio is established, we need to consider the prior odds of *oust* versus *preserve*. Lord Reid begins by noting that ouster clauses are not just rare, but unheard-of. Conferring powers on officials and tribunals is the bread and butter of legislation; Parliament has created this sort of power thousands and thousands of times. And yet '[n]o case has been cited in which any other form of words limiting the jurisdiction of the court has been held to protect a nullity'.⁵¹ In addition, ouster clauses are constitutionally disreputable. Ouster clauses shield officials from effective, external constraints. This threatens the rule of law – a principle which Parliament is thought to value and to want to uphold.⁵² If accurate, these points suggest a very low prior probability that Parliament intends to oust the courts. Suppose that the prior odds are 1:50 against *oust*.

Next, we consider the appropriate standard of proof. Given that the rule of law is of great value, and that *oust* threatens the rule of law, *Anisminic* is a high stakes case. The court should therefore demand a high standard of proof to accept that what looks like an ouster clause really is an ouster clause. Lord Reid seems to suggest as much when he invokes the 'well established principle':

that a provision ousting the ordinary jurisdiction of the court must be construed strictly – meaning, I think, that if such a provision is reasonably capable of having two meanings, that meaning shall be taken which preserves the ordinary jurisdiction of the courts.⁵³

What counts as a reasonable doubt as to the meaning of a statute Lord Reid doesn't say. But the rule of law is often said to be of prime importance in the British constitution.⁵⁴ If that's right, then a very high standard would be appropriate, similar to the standard required by the old rule of lenity, i.e., > 19:1 or > 0.95.

⁵¹ *Anisminic* (n 2) 170 (per Lord Reid).

⁵² See e.g. M. Elliott, *The Constitutional Foundations of Judicial Review* (Hart 2001) 111 (Parliament intends 'that the rule of law should be upheld'). Judges often express similar claims, e.g., *R v Secretary of State for the Home Department, ex p Pierson* [1998] AC 539, 591 (per Lord Steyn) ('Parliament must be presumed not to legislate contrary to the rule of law'). For a more sceptical view, see T. Endicott, 'Constitutional Logic' (2003) 53 *University of Toronto Law Journal* 201, 208 ('There is no general reason to believe that Parliament at Westminster does not intend to act contrary to the rule of law'). If a view like Endicott's is correct, then the prior odds against *oust* should be shorter than I suggest in the main text, but it wouldn't detract from my main point – i.e., given the court's starting assumptions, it's possible to justify a surprisingly strained interpretation of s 4(4).

⁵³ *Anisminic* (n 2) 170.

⁵⁴ For recent judicial paens to the rule of law, with special reference to the judicial role, see *R (Evans) v Attorney General* [2015] UKSC 21, [52]-[58] (per Lord Neuberger); *R (UNISON) v Lord Chancellor* [2017] UKSC 51, [66]-[89] (per Lord Reed).

At this point, Lord Reid's adoption of the second, preserving interpretation should look obvious. Bayes' theorem tells us that the posterior odds are:

$$\frac{P(oust)}{P(preserve)} \times \frac{P(text|oust)}{P(text|preserve)} = \frac{P(oust|text)}{P(preserve|text)}$$

With the odds we've estimated, we know the following:

$$\frac{1}{50} \times \frac{100}{1} = \frac{100}{50} = \frac{2}{1}$$

In other words, it's twice as probable that *oust* is Parliament's intended meaning than that *preserve* is. But twice as probable is far short of the high standard of proof necessary. Even though *oust* is the natural interpretation, it's not nearly proved. The court should resort to the safe choice – in this case, to take the statute to mean that the courts' reviewing role is preserved. No wonder that Lord Reid said he had 'come without hesitation to the conclusion that in this case we are not prevented from inquiring whether the order of the commission was a nullity'⁵⁵.

Anisminic has plenty of other defenders; so what's new here? One thing is the nature of my defence. Other scholars have noted the significance of the intrinsic improbability that Parliament intended to oust the courts.⁵⁶ But, if I'm right, that factor alone doesn't justify the preserving interpretation. Even assuming long prior odds of 1:50, the statute was clear enough (i.e., the likelihood ratio was large enough) to prove *oust* on a civil standard. Nor, for that matter, would it have been enough to demand high posterior odds to adopt *oust*. Starting with even prior odds, the statute was clear enough to prove *oust* on even a high standard of > 19:1. The best defence is a compound defence. The court should adopt *preserve* because it should *both* set the prior odds of *oust* low and *and* demand high posterior odds to adopt *oust*. Both elements are crucial.

Anisminic is often said to be at odds with parliamentary sovereignty. The House of Lords knew that Parliament intended to oust the courts, and it chose not to give effect to that intention. The House of Lords didn't reject parliamentary sovereignty explicitly, but that's what it did in substance. Legislative intent was a "fig leaf"; in reality, the court invented a meaning and attributed it to Parliament.⁵⁷ That's the thought, anyway. But my analysis suggests there's no tension between *Anisminic* and parliamentary sovereignty. The House of Lords didn't know that Parliament intended to oust the courts. So, by interpreting the statute to mean that Parliament intended to preserve the role of courts, the House of Lords didn't defy Parliament's will.

⁵⁵ *Anisminic* (n 2) 171.

⁵⁶ See e.g. J. Goldsworthy, *Parliamentary Sovereignty: Contemporary Debates* (CUP 2010) 286.

⁵⁷ For suggestions along these lines, see HWR Wade and CF Forsyth, *Administrative Law* (11th edn, OUP 2011) 614; V. Bogdanor, 'Imprisoned by a Doctrine: The Modern Defence of Parliamentary Sovereignty' (2012) 32 OJLS 179, 191; M. Gordon, *Parliamentary Sovereignty in the UK Constitution: Process, Politics, and Democracy* (Hart 2015) 122.

Anisminic is often seen as a hard case, with the arguments on either side finely balanced.⁵⁸ But I think that *Anisminic* was an easy case, and that the argument for *preserve* was overwhelming. Given the low prior odds (1:50) and high standard of proof (>19:1), it would require a likelihood ratio of ~1000 to justify *oust*. Another way to express that ratio is that the odds would need to be about 1000:1 against seeing the language in s 4(4) given an intention to enact *preserve*. It helps to put those odds in perspective. At the time of writing, you can place a bet at 250:1 that Nigel Farage, a nemesis of the Conservative party, will be the next leader of that party.⁵⁹ Just before William and Kate, the Duke and Duchess of Cambridge, had their third baby, you could place a bet at 500:1 that they would name the baby “Chardonnay”.⁶⁰ It’s admittedly improbable that Parliament intended to preserve the role of the courts and simply chose its words poorly in s 4. Surely, though, it’s not as improbable as Farage leading the Conservative party or a royal baby named Chardonnay. If that’s right, then the necessary likelihood ratio wasn’t nearly achieved. (It’s no surprise that *Anisminic* is thought of as a hard case, even though it was an easy case. *Anisminic* is an easy case because of the effect of combining low prior odds and a demand for high posterior odds. And, as I said in the last section, it’s difficult to appreciate the significance of combining these two factors without using Bayes’ theorem.)

This last point has a forward-looking implication. Because *Anisminic* was an easy case, it would have required much clearer language than in s 4 to oust the courts. That should give courts in later cases pause before concluding that *Anisminic* has been distinguished, and that the courts have been ousted.⁶¹ Ouster clauses are harder to overcome than even defenders of *Anisminic* have realised.

11. Impossible interpretations?

There are some meanings which it may be effectively impossible for a legislature to enact. Equally there may be some presumptions which are effectively irrebuttable. Suppose that the prior odds that Parliament intends to enact some meaning are 1:200. And suppose that, given the nature of the meaning, we demand posterior odds of 100:1 or better to adopt it. If legislative language is used imperfectly on more than 1 out of 20,000 occasions, that is, more than 0.005% of the time, then it will be impossible for Parliament to communicate the intention at issue. The necessary likelihood ratio can never be achieved. It doesn’t matter what

⁵⁸ See e.g. DM Gordon, ‘What did the *Anisminic* Case Decide?’ (1971) 34 MLR 1, 11.

⁵⁹ ‘British Politics – Next Conservative Leader Betting Odds’ at <<https://www.oddschecker.com/politics/british-politics/next-conservative-leader>> accessed 21 April 2018.

⁶⁰ ‘William and Kate – Royal Baby Name Betting Odds’ at <<https://www.oddschecker.com/novelty/specials/william-and-kate/royal-baby-name>> accessed 21 April 2018.

⁶¹ *R (Privacy International) v Investigatory Powers Tribunal* [2017] EWCA Civ 1868 (finding that a clause effectively ousts the courts).

words we see in the statute; a court should never find that Parliament has sufficiently clearly indicated that intention. Parliament is limited in what it can enact into law.⁶²

This claim will sound very bold, so let me immediately emphasise three points. The first point is about the nature of the limit I'm proposing. We're used to saying that Parliament's law-making powers are unlimited. And so they are, in a sense: the law gives Parliament the power to make or unmake law on any matter. But there's a difference between having a legal power and having the practical ability or capacity to use that power. I may have the legal power to accept your offer to contract; but, if I'm unable to convey my acceptance (because, say, I'm out of mobile phone range), then I have a power which I am, for purely practical reasons, unable to use. Analogously, Parliament may lack the ability to use its law-making powers simply because it can't convey its intention sufficiently clearly. Parliament would be limited, not by law, but by probability.

Second, my claim supposes very low prior odds and a very high standard of proof. Odds of 1:200 and 100:1 are longer and shorter, respectively, than the prior and posterior odds in *Anisminic*. So whatever intention we're imagining, it must be both rarer and more troubling than ousting the courts in a specific context. For public lawyers, the natural thought at this point will be some of the provocative *obiter dicta* from senior judges in recent years. In *Jackson v Attorney General*, Lord Steyn speculated that Parliament lacked the power to 'abolish judicial review or the ordinary role of the courts'⁶³. In the same case Lady Hale said that the courts might reject 'any attempt to subvert the rule of law by removing governmental action affecting the rights of the individual from all judicial scrutiny'⁶⁴. In *Moohan v Lord Advocate*, Lord Hodge refused to 'exclude the possibility' that Parliament lacked the power to 'entrench its power by a curtailment of the franchise or similar device'⁶⁵. These judges are suggesting that Parliament may be unable to abolish judicial review and the like because Parliament may not be sovereign. I suggest a different basis for a similar conclusion: there may be some things which Parliament can't do, because some things are so beyond the pale that no statutory language can convince us that Parliament truly meant them.

Third, my claim is *not* that there are things which Parliament is unable to enact under *any* conditions. It's that there are things which Parliament is unable to enact under *certain* conditions, i.e., given certain facts about the prior odds, standard of proof, and imperfection of language. When those conditions change, what Parliament can do changes. Assume (purely for the sake of argument) that the prior odds that Parliament intends to abolish judicial review are 1:200. Assume we

⁶² My thinking along these lines was sparked by Listokin (n 1) 371

⁶³ *R. (Jackson) v Attorney General* [2005] UKHL 56, [2006] 1 A.C. 262, [102] (per Lord Steyn). See also *AXA General Insurance Ltd v HM Advocate* [2011] UKSC 46, [2012] 1 A.C. 868, [50]-[51] (per Lord Hope).

⁶⁴ *Jackson v Attorney General* (n 61) [159] (per Lady Hale).

⁶⁵ *Moohan v Lord Advocate* [2014] UKSC 67, [2015] AC 901, [35] (per Lord Hodge).

apply a standard of proof equivalent to 100:1. Finally, assume that there is a 1 in 20,000 chance that in this context Parliament will use language imperfectly (because of omission, drafting error, etc.). Under these assumptions, Parliament can't abolish judicial review. Now imagine a world in which Parliament has a long track record of systematically violating the rule of law. It would be unreasonable to continue to estimate prior odds of 1:200. The odds should be much shorter. In this dystopia, it *is* possible for Parliament to convey its intention to abolish judicial review. What's possible in our world depends on the factual conditions we find here from time to time.

12. Summary and conclusion

I had two aims in this article. The main aim was to answer a series of fundamental questions about interpretation. What makes an interpretation strained, I asked? That the odds are better that we'd see a certain statutory text and context assuming Parliament intended to enact the alternative, "natural" interpretation. What justifies a strained interpretation? One justification is that the prior odds of the natural meaning are so low that the strained meaning is proved instead. Another justification is that neither meaning is proved and the strained choice is the safe choice. When we combine these justifications, it becomes possible to justify highly strained interpretations. Is there always a point past which an interpretation is *too* strained? It depends on the prior odds and the necessary posterior odds. But, at least in principle, there could be meanings which it's impossible for the legislature to enact. In the process of showing these points, I've also proposed a partial typology of presumptions and offered a new perspective on *Anisminic*, an old yet still controversial case.

My secondary aim was to contribute to the larger project of bridging the gap between questions of proof and questions of interpretation. I've done so, first, by showing the relevance of Bayes' theorem, a familiar topic in evidence scholarship, to matters of interpretation; and also by showing that some interpretive issues and tools are usefully conceived in terms of the standard and burden of proof. I've said almost nothing about a range of other interesting questions about the relationship between proof and interpretation. For example: (i) who in general should bear the burden of proof on matters of interpretation? (ii) How much weight should a court give to the text of a statute compared with, for example, statements in *Hansard*? (iii) How should we think of courts' occasional willingness to defer to administrative interpretations of statutes? Is it a shift in burden of proof, an increase in the standard of proof, or something else? These are just a few examples of what I hope to have helped show is a fertile ground of future research.