What About the Precedent:
An Information-Theoretic Analysis of Common Law

Josef Valvoda[ð] **Tiago Pimentel**[ð] **Niklas Stoehr**[ſ] **Ryan Cotterell**[ð,ſ] **Simone Teufel**[ð]

[ð]University of Cambridge, [ſ]ETH Zürich
`jv406@cam.ac.uk`, `sht25@cam.ac.uk`

Abstract

In common law, the outcome of a new case is determined mostly by precedent cases, rather than by existing statutes. However, how exactly does the precedent influence the outcome of a new case? Answering this question is crucial for guaranteeing fair and consistent judicial decision-making. We are the first to approach this question computationally by comparing two longstanding jurisprudential views; Halsbury's, who believes that the arguments of the precedent are the main determinant of the outcome, and Goodhart's, who believes that what matters most is the precedent's facts. We base our study on the corpus of legal cases from the European Court of Human Rights (ECtHR), which allows us to access not only the case itself, but also cases cited in the judges' arguments (i.e. the precedent cases). Taking an information-theoretic view, and modeling the question as a case outcome classification task, we find that the precedent's arguments share 0.38 nats of information with the case's outcome, whereas precedent's facts only share 0.18 nats of information (i.e., 58% less); suggesting Halsbury's view may be more accurate in this specific court. We found however in a qualitative analysis that there are specific statues where Goodhart's view dominates, and present some evidence these are the ones where the legal concept at hand is less straightforward.

1 Introduction

Legal systems around the world can be divided into two major categories (Joutsen, 2019): civil law systems, which rely predominantly on the rules written down in statutes, and common law systems, which rely predominantly on past judicial decisions, known as the precedent. Within common law systems, jurisprudential scholars have pondered over the nature of precedent in law for at least a century (Halsbury, 1907). Is it the judges' argumentation in the precedent, or is it the
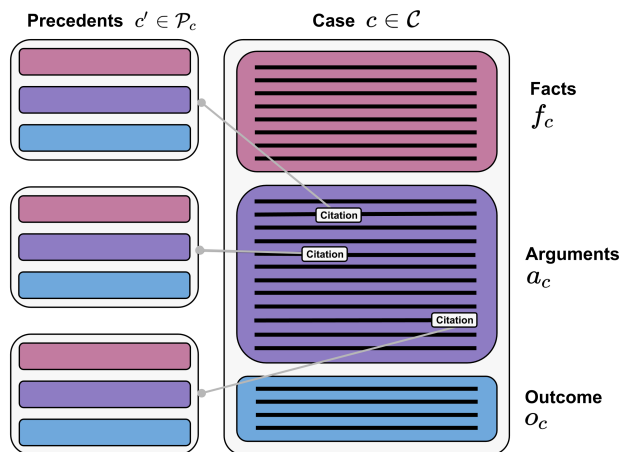

Figure 1: The text of ECtHR cases can be divided into *facts*, *arguments* and *outcome*. *Arguments* cite relevant cases, also known as the *precedent*.

claimants' specific individual circumstances that are the deciding factor in what becomes the law? Here, we present a new information-theoretical methodology that helps answer this question.

In common law countries, statutes establish the general idea of the law, but the actual scope of the law is determined by the courts during a trial. To keep case outcomes consistent and predictable in subsequent cases, judges are forced to apply the reasoning developed in prior cases with similar facts (precedent), to the facts of the new case under the doctrine of *stare decisis* (Duxbury, 2008; Lamond, 2016; Black, 2019). This is done by identifying the *ratio decidendi* (the reasons for the decision) as opposed to the *obiter dicta* (that which is said in passing). The distinction between ratio and obiter is an important one, since ratio is binding, whereas obiter is not. This means that courts will only strive to remain consistent in upholding ratio, but can freely depart from the obiter.

But what does the ratio consist of? There is no accepted overarching theory of precedent (Duxbury, 2008), but there are two tests of ratio. On the one hand, Lord Halsbury (1907) claims

that what is binding is the judge's reasoning and arguments. For instance, by using a high degree of abstraction, judges can analogise physical and psychological pain. A different view has been put forward by Goodhart (1930), who argues it is the analogy of the facts of the precedent and the case at hand, without the need for reasoning (e.g. comparing the pain caused by a knife to that caused by another instrument, requiring a far lower degree of abstraction). These give rise to the two well-known legal tests for ratio: **Halsbury's test** and **Goodhart's test**.

In this paper, we are the first to approach this problem from a data-driven perspective, using the European Court of Human Rights (**ECtHR**)[1] case law; see Figure 1. We build a citation network over this corpus in order to have access to many precedents' full text. Training our model on either the facts or the arguments of the precedent, we can put Halsbury's and Goodhart's views to the test. We cast this problem as an information-theoretic study by measuring the **mutual information** (Shannon and Weaver, 1962) between the case outcome and either the precedent facts or arguments. We find that precedent arguments and case outcome share information to the degree of 0.38 nats, whereas facts and case outcome only share information to the degree of 0.18 nats (i.e., 58% less). We therefore observe that—at least for ECtHR—Halsbury's view of the precedent is more accurate than that of Goodhart.

2 Legal Background

Despite the importance of the precedent in civil law, its operationalization remains shrouded in philosophical debate centred around *how* the precedent actually forms the binding law. Jurisprudentially, we can think of this as searching for the ratio decidendi in the judgement, i.e. separating the ratio decidendi from the obiter dicta, or binding law from merely circumstantial statements. It is the nature of ratio that distinguishes Halsbury's view from Goodhart's.

2.1 Halsbury: Arguments as ratio

The case argument contains the judge's explanation of why the case is decided the way it is. It incorporates knowledge of the precedent, facts of the case

and any new reasoning the judge might develop for the case itself. We consider the intuitive position that a legal test is formulated by the argument that the judge put forward when deciding the case.

A legal test is by its nature part of the ratio and, thus, would be binding on all subsequent cases. This is the position endorsed by Lord Halsbury (1907). Under this conception of the ratio, it is the arguments that matter, becoming the law; the facts of the case are of secondary importance. If a judge acts as Halsbury suggests they should extract the logic of the implicit legal test of the precedent, and attempt to largely ignore the specific facts of the case. Halsbury's view remains the conventional view of the precedent to this day (Lamond, 2005).

2.2 Goodhart: Facts as ratio

In contrast, Goodhart (1930) observes that many cases do not contain extensive reasoning, or any reasoning at all; judges seem to decide the outcome without these. Therefore, he claims that the facts of the case together with its outcome must form the ratio; otherwise, a hypothetical new case with the same facts as any given precedent could lead to a different outcome. Duxbury (2008) observes that judges, when in disagreement with the precedent, concentrate on the facts of a previous case more than one would expect if Halsbury's hypothesis were fully correct. Halsbury would predict that they should talk about the facts of previous cases as little as possible, and seek the most direct route to ratio in the form of argument, but they evidently do not. A potential explanation is that, when disagreement arises, it is easier for judges to claim that the facts are substantially different, than to challenge the logic of the precedent, i.e. to overrule that case. Overruling a previous judgement is a rare and significant legal event (Dunn, 2003; Spriggs and Hansford, 2001) because it threatens the stability of the legal system. By concentrating on facts rather than running the risk of overruling, the judge can avoid this problem, including the threat of overruling her own previous judgement.

In support of this view, inspection of the argumentative part of the judgement reveals judges do not usually formulate legal tests of the kind Halsbury implies (Lamond, 2005). Neither do judges usually search the precedent for such legal tests (Alexander and Sherwin, 2008). Goodhart's position suggests that the precedent operates less as an enactment of rules, but more as reasoning by

---

[1]European Court of Human Rights (ECtHR) is the court that adjudicates on cases dealing with the European Convention of Human Rights (ECHR).
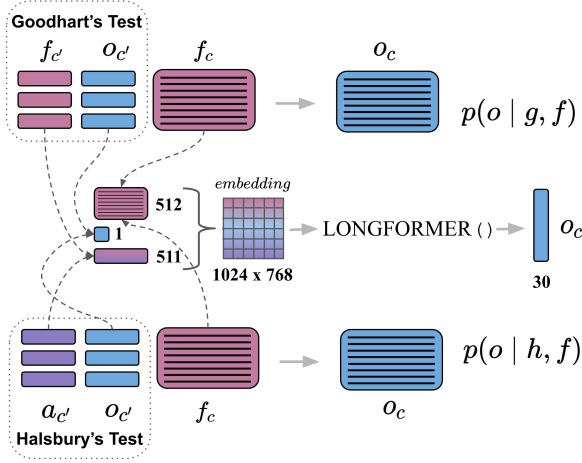
Figure 2: Our formulation of Halsbury's and Goodhart's tests as a classification task. Current case facts are truncated to 512 tokens. Outcome of the precedent is concatenated with either the precedent facts or arguments, and both are jointly truncated at 512 tokens. Finally, these are concatenated together and embedded in 768 dimensions before being fed into the LONGFORMER.

analogy; hence it is the good alignment between the facts of the two cases that leads to consistent outcomes.

3 An Information-theoretic Approach

Notation. We denote the set of cases as $\mathcal{C}$, writing each of its element as c. The set of cases that form the precedent for case c are denoted $\mathcal{P}_c \subset \mathcal{C}$. We will consider three main random variables in this work. First, we consider O, a random variable that ranges over a binary outcome space $\mathcal{O} = \{0,1\}^K$, where K is the number of Articles. An instance $o \in \mathcal{O}$ tells us which Articles have been violated. Since o is a vector of binary outcomes for all Articles, we can index it as o_k to get the outcome of a specific k^{th} Article and we analogously index the random variable O_k. We will denote o_c the outcome of a specific case c.[2] Next, we consider F, a random variable that ranges over the space of facts. We denote the space of all facts as $\mathcal{F} = \Sigma^*$, where Σ is a set of sub-word units and Σ^* is its Kleene closure. We denote an instance of F as f. We will further denote the facts of a specific case c as f_c. Finally, we consider A, a random variable that ranges over the space of Arguments. Analogously to facts, the space of all Arguments is $\mathcal{A} = \Sigma^*$. An element of $\mathcal{A}$ is denoted

as a, which we again term a_c when referring to a specific case.

Operationalising Halsbury and Goodhart. In this work, we intend to measure the use of Halsbury's and Goodhart's views in practice, which we operationalise information-theoretically following the methodology proposed by Pimentel et al. (2019). To test the hypothesis, we construct two collections of random variables, which we denote H and G. We define an instance h_c of random variable H as the union of arguments and outcomes for all precedent cases of c, i.e. $\bigcup_{c' \in \mathcal{P}_c} \{a_{c'}, o_{c'}\}$. We will denote the instance h when referring to it in the abstract (without referring to a particular case). We analogously define instances of random variable G as $g_c = \bigcup_{c' \in \mathcal{P}_c} \{f_{c'}, o_{c'}\}$. While the set-theoretic notation may seem tedious, it encompasses the essence of the distinction between Halsbury's and Goodhart's view: Each view hypothesises a different group of random variables should contain more information about the outcome O of a given case. In terms of mutual information, we are interested in comparing the following:

$$\text{MI}(O; H \mid F), \quad \text{MI}(O; G \mid F) \tag{1}$$

If $\text{MI}(O; H \mid F) > \text{MI}(O; G \mid F)$, then Halsbury's view should be more widely used in practice. Conversely if the opposite is true, i.e. $\text{MI}(O; G \mid F) > \text{MI}(O; H \mid F)$, then Goodhart's view should be the one more widely used.

The MI is calculated by subtracting the outcome entropy conditioned on the case facts and either H or G from the outcome entropy conditioned on the facts alone. Therefore, to compute the MI we need to compute the Halsbury's and Goodhart's conditional entropies first:

$$\text{H}(O \mid H, F) \tag{2}$$
$$= -\sum_{o,h,f} p(o,h,f) \log p(o \mid h, f)$$
$$\text{H}(O \mid G, F) \tag{3}$$
$$= -\sum_{o,g,f} p(o,g,f) \log p(o \mid g, f)$$

as well as the entropy conditioned on the facts of the current case alone:

$$\text{H}(O \mid F) = -\sum_{o,f} p(o,f) \log p(o \mid f) \tag{4}$$

---

[2] We note here that we overload the subscript notation in this paper. We will use subscript c to denote a specific case and subscript k to denote a specific article.

The conditional entropies above reflect the uncertainty (measured in nats)[3] of an event, given the knowledge of another random variable. For instance, if G completely determines O, then $H(O \mid G)$ is 0; there is no uncertainty left. Conversely, if the variables are independent, then $H(O) = H(O \mid G)$, where $H(O)$ denotes the unconditional entropy of the outcomes O. We now note a common decomposition of mutual information that will help with the approximation:

$$MI(O; H \mid F) = H(O \mid F) - H(O \mid H, F) \quad (5)$$
$$MI(O; G \mid F) = H(O \mid F) - H(O \mid G, F) \quad (6)$$

In this work, we consider the conditional probabilities $p(o \mid \bullet)$ as the independent product of each Article's probability, i.e. $\prod_{k=1}^{K} p(o_k \mid \bullet)$. Information-theoretically, then, they are related through the following equation:

$$H(O \mid \bullet) = \sum_{k=1}^{K} H(O_k \mid \bullet) \quad (7)$$

Following Williams et al. (2020), we further calculate the uncertainty coefficient (Theil, 1970) of each of these mutual informations. These coefficients are easier to interpret, representing the percentage of uncertainty reduced by the knowledge of a random variable:

$$U(O \mid H; F) = \frac{MI(O; H \mid F)}{H(O \mid F)} \quad (8)$$
$$U(O \mid G; F) = \frac{MI(O; G \mid F)}{H(O \mid F)} \quad (9)$$

4 Experimental Setup

We choose to work with the ECtHR corpus for three reasons. First, it can be treated as operating under precedential law, in the vein of common law countries. This is not a given, as the ECtHR is an international court of highest appeal without a formal doctrine of *stare decisis* (Jacob, 2014), but there is nevertheless strong evidence that it is precedential. This evidence comes from the court's own guidelines (ECtHR, 2014), but can also be found in the writings of a former judge of the ECtHR (Zupancic, 2016) and of legal scholars (Lupu and Voeten, 2010). Second, there is existing research on the neural modeling of ECtHR case law we can build upon (Aletras et al., 2016; Chalkidis et al., 2019, 2020). Third, the documents of the ECtHR

case law, unlike those of most other courts, textually separate the *facts* from the *arguments*, which is crucial for our experiments.

Case facts are descriptions of what had happened to the claimant before they went to the court; they include domestic proceedings of their case before it was appealed to the ECtHR as a form of a last resort. They do not contain any reference to European Convention of Human Rights (ECHR) Articles or ECtHR case law. Arguments on the other hand contain judges' discussion of ECHR articles and ECtHR case law in relation to the facts. The ECtHR corpus has been scraped from the HUDOC[4] database and contains 11,000 cases reported in English (Chalkidis et al., 2019).[5] Judges decide for each Article of ECHR whether it has been violated with respect to the claimant's circumstances. In the ECtHR corpus, each case therefore comes with a pre-extracted decision in form of a set of violated ECHR Article numbers. We refer to this set as the *outcome* of a case. Out of 30 Articles, 18 are from the Convention itself (Articles 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 18, 34, 38, 41, 46), while the rest (1.1, 1.2, 1.3, 4.2, 4.4, 6.1, 6.3, 7.1, 7.2, 7.3, 7.4, 12.1) comes from the Protocols to the Convention.

For our experiment, we need a sub-corpus where each case has at least one outgoing citation where the full text is contained in our corpus. In practice, there will be other outgoing citations we cannot resolve, for instance because the document is not in English or HUDOC happens not to contain them. We also need our citations to be de-duplicated. We create such a sub-corpus, which contains 9,585 documents (i.e., citing documents), with 94,167 in-corpus links (tokens) to 7,113 cases (types) and 65,495 out-of-corpus links to 22,328 types (cited documents). We start from the original ECtHR split of 9,000 training, 1,000 validation and 1,000 test cases, and after citation filtering arrive at 7,627 training, 976 validation and 982 test cases. For every citation, we extract the text under headings with regular expressions such as "THE FACTS" and "THE LAW", labelling it as facts and arguments, respectively.

4.1 Approximations

The mutual information values that we intend to analyse need to be approximated. We follow Pi-

---

[3]Nats are computed with ln, while bits use $\log_2$.

[5]ECtHR cases are reported either in English, French or both. Additionally, some cases are also reported in the language of the state they take place in.

mentel et al.'s (2019; 2021) methodology for this, approximating them as the difference between two cross-entropies:

$$\mathrm{MI}(O; H \mid F) \approx \mathrm{H}_\theta(O \mid F) - \mathrm{H}_\theta(O \mid H, F)$$
$$\mathrm{MI}(O; G \mid F) \approx \mathrm{H}_\theta(O \mid F) - \mathrm{H}_\theta(O \mid G, F)$$

Indeed, although several estimates for the mutual information exist, McAllester and Stratos (2020) argues that estimating it as this difference is the most statistically justified way. These conditional entropies are themselves approximated through their sample estimate. For instance, we compute:

$$\mathrm{H}_\theta(O \mid G, F) \approx -\frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \log p_\theta(o_c \mid g_c, f_c)$$
(10)

which is exact as $|\mathcal{C}| \to \infty$. We note that the cross-entropy is an upper bound on the entropy, which uses a model $p_\theta(o \mid \bullet)$ for its estimate. The better this model, the tighter our estimates will be. The only thing left to do now, is to obtain these probability estimates. We thus model Halsbury's view as a classification task (see Figure 2) estimating the probability:

$$p_\theta(o \mid h, f) = \prod_{k=1}^{K} p_\theta(o_k \mid h, f)$$
(11)

We analogously model Goodhart's view as:

$$p_\theta(o \mid g, f) = \prod_{k=1}^{K} p_\theta(o_k \mid g, f)$$
(12)

Finally, we model the p_θ of the model conditioned only on the facts of the case at hand as:

$$p_\theta(o \mid f) = \prod_{k=1}^{K} p_\theta(o_k \mid f)$$
(13)

These models can be approximated using deep neural networks as introduced in the next section. We train deep neural networks on our training sets, using a cross-entropy loss function and a sub-gradient descent method. Given the trained models, we can then answer if it is Halsbury's view or Goodhart's that is more widely used by the ECtHR judiciary.

4.2 Implementation Details

All experiments are conducted using a LONG-FORMER classifier (Beltagy et al., 2020).[6] The

---

[6]Our code is available here: https://github.com/valvoda/Precedent.

LONGFORMER is built on the same TRANS-FORMER (Vaswani et al., 2017) architecture as BERT (Devlin et al., 2019), but allows for up to 4,096 tokens, using an attention mechanism which scales linearly, instead of quadratically. We choose this architecture in particular as it achieves state-of-the-art performance in tasks similar to ours, e.g. on the IMDB sentiment classification (Maas et al., 2011) and Hyperpartisan news detection (Kiesel et al., 2019).

To find the probability of violation of the K Articles we compute:

$$\mathbf{h} = \text{LONGFORMER}(g, f)$$
(14)
$$p_\theta(o \mid g, f) = \sigma(W^{(1)} \,\mathrm{ReLU}(W^{(2)} \,\mathbf{h}))$$

where $\mathbf{h} \in \mathbb{R}^{d_1}$ is a high dimensional representation, $W^{(1)} \in \mathbb{R}^{K \times d_2}$ and $W^{(2)} \in \mathbb{R}^{d_2 \times d_1}$ are learnable parameters in linear projections, and σ is the sigmoid function. Eq. (14) will thus output a K-dimensional vector with the probabilities for all articles, by indexing this vector we retrieve the probabilities of the individual articles applying. Due to resource limitations we set the models' hidden size to 50 and batch size to 16, and also truncate individual cases to 512 tokens. For the models $p_\theta(o_k \mid g, f)$ and $p_\theta(o_k \mid h, f)$, which are trained on the combination of f and either h or g, we concatenate cases to the maximum length of 1,024 tokens (as exemplified in Figure 2). While we do not fully utilise the 4,096 word limit of the LONGFORMER, we are able to process twice as many tokens as standard BERT without pooling; memory limitations prevent us from using the full 4,096 tokens, though.

Our LONGFORMER models are implemented using the Pytorch (Paszke et al., 2019) and Hugginface (Wolf et al., 2020) Python libraries. We train all our models on 4 Nvidia P100 16GiB GPU's for a maximum of 6 hours using LONGFORMER-base model. Our results are reported in terms of the models cross entropy.

Model Input	H_θ	MI	U
Facts	2.99	-	-
Goodhart	2.81	0.18	6%
Halsbury	2.68	0.31	10%

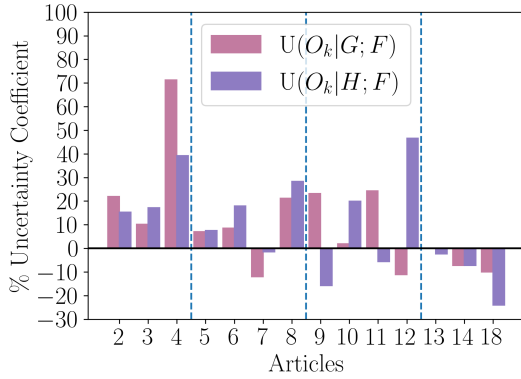Table 1: The cross entropy H_θ, mutual information MI and uncertainty coefficient U results.

Figure 3: Uncertainty coefficient for the Articles of the ECHR Convention.

5 Results

Our experimental results are contained in Table 1. We first note that both our mutual information estimates are statistically larger than zero, i.e. Goodhart's and Halsbury's cross-entropies are statistically smaller than that of the Facts.[7] The question we asked ourselves at the outset, though, concerns whether the data supports Halsbury's or Goodhart's view. We find that our estimate of $\mathrm{MI}(O; H \mid F)$ is significantly larger at 0.31 nats than our estimate of $\mathrm{MI}(O; G \mid F)$ at 0.18 nats. These results suggest that the information contributed by the precedent arguments give us nearly 58% more information about the outcome of the case than the information contained in the facts of the precedent. In terms of the uncertainty coefficient, the outcome entropy is reduced by 6% for facts and by 10% for arguments. We therefore observe that Halsbury's view is more widely used in the domain of ECtHR than Goodhart's.

6 Discussion & Analysis

A more nuanced story can be told if we inspect the individual Articles even though the small number of cases per Article does not allow for conclusive significance tests. The core rights of the Convention are contained in (Articles 2-18).[8] Figure 3 shows that for some of the core Articles, we see the opposite effect from what we observed

for the entirety of Articles, namely that facts outperform arguments, in particular for Articles 2, 4, 9, 11, 13, 18.

We hypothesise that the reason for this is either that the judges have not yet developed a functional legal method for these Articles, that the relevant precedent has been placed late in the list of precedents (and thus was truncated away by our methodology), or that the complexity of the arguments requires a reasoning ability our models are simply not capable of. We consider each hypothesis separately below.

6.1 Conceptual Uncertainty

For some Articles, it is more difficult to develop a legal method than for others because the logic of the argument is elusive for some reason. This holds, for instance, for Articles encoding a vague concept such as "right to life", cf. the discussion below. If a case deals with such an Article, the argument of a potential precedent will be less useful to determine the outcome. We hypothesise that in such a case the judges will be more willing to depart from the logic of past cases, which they might perceive as unsatisfactory in search of a better legal reasoning. However, judges strive to maintain consistency between decisions as their authority is based on this consistency. Under these conditions, a judge might take the approach of trying to find precedent cases that match the current case in terms of facts even if not in terms of logic. Case law dealing with such Articles would therefore be more likely to follow Goodhart's view.

To support or disprove this hypothesis would require an in-depth legal analysis far beyond the scope of this paper; one would need to robustly argue why judges find it relatively more difficult to develop legal reasoning for certain articles. However, looking at the Articles where our data indicate that Goodhart's view is the one more widely used, it seems to us that they indeed concern legal concepts that are more slippery than others, which we categorised as follows.

6.1.1 Corporal Articles

We can contrast Articles 2 and 4, where judges follow Goodhart's view, to Article 3, for which judges follow Halsbury's view instead, see Table 2. All three Articles are concerned with the fundamental respect of human life, and we therefore consider them together as the *corporal* Articles.

Article 2: Right to Life prohibits the *intentional*

---

[7]We measure significance using the two tailed paired permutation tests with $p < 0.05$ after Benjamini and Hochberg's (1995) correction.

[8]The Convention Section 1 is the first section of the ECHR and elevates some of the Universal Declaration of Human Rights principles into actionable rights of European citizens (Schindler, 1962).

		Goodhart		Halsbury	
Art	$H_\theta(O_k \mid F)$	MI	U	MI	U
2	0.065	0.014	21.97%	0.010	15.27%
3	0.272	0.028	10.15%	0.047	17.23%
4	0.028	0.020	71.26%	0.011	39.27%
5	0.275	0.019	7.05%	0.021	7.53%
6	0.493	0.042	8.50%	0.089	17.95%
7	0.024	-0.003	-12.01%	-0.000	-1.52%
8	0.298	0.063	21.15%	0.084	28.33%
9	0.022	0.005	23.14%	-0.003	-15.74%
10	0.173	0.003	1.92%	0.034	19.90%
11	0.074	0.018	24.29%	-0.004	-5.66%
12	0.006	-0.001	-11.09%	0.003	46.60%
13	0.235	-0.000	-0.10%	-0.006	-2.38%
14	0.071	-0.005	-7.30%	-0.005	-7.28%
18	0.031	-0.003	-10.00%	-0.007	-24.01%

Table 2: The cross-entropy H_θ, mutual information MI and uncertainty coefficient U results of each of the core ECHR Articles. We note that these values are empirical estimates, so negative MI results are caused by an approximation error in our models.

deprivation of life, save for circumstances where it is a penalty for a crime, in defence, during an arrest, or riot suppression. In the context of the criminal code of Europe, this is a very restricted prohibition. Every country already encodes these rules. On the other hand, it raises the difficult issues of beginning and end of life. Is Article 2 for or against abortion (Cosentino, 2015)? What is its stance on euthanasia (Hendriks, 2019)? Developing a legal test for Article 2 seems very hard indeed.

Similarly, Article 4: *Prohibition of slavery and forced labour*, excludes work forced in detention, compulsory military service, any service during emergency or "normal" civic obligations. Due to the large number of exceptions to the general rule it seems very hard to establish what exactly this Article does prohibit.

Let us compare these to Article 3: *Prohibition of torture*, where Halsbury's view prevails. This Article simply states that *no one shall be subjected to torture or inhuman or degrading treatment or punishment*. No exceptions are given. It seems much easier to develop a legal test for Article 3 than for Articles 2 and 4. The judges are free to establish what constitutes torture; whereas when it comes to Articles 2 and 4, they are facing many restrictions—both legal and political.

6.1.2 Faith and Family Articles

Above, we compare Articles concerned with corporal matters. In a similar way we can also group Articles 8, 9, 10, 11 and 12 as the Articles broadly concerning belief, family and religion.

The two outliers here are Articles 9 and 11. Article 9 provides the *freedom of thought, conscience and religion*, Article 11 provides the *freedom of assembly and association*. For both Articles, Goodhart's test outperforms Halsbury's.

Just like above, the nature of Articles 9 and 11 seems more complicated compared to Article 8, which is similar, but narrower in scope: *Right to respect for private and family life*, Article 10: *Freedom of expression* and Article 12: *Right to marry*.

We would argue that since Articles 8 and 12 provide a *right* as opposed to a *freedom*, they define more narrowly the obligation on the part of the State. Compared to the *freedom of thought* and *association* (Articles 9 and 11), the *right to marry* and the right to *privacy* (Article 8 and 12) seem to be more concrete and testable obligations.

We can further view Article 10: *freedom of expression*, as dealing with an action brought about by the exercise of Article 9: *freedom of thought*. While similar in concept, regulating speech seems far easier in practice than regulating thought.

Finally, an inspection of the ECHR guidelines to Article 11 reveals that judges seem to be often torn between Articles 10 and 11.[9] This is because much of the cases dealing with Article 11 concern themselves with disentangling what constitutes an expression during an assembly and conversely which assembly is a form of an expression. Many cases deal with the question of religious gathering as an assembly. This is obviously not an easy position for a judge to divine a legal test for, and perhaps a good reason for turning to the facts of the precedent cases for consistency instead.

6.2 Late Precedent

There is a group of Articles in the last quarter of Figure 3 (13, 14, 18) for which neither Goodhart's nor Halsbury's view seem to hold. We speculate that the reason for this is that these Articles never appear alone, and instead always appear in conjunction with another Article, and also that they appear late in the list of precedents, so get truncated with our methodology.

Articles 13: *Right to an effective remedy*, 14: *Prohibition of discrimination* and 18: *Limitation on use of restrictions on rights*, are designed to

---

[9]Article 11 guidance: `https://www.echr.coe.int/Documents/Guide_Art_11_ENG.pdf`.

ensure that states provide remedy for their wrongdoing, equal access to the rights, and do not use the restrictions in Articles for Human Rights abuse.

To claim any one of these Articles, the claimant will also have to claim a violation of one of the primary Articles as their core grievance for which they seek the remedy or equal treatment, for instance Article 3: *Prohibition of torture*. This means that any case dealing with Articles 13, 14 and 18 is likely to focus on the violation of that primary right.

While there might be a precedent present for the secondary Articles, the probability is high that our models will not have the chance to train on them because they appear late and because our method truncates text due to computational complexity reasons. This could explain why for these Articles, all our models trained on the precedent cases underperform when compared to the models trained on the facts of the case alone.

6.3   Model Limitations

Another possible explanation for the different behaviour between Articles could lie within the limitations of the neural architecture. There could be a model bias for facts in precedent since they are more similar to the facts at hand as opposed to the arguments. If this is the case our results understate the value of arguments. While this is a concern, the overall results of our paper would not change even if we could remove this bias since we find arguments more important than facts despite this potential handicap.

On a more nuanced level, Articles 2 and 4 above might require a higher level of reasoning than their Article 3 counterpart. So while the judges might have developed a satisfying legal test for them, our models simply aren't able to learn it. For example for Article 7: *No punishment without law*, our precedent models fail to learn any additional information from the precedent facts or arguments.

This might simply be the result of an insufficient representation of Article 7 in training cases, or of its appearance truncated out of the input. However it also raises the question of what a TRANS-FORMER model can learn.

The nascent field of *BERTology* has explored exactly this question (Rogers et al., 2020; Pimentel et al., 2020). In particular the work of Niven and Kao (2019), examining BERT performance on the English Argument Reasoning Comprehension Task (Habernal et al., 2018), suggest that instead of BERT being able to reason, it is merely very good at utilising the artefacts in the data when compared to previous approaches. As Bender and Koller (2020) contend a system can't ever learn meaning from form alone. According to their view, description of the case facts alone will never fully capture the reality of the world the claimant inhabits.

On the other hand, there is some evidence towards transformers being able to reason over simple sentences Clark et al. (2020). While this is encouraging, legal documents are far more complicated than the simple sentences considered in the study above. Either way, the models' ability to reason in the way a human lawyer would is certainly limited and could explain the diminished performance for the more complicated Articles.

7   Related work

In this section, we contextualise our work with relation to the related research on legal AI. Computational approaches to solving legal problems go back at least as far as the late 1950's (Kort, 1957; Nagel, 1963). Early research has focused on crafting rule-based systems for case outcome prediction, achieving human-like performance by the early 2000's (Ashley, 2017). These systems however proved too brittle to keep up with the ever-changing legal landscape and never transitioned from research into industry.

More recently, a new wave of deep learning methods has reinvigorated the research interest in legal AI. The majority of this new work has been conducted on statutory legal systems which do not rely on the doctrine of precedent to nearly the same extent as their common law counterparts. For instance, in Chinese law the use of neural models for case outcome classification has already been investigated extensively (Hu et al., 2018; Zhong et al., 2018; Xu et al., 2020). In the precedential legal domain, smaller corpora of annotated cases have been investigated over the years (Grover et al., 2003; Valvoda et al., 2018). However, large-scale corpora necessary for deep learning architectures have become available only recently. The Caselaw Access Project[10] introduced a large dataset of American case law in 2018. Aletras et al. (2016) have introduced the ECtHR corpus, and Chalkidis et al. (2019) have run deep neural networks on it in order to predict outcome. Similarly, the Canadian Supreme Court Case corpus has been used in infor-

---

[10]Caselaw Access Project:, `https://case.law`

mation retrieval for the first time by Rabelo et al. (2020). This improved access to a high quality common law datasets has opened up a potential for new work in the field of legal AI.

Particularly similar to our work is the study done by Sim et al. (2016), who have considered the influence of petitioners and responders (amicus) briefs on the US Supreme Court decision and opinions.

8 Conclusion

In this paper, we have shifted the focus of legal AI research from practical tasks such as precedent retrieval or outcome prediction, to a theoretical question: which aspect of the precedent is most important in forming the law? To this end, we trained a similar neural modeling approach as Chalkidis et al. (2019) to predict the outcome of a case on the ECtHR dataset, and inspected the difference in the mutual information between our operationalisations of Halsbury's and Goodhart's view. We have used a method inspired by Pimentel et al. (2019) to approximate the MI. We observe that out of the two archetypal views on precedent, that of Halsbury and Goodhart, the former has a better empirical support in the domain of ECtHR case law.

This study has demonstrated a novel method of approaching jurisprudential questions using the information-theoretic toolkit. We hope that future work can leverage our methodology towards answering other questions of legal philosophy. However, our results are not only of an interest in the context of legal theory, but they can also inform a development of better legal models in practice. Since most precedential reasoning is conducted using the arguments in the precedent, outcome prediction models should take advantage of the case arguments, instead of relying solely on the facts.

Ethical Considerations

While our work is not concerned with a legal application, it is important to note that the results presented here are qualified by the limitations of contemporary NLP models' ability to process language. It should therefore serve as no indication that judges could (or should) be replaced by models or techniques discussed in this paper.

Acknowledgements

References

Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preoţiuc-Pietro, and Vasileios Lampos. 2016. Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective. *PeerJ Computer Science*, 2:e93.

Larry Alexander and Emily Sherwin. 2008. *The Mystification of Common-Law Reasoning*, Cambridge Introductions to Philosophy and Law, page 64–103. Cambridge University Press.

Kevin D. Ashley. 2017. *Artificial Intelligence and Legal Analytics*. Cambridge University Press, Cambridge.

Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020. Longformer: The Long-Document Transformer. *arXiv*.

Emily M. Bender and Alexander Koller. 2020. Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5185–5198, Online. Association for Computational Linguistics.

Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300.

Henry Black. 2019. *Black's Law Dictionary*, 11th edition. Thomson Reuters.

Ilias Chalkidis, Ion Androutsopoulos, and Nikolaos Aletras. 2019. Neural legal judgment prediction in English. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4317–4323, Florence, Italy. Association for Computational Linguistics.

Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. 2020. LEGAL-BERT: The muppets straight out of law school. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2898–2904, Online. Association for Computational Linguistics.

Peter Clark, Oyvind Tafjord, and Kyle Richardson. 2020. Transformers as Soft Reasoners over Language. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 3882–3890. International Joint Conferences on Artificial Intelligence Organization. Main track.

Chiara Cosentino. 2015. Safe and legal abortion: An emerging human right? the long-lasting dispute with state sovereignty in ECHR jurisprudence. *Human Rights Law Review*, 15(3):569–589.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Pintip Hompluem Dunn. 2003. How judges overrule: Speech act theory and the doctrine of stare decisis. *The Yale Law Journal*, 113(2):493–531.

Neil Duxbury. 2008. *Distinguishing, overruling and the problem of self-reference*, page 111–149. Cambridge University Press.

ECtHR. 2014. European court of human rights: Questions & answers for lawyers.

Arthur L. Goodhart. 1930. Determining the ratio decidendi of a case. *The Yale Law Journal*, 40(2):161–183.

Claire Grover, Ben Hachey, Ian Hughson, and Chris Korycinski. 2003. Automatic summarisation of legal documents. In *Proceedings of the 9th International Conference on Artificial Intelligence and Law*, ICAIL '03, page 243–251, New York, NY, USA. Association for Computing Machinery.

Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, and Benno Stein. 2018. The argument reasoning comprehension task: Identification and reconstruction of implicit warrants. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1930–1940, New Orleans, Louisiana. Association for Computational Linguistics.

Lord Halsbury. 1907. *Halsbury's Laws of England*, 1 edition. LexisNexis Butterworths.

Arend Cornelis Hendriks. 2019. End-of-life decisions. Recent jurisprudence of the European Court of Human Rights. *ERA Forum*, 19(4):561–570.

Zikun Hu, Xiang Li, Cunchao Tu, Zhiyuan Liu, and Maosong Sun. 2018. Few-shot charge prediction with discriminative legal attributes. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 487–498, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Marc Jacob. 2014. *Precedents and Case-Based Reasoning in the European Court of Justice: Unfinished Business*. Cambridge University Press.

Matti Joutsen. 2019. *International and Transnational Crime and Justice*, 2 edition. Cambridge University Press.

Johannes Kiesel, Maria Mestre, Rishabh Shukla, Emmanuel Vincent, Payam Adineh, David Corney, Benno Stein, and Martin Potthast. 2019. SemEval-2019 task 4: Hyperpartisan news detection. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 829–839, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

Fred Kort. 1957. Predicting supreme court decisions mathematically: A quantitative analysis of the "right to counsel" cases. *American Political Science Review*, 51(1):1–12.

Grant Lamond. 2005. Do precedents create rules? *Legal Theory*, 11(1):1–26.

Grant Lamond. 2016. Precedent and Analogy in Legal Reasoning. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*, spring 2016 edition. Metaphysics Research Lab, Stanford University.

Yonatan Lupu and Erik Voeten. 2010. The role of precedent at the european court of human rights: A network analysis of case citations.

Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. 2011. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 142–150, Portland, Oregon, USA. Association for Computational Linguistics.

David McAllester and Karl Stratos. 2020. Formal limitations on the measurement of mutual information. In *International Conference on Artificial Intelligence and Statistics*, pages 875–884.

Stuart S. Nagel. 1963. Applying correlation analysis to case prediction. *Texas Law Review*, 42:1006.

Timothy Niven and Hung-Yu Kao. 2019. Probing neural network comprehension of natural language arguments. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4658–4664, Florence, Italy. Association for Computational Linguistics.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, pages 8026–8037.

Tiago Pimentel, Arya D. McCarthy, Damian Blasi, Brian Roark, and Ryan Cotterell. 2019. Meaning to form: Measuring systematicity as information. In

Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy. Association for Computational Linguistics.

Tiago Pimentel, Brian Roark, Søren Wichmann, Ryan Cotterell, and Damian Blasi. 2021. Finding concept-specific biases in form–meaning associations. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Virtual. Association for Computational Linguistics.

Tiago Pimentel, Josef Valvoda, Rowan Hall Maudslay, Ran Zmigrod, Adina Williams, and Ryan Cotterell. 2020. Information-theoretic probing for linguistic structure. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4609–4622, Online. Association for Computational Linguistics.

Juliano Rabelo, Mi-Young Kim, Randy Goebel, Masaharu Yoshioka, Yoshinobu Kano, and Ken Satoh. 2020. A summary of the COLIEE 2019 competition. In *New Frontiers in Artificial Intelligence*, pages 34–49, Cham. Springer International Publishing.

Anna Rogers, Olga Kovaleva, and Anna Rumshisky. 2020. A primer in BERTology: What we know about how BERT works. *Transactions of the Association for Computational Linguistics*, 8:842–866.

Dietrich Schindler. 1962. The European convention on human rights in practice. *Washington University Law Review*, page 152.

C. E. Shannon and W. Weaver. 1962. *The Mathematical Theory of Communication*. University of Illinois Press.

Yanchuan Sim, Bryan Routledge, and Noah A. Smith. 2016. Friends with motives: Using text to infer influence on SCOTUS. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1724–1733, Austin, Texas. Association for Computational Linguistics.

James F. Spriggs and Thomas G. Hansford. 2001. Explaining the overruling of u.s. supreme court precedent. *The Journal of Politics*, 63(4):1091–1111.

Henri Theil. 1970. On the estimation of relationships involving qualitative variables. *American Journal of Sociology*, 76(1):103–154.

Josef Valvoda, Oliver Ray, and Ken Satoh. 2018. Using agreement statements to identify majority opinion in UKHL case law. In *Legal Knowledge and Information Systems - JURIX 2018: The Thirty-first Annual Conference, Groningen, The Netherlands, 12-14 December 2018*, volume 313 of *Frontiers in Artificial Intelligence and Applications*, pages 141–150. IOS Press.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008.

Adina Williams, Tiago Pimentel, Hagen Blix, Arya D. McCarthy, Eleanor Chodroff, and Ryan Cotterell. 2020. Predicting declension class from form and meaning. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6682–6695, Online. Association for Computational Linguistics.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Nuo Xu, Pinghui Wang, Long Chen, Li Pan, Xiaoyan Wang, and Junzhou Zhao. 2020. Distinguish confusing law articles for legal judgment prediction. *arXiv preprint arXiv:2004.02557*.

Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, and Maosong Sun. 2018. Legal judgment prediction via topological learning. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3540–3549, Brussels, Belgium. Association for Computational Linguistics.

Bostjan Zupancic. 2016. In the Context of the Common Law: The European Court of Human Rights in Strasbourg Transcript.

A   Glossary:

Legal Terms	
Facts	The description of what had happened to the claimant. This includes more general description of who they are, circumstances of the perceived violation of their rights and the proceedings in domestic courts before their appeal to ECtHR.
Arguments	The judges explanation of why did they decide the case the way they did. This includes citations of previous cases, application of any relevant legal test, development of a new legal test, analysis of the facts etc.
Precedent	Cases that have been cited by the judges as part of their arguments.
Ratio Decidendi	The reasons for the decision in a case that is binding on the subsequent cases. Also known as the ratio. What exactly is ratio is contested by legal scholars.
Obiter Dicta	The non-binding discussions in the case. Whatever is not ratio.
Binding	Judges are expected to adhere to the binding rules of law and decide future access accordingly.
Stare Decisis	New cases with the same facts to the already decided case should lead to the same outcome. This is the doctrine of precedent by which judges can create law.
Caselaw	Transcripts of the court proceedings.
ECHR	European Convention of Human Rights, comprises of the Convention and the Protocols to the convention. The Protocols are the additions and amendments to the Convention introduced after the signing of the original Convention.
ECtHR	European Court of Human Rights, adjudicates ECHR cases.

Selected ECHR Articles

Article 2: *Right to life*	Everyone's right to life shall be protected by law. No one shall be deprived of his life intentionally save in the execution of a sentence of a court following his conviction of a crime for which this penalty is provided by law.
Article 3: *Prohibition of torture*	No one shall be subjected to torture or to inhuman or degrading treatment or punishment.
Article 4: *Prohibition of slavery and forced labour*	No one shall be held in slavery or servitude.
Article 8: *Right to respect for private and family life*	Everyone has the right to respect for his private and family life, his home and his correspondence.
Article 9: *Freedom of thought, conscience and religion*	Everyone has the right to freedom of thought, conscience and religion; this right includes freedom to change his religion or belief and freedom, either alone or in community with others and in public or private, to manifest his religion or belief, in worship, teaching, practice and observance.
Article 10: *Freedom of expression*	Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. This Article shall not prevent States from requiring the licensing of broadcasting, television or cinema enterprises.
Article 11: *Freedom of assembly and association*	Everyone has the right to freedom of peaceful assembly and to freedom of association with others, including the right to form and to join trade unions for the protection of his interests.
Article 12: *Right to marry*	Men and women of marriageable age have the right to marry and to found a family, according to the national laws governing the exercise of this right.
Article 13: *Right to an effective remedy*	Everyone whose rights and freedoms as set forth in this Convention are violated shall have an effective remedy before a national authority notwithstanding that the violation has been committed by persons acting in an official capacity.
Article 14: *Prohibition of discrimination*	The enjoyment of the rights and freedoms set forth in this Convention shall be secured without discrimination on any ground such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status.
Article 18: *Limitation on use of restrictions on rights*	The restrictions permitted under this Convention to the said rights and freedoms shall not be applied for any purpose other than those for which they have been prescribed.

B Facts & Arguments Examples

Fact	*The applicants, D.P. and J.C., who are sister and brother, are United Kingdom nationals, born in 1964 and 1967 and living in London and Nottingham, respectively...*
Argument	*Article 2 of the Convention provides, in its first sentence: "1. Everyone's right to life shall be protected by law. ..." 46. The applicants complain that the authorities failed to protect the life of their son and were responsible for his death...*