A Theory of Judicial Torture

Kong-Ping Chen\textsuperscript{a}  Tsung-Sheng Tsai\textsuperscript{b}  Angela Leung\textsuperscript{c}

January 25, 2006

Abstract

Judicial torture to elicit information or extract confession was a common practice in pre-modern societies, both in the East and the West. Moreover, often it was applied not only on the suspects, but also on the witnesses and plaintiffs as well. This paper proposes a theory for judicial torture. It is shown that if the judge aims to balance type I and type II errors in decision-making, then torture can improve social welfare by forcing the guilty to confess with higher probability than the innocent, and thereby decreases type I error, although at the cost of type II error. In that case torturing the witnesses might also be welfare-improving, as it helps to screen the cases so that only those with greater merits enter the court. When the information revealed during investigation improved as a result of technological advance, a judicial system based on torture became inferior to one based on evidence. This explains the historical development of the judicial system in Europe.

Keywords: Torture; Type I and Type II errors; Evidence.

\textsuperscript{a}Corresponding author. Research Center for Humanities and Social Sciences, Academia Sinica, Taipei, 11529, Taiwan. Email: kongpin@gate.sinica.edu.tw. Tel: 886-2-2789-8160. Fax: 886-2-2789-4160.

\textsuperscript{b}Institute of Economics, Academia Sinica. Email: tstsai@econ.sinica.edu.tw

\textsuperscript{c}Institute of History and Philology, Academia Sinica. Email: tpleung@ccvax.sinica.edu.tw
1 Introduction

Judicial torture for the purpose of eliciting information was a common practice in pre-modern societies. In the West, it emerged in Greek law and continued in Roman law. Its history can then be traced through the Middle Ages, down to the legal reforms of the eighteenth century and the abolition of torture in criminal legal procedure in the nineteenth century in most parts of Europe. (Peters 1985, p.5.) In China, judicial torture formed an important part of the imperial legal codes from the first Empire (Qin dynasty, 221-207 BC) to the last Qing dynasty (1644-1911). (Shen 1985.) Various forms of judicial torture were also widely implemented in Moslem, African and various Asian societies before our time. (Lea 1971, 203-208.)

It is indeed hard to find any pre-modern society that did not rely on torture for the gathering of evidence in judicial proceedings.

The modern man, for whom judicial torture is not only immoral but also irrational, may imagine that it was inflicted exclusively on the accused to obtain his confession. Nothing is further from the truth. Judicial torture in pre-modern times was applied not only to the suspect, but also to the witness, and even to the plaintiff. Witnesses were tortured already in Greek and Roman cases, and continued to be so in the medieval period and until the late 18th century. (Peters 1985, p.18, p.69.) Even the accuser or plaintiff would undergo an ordeal to substantiate his veracity or be tortured when he was unable to make good his accusation. In Roman law, the accuser could be exposed to the *lex talion* (law of retaliation) in case he failed to prove the justice of the charge. (Lea 1971, p.333.) In primitive Russian laws, the accuser was obliged to undergo the ordeal of a red-hot iron if he could not substantiate his case with witnesses. Archbishop Hincmar of Rheims of the 9th century required that cases of complaint against priests be supported by seven witnesses, of whom one must be tortured to prove the truth of his companions’ oath, as a wholesome
check upon perjury and subornation. (Lea 1971, p.290.)

The same principles could be observed in China. The legal code of the Qing dynasty clearly allowed torture to be used not only on the accused, but also on “secondary suspects and to witnesses as well as to principals.” (Bodde and Morris 1967, p.97.) In fact, the tradition could be traced back to the Tang code of 653 AD that constituted the basis for all subsequent imperial law codes. According to Tang law, if the accused insisted on his innocence even after having suffered the maximum amount of torture allowed by the law, the plaintiff would in turn be tortured. (Shen 1985, p.511.) Moreover, since at least the Tang dynasty, a plaintiff who bypassed the immediate authority and presented his complaints to higher administrative levels was required to undergo torture before the examination of the case (Xue 1998, pp.636-639), again as a check upon perjury and subornation, or as a measure to discourage such actions. During the early years of the Ming dynasty (1368-1644), frequent complaints brought directly to the capital prompted the government to implement an extremely severe punishment - banishment to the frontier - on plaintiffs, even at times on those who could substantiate their cases, to control the number such acts. (Shen 1971, p.1142.) In brief, judicial torture in pre-modern societies was legitimately applied almost universally to suspects, witnesses, and plaintiffs for various reasons before the mid-nineteenth century. Moreover, it was an institutionalized practice, rather than a manifestation of judge’s abuse of power.

Starting from around mid-18th century, the practice of judicial torture in the West was gradually replaced by a system which was based on evidence. Its abolition was a long and gradual process that lasted from the mid-18th to the early 19th centuries. The conventional historical explanation of this abolition movement relies heavily on the influence of the humanists of the Enlightenment. However, experts
on jurisprudence usually find this explanation too loose and prefer to explain the phenomenon by changes in the judicial system itself. According to Langbein, the abolition was largely due to the emergence of the new law of proof. During this period, the courts gradually imposed the new and less rigorous punishments according to a less strict standard of proof, one of persuasion rather than certainty. Since certainty was no longer the only requirement to put an accused in prison, a lesser punishment, torture became unnecessary. “Only when confession evidence was no longer necessary to convict the guilty could European law escape its centuries of dependence on judicial torture.” (Langbein 1983, pp.1555-1556.) This explanation, however, is incomplete because, as noted by Damaška (1978) and Silverman (2001), torture continued to be used well after the change in law of proof.

An indirect but equally important factor in the process of the abolition of judicial torture in the West should be the systematic application of scientific knowledge in criminal investigation beginning in the early 19th century. Experts in criminalistics agree that as early as in the 1820s, the basis was laid for the introduction of empirical science into criminal justice and for the redefinition of criminalistics. The 19th century witnessed important breakthroughs in the application of scientific knowledge in criminal investigations including the use of precise measurements of the human body structure (anthropometry) and fingerprinting. The application of scientific methods in criminal investigations had obviously reached a mature stage in the late 19th century when Hans Gross published the first classic in the specialty: Criminal Investigation in 1883. With such technological breakthroughs, past human activity can be specified without relying exclusively on the confessions of the guilty or witnesses, often obtained by torture in earlier times. (Parker 1983, 432-433; Fullmer 1980, p.27.) Indeed, it is the purpose of this paper to show that it is exactly this technological advancement and its implications on information revealed
in investigation that rendered torture obsolete.

We propose a model for judicial torture based on the theory of informational economics. The model serves two purposes. First, it explains why under a certain environment a system based on torture becomes the dominant judicial system. There are two consequences of torture. On the one hand, since some innocents are tortured to confession, torture increases the chance that an innocent suspect is wrongfully convicted (i.e., type I error). On the other hand, if investigation is of informational value in the sense that guilty suspects are more likely to be found to be so, then (if the judge uses torture as a threat) the expected cost to deny a crime will be greater for the guilty than the innocent, since he will face higher probability of being tortured after investigation.\(^1\) This reduces the chance that a guilty suspect is wrongfully released (i.e., type II error). It is shown that if little information is revealed during investigation, then the gain from reduction in type II error (by torture) will outweigh the cost of increase in type I error. In other words, if the goal of the judge is to balance the costs of type I and type II errors, torture increases social welfare. Moreover, the judge can sometimes further improve social welfare by torturing the witness or plaintiff as well. The reason is that if the witness or the plaintiff is also tortured, then only those witnesses or plaintiffs who are more sure of the suspects’ crime will come to the court. Thus the average quality of the cases entering the court will increase. This implies a lower type I error when the judge also tortures the suspect in the legal proceedings. If this reduction in type I error is sufficiently large, then torturing the witness or suspect is indeed welfare-improving.

The second purpose of our theoretical model is to apply our results to explain the historical development of the judicial system. Since the advantage of torture is weakened when the information revealed during investigation improves, a system

\(^1\) Note that without torture, the legal penalty of confession and being proven guilty are the same. The guilty suspect will then not have greater incentive to confess than the innocent.
based solely on evidence will overtake torture as a better system when technological progress reached a certain threshold. This occurred in the 19th century in the West when important breakthroughs occurred in applying scientific knowledge to criminal investigations such as blood type tests and fingerprint identification.

The underlying principle behind our model is similar to that of the war of attrition, a phrase introduced in theoretical biology by Maynard Smith (1974) to explain animals’ behavior in fights for prey, and was later used to explain the exit behavior of firms by economists.\(^2\) In our context the judge imposes a war of attrition between the innocent and the guilty. The former has greater strength in that he is more likely to survive an investigation, and can thus afford to deny the crime (win out the war of attrition) with higher probability. Another strand of research, also related to our model, argues that a strategy of a firm that raises the costs of all competing firms in the market (e.g., advertising wars, adopting incompatible technologies, lobbying for legislations that raise labor costs, etc.) can sometimes strategically increase the profit of the firm which has a certain cost advantage.\(^3\) In our model, imposing a cost (torture) to all suspects sometimes can lend a certain advantage to the innocent. In a broader context, our model is also related to the signaling game literature (e.g., Spence 1973), in which an economic agent (in our model, the suspect) with private information of his own characteristic (in our model, whether he is guilty) must signal himself through a costly action (in our model, enduring torture).

The paper that is most relevant to ours is Wantchekon and Healy (1999). Similar to our paper, they also view torture as a device to elicit information under uncertainty. In their model, there is uncertainty both about the types of torturer and the victim. They mainly try to argue that even when the torturer cares about the pains

\(^2\) See, for example, Ghemawat and Nalebuff (1985) and Whinston (1988).

\(^3\) See for example, Salop (1979) and Salop and Scheffman (1983, 1987).
suffered by the victims, there is still the possibility that the (informed) victim will reveal the information, out of fear of facing a sadistic torturer. Moreover, they also argue that when torture is used by the state as a way to intimidate, it has a tendency to become more widespread and more cruel. In our model, the torturer aims to maximize a measure of social welfare. We also compare the relative advantage of systems based on evidence and torture, and use this comparison to explain the historical development of judicial proceedings.\textsuperscript{4}

2 The Model

A crime has been committed and a suspect is brought to the court. He is either guilty (\(\theta = G\)) or innocent (\(\theta = I\)), which is his private information. The judge needs to make a verdict on whether he is guilty. The prior belief of the judge that the suspect is guilty is \(q\), with \(0 < q < 1\). That is, \(\text{Prob}(\theta = G) = q\).

We consider two types of judicial system. In the first system, which we call the evidence-based system, the judge conducts an investigation, which can improve the precision of his information but not resolve the uncertainty he faces. The precision of investigation is modeled in the following way: The judge draws a signal (evidence) from the set \(\{g, i\}\). If the suspect is indeed guilty, then with probability \(q_G\) (resp. \(1 - q_G\)) the judge will draw \(g\) (resp. \(i\)). If the suspect is innocent, then with probability \(q_I\) (resp. \(1 - q_I\)) the judge draws \(g\) (resp. \(i\)). Assume that \(1 > q_G > q_I > 0\), i.e., investigation is informative in the sense that the judge is more likely to draw a \(g\) when he faces a guilty suspect than an innocent. The judge then decides whether to convict the suspect based on the evidence collected. If a suspect is convicted,

\textsuperscript{4} Other papers that utilize game-theoretical models to investigate how rules of proceeding in criminal cases affect information transmission are Seidmann (2005), Seidmann and Stein (2000), and Sanchirico (2000, 2001). In particular, Sanchirico (2000), as our paper, also tries to explain a certain aspect of legal history with a theoretical framework.
he is subject to a legal penalty causing a disutility $P$. Otherwise he is released, in which case his utility is 0.

Since investigation is imperfect, there are two possible errors that the judge can commit. First, he might wrongfully convict a suspect who is actually innocent. We call this a type I error. Second, he might wrongfully release a suspect who is actually guilty. We call this a type II error. Assume that type I error causes a loss of social welfare by $L_1$, and type II error $L_2$. The objective of the judge is to minimize the sum of the two types of cost by choosing whether to convict the suspect after he draws the evidence. This is a simple optimization problem for the judge.

In the second type of judicial system, which we call the torture-based system, confession from the suspect is required for conviction. For that purpose the judge is allowed to torture the suspect. We model this as follows. The suspect is given an option of whether to confess. If he does, he is subject to a legal penalty of the crime which causes a disutility of $P$. If not, the judge will go through the same investigation as in the evidence-based system. Based on the evidence drawn, he decides whether to torture the suspect. If he does not, then the suspect is released with a utility 0. If he tortures, let $T$ denote the disutility of being tortured for the suspect. Thus if the suspect is tortured to confess, his total disutility will be $T + P$. If the suspect does not confess, the judge can again decide whether to torture or to release, and the procedure repeats *ad infinitum*. This setup requires some explanation. First, the optimal strategy of a suspect, if he intends to confess, is actually to deny his crime initially, and to confess immediately at the moment the judge decides to torture. Our specification thus assumes that the judge can make a credible commitment in that if he decides to torture, then he commits to torture by an amount $T$, regardless of whether the suspect confesses during (or before) torture. Second, if the suspect has not confessed after torture, then (since past torture is a
sunk cost which is irrelevant to current decision, in the spirit of subgame perfection) the judge needs to keep torturing the suspect until the latter confesses,\textsuperscript{5} otherwise previous torture is a pure social loss, and is never an optimum for the judge. In this case the optimal strategy of the suspect after he has been tortured once (by an amount $T$) is to confess immediately, and there will never be a second round of torture. This result greatly simplifies the model, as a suspect who does not confess only faces two possible consequences: either he is released, or he is tortured exactly once and confesses.\textsuperscript{6}

Assume that when a level of torture $T$ is imposed on the suspect, it also causes a social loss of $kT$, $k < 1$. The judge is assumed to be benevolent in that his aim is to minimize both the sum of expected social loss caused by the two types of error and, if he tortures, the social loss due to torture. Since the judge’s decision will affect the suspect’s confession decision and vice versa, this is a game played between the judge and the suspect. The torture-based system is substantially more complicated than the evidence-based system, as it involves strategic interaction of the suspect and the judge.

To summarize, the events under the evidence-based system proceed in the following order: (1) Nature determines whether the suspect is guilty (i.e., whether $\theta = G$ or $\theta = I$). (2) The judge draws an evidence. (3) Based on evidence, the judge decides whether to release the suspect (and the suspect receives a utility 0), or to convict him (and the utility of the suspect is $P$). The events under the torture-based system proceed in the following order: (1) Nature determines whether the suspect

\textsuperscript{5} Technically, this means that the judge’s belief of the probability that the suspect is guilty does not change during or after torture.

\textsuperscript{6} Note that since a suspect who confesses before investigation suffers a disutility $P$ while one who confesses after being tortured suffers $P + T$, the specification is equivalent to the real world plea bargaining, in which the suspect’s penalty is reduced (by an amount of $T$ in our specification) if he chooses to confess before investigation.
is guilty. (2) The suspect decides whether to confess. If he confesses, he is subject
to a legal penalty of $P$. If he does not, then (3) the judge conducts an investigation
by drawing an evidence. Based on the evidence, he decides whether to torture the
suspect. If not, the suspect is released. (4) If tortured, then by our argument earlier,
the suspect confesses (and is penalized by $P$, in addition to the disutility of torture,
$T$).

Section 3 derives the optimal decision of the judge for an evidence-based system.
Section 4 derives the equilibrium under the torture-based system.

## 3 The Evidence-Based System

Under the evidence-based system, the judge draws an evidence, and decides whether
to convict the suspect solely based on the posterior derived from the evidence. If he
draw an $i$, his posterior that the suspect is guilty will be

$$ q'_I \equiv \frac{q(1-q_G)}{q(1-q_G) + (1-q)(1-q_I)}. $$

Similarly, if he draws $g$, the posterior is

$$ q'_G \equiv \frac{qq_G}{qq_G + (1-q)q_I}. $$

Thus if the evidence collected is $\theta$ ($\theta = I, G$), then the expected social loss will be
$(1-q'_\theta)L_1$ if the judge convicts the suspect, and is $q'_\theta L_2$ if he releases the suspect.

Whether he should convict the suspect depends on the evidence he draws and the
relative size of $(1-q'_\theta)L_1$ and $q'_\theta L_2$. To make our discussion non-trivial, we assume
that

**A1.** $\frac{qq_G}{q_I} L_2 > \frac{1-q}{q} L_1 > \frac{1-q_G}{1-q_I} L_2.$

If A1 does not hold, then regardless of evidence drawn, either he never convicts
(if the second inequality fails to hold) or he never releases a suspect (if the first
inequality fails). Investigation thus does not have any function. If A1 holds, then it can be easily checked that $q'_G L_2 > (1 - q'_G) L_1$ and $q'_I L_2 < (1 - q'_I) L_2$. That is, the judge will convict (release) the suspect if he draws $g(i)$. We thus have the following lemma.

**Lemma 1.** Assuming A1, then under an evidence-based system, the judge convicts the suspect if and only if he draws $g$.

### 4 The Game of Torture

We first consider the case without investigation. Since the prior belief of the judge that the suspect is guilty is $q$, the probability of making a type I error is $1 - q$ if the suspect is convicted, and the probability of making a type II error is $q$ if the suspect is released. Thus without an investigation, the expected social loss when a suspect is tortured to confess is $(1 - q) L_1 + kT$, and the expected social loss when he is released is $q L_2$. We make the following assumption:

**A2.** $\frac{1-q}{q} L_1 + \frac{kT}{q} > L_2 > kT$.

The first inequality means that under the prior belief $q$, the expected social loss of torture is greater than that of releasing the suspect. This implies that the judge will not torture without an investigation. This is a reasonable assumption to make. Without this assumption there will be no interesting trade-off under torture: The judge will always torture the suspect, and the latter will then always confess. The second inequality means that the social loss of type II error is greater than the loss of torture. If this inequality is violated, there will never be torture, as the judge will

---

7 In Section 5 of the paper, we will investigate how the plaintiff or witness helps to screen the cases by deciding whether to bring the cases to court. As a result, the value of $q$ can influenced by the behavior of the plaintiff.

8 If $\frac{kT}{q} > \frac{qG - qI}{1 - qI}$, then A1 implies the first inequality of A2.
not torture even if he believes the suspect is guilty with probability 1. Naturally, in this case no suspect will confess.

4.1 The Optimal Decision of the Judge

Suppose now the judge considers whether to torture after he investigates. Since the suspect’s type is his private information, his decision will be based on whether he is innocent or guilty, and this decision will in turn play a role in determining the posterior of the judge. Suppose the innocent suspect denies the crime with probability \( \nu_I \), and the guilty denies the crime with probability \( \nu_G \). By the rule of the torture-based system, if the suspect does not confess, the judge will conduct an investigation to draw an evidence. Given the strategy of the suspect and the evidence drawn by the judge, the posterior of the judge that the suspect is guilty is

\[
\hat{q}(\theta = G|\nu_G, \nu_I, g) = \frac{q\nu_G q_G}{q\nu_G q_G + (1-q)\nu_I q_I} \equiv \hat{q}_G \text{ if he draws } g.
\]

On the other hand, the posterior, when the judge draws \( i \), is

\[
\hat{q}(\theta = G|\nu_G, \nu_I, i) = \frac{q\nu_G (1-q_G)}{q\nu_G (1-q_G) + (1-q)\nu_I (1-q_I)} \equiv \hat{q}_I.
\]

It can be easily seen that \( \hat{q}_G > \hat{q}_I \). Based on this posterior, the expected social loss is \((1 - \hat{q}_\theta)L_1 + kT\) if the judge tortures the suspect; \( \theta = I, G \). If he decides not to torture, and release the suspect, the expected social welfare loss is \( \hat{q}_\theta L_2 \). Therefore, the judge tortures the suspect if and only if

\[
(1 - \hat{q}_\theta)L_1 + kT \leq \hat{q}_\theta L_2.
\]
4.2 Optimal Decision of the Suspect

The decision of a suspect of type $\theta$ ($\theta = I, G$) is to choose the value of $\nu_\theta$ to maximize his expected utility, which is

$$-\nu_\theta(q_\theta x_g + (1-q_\theta)x_i)(T+P) - (1-\nu_\theta)P;$$

where $x_g$ and $x_i$ are the probability the judge will torture after he draws $g$ and $i$, respectively.\footnote{Presumably the guilty and the innocent suspects will face different chances of being tortured; that is, both $x_i$ and $x_g$ are functions of $\theta$. This, however, happens only for a separating equilibrium. We will show in Lemma 2 that there does not exist a separating equilibrium. Thus the innocent and the guilty face the same $x_i$ and $x_g$.}

Since $\hat{q}_G > \hat{q}_I$, it is obvious that $x_g \geq x_i$ in equilibrium. The following two lemmas are very useful for solving equilibria later.

**Lemma 2.** There does not exist a separating equilibrium. That is, there does not exist an equilibrium in which $\nu_G = 0$ and $\nu_I = 1$.

**Proof.** Consider a separating equilibrium in which $\nu_I = 1$ and $\nu_G = 0$. Then by Bayes’ rule it is easy to see that the posterior for the judge is $\hat{q}_I = \hat{q}_G = 0$. That is, he believes a denying suspect is innocent, regardless of the evidence drawn. By Assumption A2 he will not torture the denying suspect (i.e., $x_i = x_g = 0$). But if that is the case, then the guilty suspect will surely deviate from $\nu_G = 0$.

Lemma 2 implies that, in any equilibrium, the pool of denying suspects (if there are any) will comprise of both innocent and guilty suspects. An important consequence of this is that both types of denying suspect will face the same posterior of the judge, and will be tortured with the same probability. That is, as we have already mentioned, $x_g$ and $x_i$ are independent of $\theta$. We then have the following useful result.

**Lemma 3.** In any equilibrium, the guilty suspect will confess with a higher probability than the innocent one; i.e., $\nu_I \geq \nu_G$. 
Proof. If a suspect confesses, then he is subject to a legal penalty $P$, regardless of his type. If he denies, then his expected utility is $-\[q_Gx_g + (1 - q_G)x_i\](T + P)$ if he is guilty, and $-\[q_Ix_g + (1 - q_I)x_i\](T + P)$ if innocent. Since $q_G > q_I$ and $x_g \geq x_i$, the consequence of denying is more serious for the guilty suspect. As a result, it must be the case that $\nu_I \geq \nu_G$.

Although simple, Lemma 3 is actually an important result. If, contrary to Lemma 3, $\nu_I = \nu_G$, then torture will not change the composition of the pool of denying suspects. That is, given evidence drawn, the probability that the denying suspect is guilty is exactly the same as the prior. In that case torture will be necessarily inferior to an evidence-based system, as the judge can directly convict the suspect, which results in exactly the same type I and type II errors but saves the loss of pain of torture. The informational advantage brought about by torture can be seen clearly by observing that $\nu_I \geq \nu_G$ implies $\hat{q}_G > q'_G$ and $q'_I > \hat{q}_I$. That is, compared with the evidence-based system, under the torture-based system the judge is more sure that the suspect is guilty when he draws $g$, and is more sure that the suspect is innocent when he draws $i$. Lemma 3 also implies that $\hat{q}_G > q > \hat{q}_I$. That is, the judge’s posterior that suspect is guilty, after he draws $g$ ($i$), is greater (smaller) than the prior. Therefore, since the judge will not torture under the prior belief (by A1), he will not torture when he draws an $i$ either:

**Lemma 4.** If the judge draws $i$, then the suspect is released.

Proof. By A1 we know that $(1 - q)L_1 + kT > qL_2$. Since $q > \hat{q}_I$, by the second inequality of A2, we know that $(1 - \hat{q}_I)L_1 + kT > \hat{q}_IL_2$. That is, under the posterior $\hat{q}_I$, the expected social welfare loss is greater if he tortures. This implies that the judge will release the suspect if he draws an $i$.

By Lemma 4, the judge will torture with positive probability only if he draw a
As a result, $x_i$ is always 0, and we will use $x$ as a shorthand for $x_g$. That is, $x$ is the probability of torture when the judge draws a $g$.

4.3 The Equilibrium

Given the behavior of the suspect derived in the previous subsection, the judge designs the value of $x$ to minimize expected social welfare lost:

$$\min_x W^c(x; q, \nu_G, \nu_I) \equiv x(kT + (1 - \hat{q}_G)L_1) + (1 - x)\hat{q}_GL_2.$$  \hspace{1cm} (2)

The first term in (2) is the expected social welfare loss of torture (pain of torture plus cost of type I error). The second term is the expected social welfare loss of type II error. The equilibrium behavior of the game of torture is characterized in the following proposition.

**Proposition 1.** The equilibrium of the game of torture is as follows.

The judge will release the suspect if he draws $i$. Moreover,

1. If $\frac{q_G}{q_I} > \frac{(1-q)(L_1+kT)}{q(L_2-kT)}$, then
   
   (i) when $\frac{P}{P+T} < q_G$, then $\nu^*_I = 1$, $\nu^*_G = \nu$, and $x^* = \frac{P}{q_G(P+T)}$; 
   
   (ii) when $\frac{P}{P+T} = q_G$, then $\nu^*_I = 1$, $\nu^*_G \in [\nu, 1)$, and $x^* = 1$; 
   
   (iii) when $\frac{P}{P+T} > q_G$, then $\nu^*_I = \nu^*_G = 1$, and $x^* = 1$; 

   where $\nu = \frac{qI(1-q)(L_1+kT)}{q_G(qL_2-kT)}$.

2. If $\frac{q_G}{q_I} = \frac{(1-q)(L_1+kT)}{q(L_2-kT)}$, then
   
   (i) when $\frac{P}{P+T} < q_G$, then $\nu^*_I = \nu^*_G = 1$, and $x^* \in [0, \frac{P}{q_G(P+T)}]$; 
   
   (ii) when $\frac{P}{P+T} \geq q_G$, then $\nu^*_I = \nu^*_G = 1$, and $x^* \in [0, 1]$.

3. If $\frac{q_G}{q_I} < \frac{(1-q)(L_1+kT)}{q(L_2-kT)}$, then $\nu^*_I = \nu^*_G = 1$, and $x^* = 0$. 

14
In the case when $\frac{P}{P+T} \leq q_1$, there also exists a pooling equilibrium in which
\( \nu^*_I = \nu^*_G = 0 \), and \( x^* \in \left[ \frac{P}{q_1(P+T)}, 1 \right] \). This is supported by the posterior belief of the judge that any non-confessing suspect is guilty with probability \( q^* \in \left[ \frac{L_1+kT}{L_1+L_2}, 1 \right] \).

**Proof.** See Appendix.

Broadly speaking, there are two kinds of equilibrium. The first is a pooling equilibrium in which both types of suspect confess. This is supported by the judge’s belief that a non-confessing suspect is highly likely to be guilty. Under this belief, he tortures with so high a probability that both types of suspects confess before investigation. The second, and more interesting, type of equilibrium is one in which the innocent type never confesses (\( \nu^*_I = 1 \)). The judge’s probability to torture, \( x^* \), depends on the denying probability of the guilty, \( \nu^*_G \). The higher its value, the greater his posterior that a denying suspect is guilty and, therefore, the greater the chance he will torture. Figure 1 plots the non-pooling equilibrium as a function of the parameters $\frac{P}{P+T}$ and $\frac{(1-q)(L_1+kT)}{q(L_2-kT)}$.

The value of $\frac{P}{P+T} = \frac{1}{1+T}$ can be seen as the relative disutility of penalty to torture for a suspect. The greater its value, the less likely the suspect will confess. Indeed, as we move up vertically from any point in Figure 1, the value of $\nu_G$ increases. This, however, is also accompanied by an increase of $x^*$, as the posterior that a denying suspect being guilty also rises when $\nu_G$ increases. When $\nu_G$ equals 1,\(^\text{10}\) the judge’s posterior that the suspect is guilty, after he draws a $g$, is so high that he will torture with probability 1. This is the equilibrium outcome in region A of Figure 1. Note that in this region, all suspects still deny with probability 1 even if they know that they will be tortured with certainty when the judge draws $g$. The reason for this is that $P$ is so large compared to $T$ that they are willing to take the chance of going through investigation.

\(^{10}\) Or equivalently, when $\frac{P}{P+T}$ is greater then $q_G$. 

15
The value of $(1-q)(L_1+kT)q(L_2-kT)$ is the expected cost of type I error relative to that of type II error. As its value increases, the judge will be less inclined to torture. This can be seen by moving from any point in Figure 1 horizontally to the right. In that case, although the value of $x^*$ remains fixed, the value of $\nu_G$ will increase, meaning that the guilty suspect becomes less likely to confess. If the value of $(1-q)(L_1+kT)q(L_2-kT)$ is large enough (specifically, larger than $\frac{qG}{qI}$), either because the prior $q$ is low or loss of type I error is large, then the judge will be unwilling to torture even if he draws a $g$. In that case $x^* = 0$, and the suspect will never confess. This is the equilibrium outcome in region C of Figure 1. In region B, the value of legal penalty is low so that the guilty suspect is inclined to deny guilt. This, however, is tempered by the high probability that the judge will torture (since the value of $\frac{(1-q)(L_1+kT)}{q(L_2-kT)}$ is low). In the equilibrium, he confesses with probability greater than 0 but less than 1.
It can be seen clearly that of all equilibria, the guilty and the innocent act differently only in region B of the non-pooling equilibrium. That is, only in region B does the torture-based system force the guilty suspect to confess more readily. This indicates that the torture-based system can be superior only in region B, a fact that will be proved in the next section.

Proposition 1 also implies that the judge uses a more lenient rule to convict a suspect under the torture-based system. Recall that under the evidence-based system, the judge convicts the suspect if and only if he draws a $g$. Here it is still the case that the suspect will be released when $i$ is drawn. However, the judge will not necessarily torture (and convict) the suspect even if a $g$ is drawn.

5 Comparison

The central question that the paper intends to answer is which of the two systems we consider renders higher social welfare. Torture, by forcing a suspect to confess regardless of whether he is guilty or not, decreases the chance of type II error at the expense of both the pain imposed on the suspect and increase in type I error. As a result, the relative merit of the two systems critically depends on two factors. The first is the relative size of social welfare loss brought about by type I and type II errors. The second is the accuracy of the judge’s investigation. In this section we will first derive precisely under what condition one system is better than the other. Based on this comparison, we will then show how their relative merit will change in response to improvement in information revealed during investigation. Our result is then used to explain certain aspects of the historical evolution in the judicial system in Section 6.
5.1 Welfare Comparison

Under the evidence-based system, with probability $q(1 - q_G) + (1 - q)(1 - q_I)$ the judge will draw an $i$, in which case the suspect will be released. Since there will not be type I error, the expected social loss is $q(1 - q_G)L_2$. If he draws a $g$, which occurs with probability $qq_G + (1 - q)q_I$, the judge will convict the suspect. This results in an expected social loss of $(1 - q)q_I L_1$. The total expected social loss under the evidence-based system is thus

$$W^e = q(1 - q_G)L_2 + (1 - q)q_I L_1.$$  \hspace{1cm} (3)

The expected social loss of the torture-based system depends on which equilibrium outcome we consider. For the pooling equilibrium in which $\nu^*_I = \nu^*_G = 0$, $x^* \in \left[\frac{p}{q_I(P+T)}, 1\right]$, the expected social welfare loss is $(1 - q)L_1$.\(^{11}\) As a result, the difference in social welfare loss between the two systems is

$$\Delta W = (1 - q)L_1 - [(1 - q)q_I L_1 + q(1 - q_G)L_2]$$

$$= (1 - q)(1 - q_I)L_1 - q(1 - q_G)L_2.$$ \hspace{1cm} (4)

As a result, $\Delta W \leq 0$ if and only if

$$\frac{1 - q_G}{1 - q_I} \geq \frac{(1 - q)L_1}{q L_2}.$$ \hspace{1cm} (5)

Inequality (5), however, violates assumption A1. This means that if an investigation is of certain informational value in the sense of A1, then an evidence-based system always results in lower social welfare loss than the pooling equilibrium in the torture-based system. The reason for this is as follows. Under the pooling equilibrium, the suspect always confesses, and the judge never needs to investigate. The torture-based system thus has not utilized the information that would have been revealed

\(^{11}\) Note that under the pooling equilibrium, both confess at the equilibrium, so the judge does not need to investigate. We thus do not need to distinguish the cases when he draws $g$ and $i$.  

18
had there been an investigation. Given that the investigation is of informational value, the evidence-based system, as one in which the judge’s decision is based on evidence, is naturally better.

For the non-pooling equilibrium, the comparison becomes much more complicated, because the equilibrium outcome depends on parametric configuration. In general, the expected social welfare loss under the torture-based system is

\[
W^T \equiv (1 - q)(1 - \nu_I^* )L_1 + (1 - q)\nu_I^* q_1 x^* L_1 + q\nu_G^* [(1 - q_G) + q_G (1 - x^*)] L_2 \\
+ [(1 - q)q_1 \nu_I^* + \nu_G^* qq_G] x^* kT
\]

where the last equality comes from the fact that \( \nu_I^* \) is always 1 in non-pooling equilibrium. In region A of Figure 1 where \( \nu_I^* = \nu_G^* = x^* = 1 \), \( W^T \) reduces to

\[
(1 - q)q_1 x^* L_1 + q\nu_G^* [(1 - q_G) + q_G (1 - x^*)] L_2 + [(1 - q)q_1 + \nu_G^* qq_G] x^* kT;
\]

Thus we have

\[
\Delta W = W^T - W^e = [(1 - q)q_1 + qq_G] kT > 0.
\]

That is, if the suspect always denies and the judge always tortures (when he draws \( g \)) under the torture-based system, then the evidence-based system is always superior. The reason for this is that under the equilibrium in region A, the judge will torture (and convict) a suspect if and only if he draws \( g \). Thus the criteria to release or to convict a suspect are exactly the same in the two systems. This implies the torture-based system is inferior by exactly the magnitude of the expected cost of torture.

In region C of Figure 1 where \( \nu_I^* = \nu_G^* = 1 \) and \( x^* = 0 \), \( W^T \) reduces to \( q L_2 \). Thus

\[
\Delta W = q L_2 - (1 - q)q_1 L_1 - q(1 - q_G) L_2 \\
= qq_G L_2 - (1 - q)q_1 L_1.
\]
Therefore, $\Delta W > 0$ if and only if

$$\frac{(1 - q)L_1}{qL_2} < \frac{q_G}{q_I},$$

which always holds by assumption A1. This implies that if the suspect always denies and the judge never tortures under the torture-based system, then the evidence-based system is also superior. The reason for this is actually very simple. In the case when $\nu^*_I = \nu^*_G = 1$ and $x^* = 0$, a system based on torture actually does nothing. Suspects are routinely summoned to the court and released. Given that the investigation is of certain informational value (assumption A1), a system that convicts a suspect based on evidence is naturally superior.

All the three cases above share a common feature: Information that is available through investigation is not utilized in the tortured-based system in an efficient way. Either all types of suspect confess so investigation is not needed, or all who deny are released, or the judge convicts the suspect in exactly the same fashion as the less costly evidence-based system. Since information offered by investigation is valuable, and since an evidence-based system does utilize the information revealed in investigation, the torture-based system always incurs greater social welfare loss.

To summarize, if the judge’s decision is not altered by the outcomes of investigation in a torture-based system, or the criterion to convict a suspect is the same as in the evidence-based system, then it is inferior.

The more interesting case, however, occurs when the equilibrium outcome falls in region B of Figure 1, i.e., when the evidence collected will affect the judge’s decision in a way different from the evidence-based system. In that case, the difference in social welfare loss is

$$\Delta W = - (1 - q)(1 - x^*)q_t L_1 + q[\nu(1 - x^* q_G) - (1 - q_G)]L_2$$

$$+ [(1 - q)q_t + \nuqq_G]x^*kT$$

20
\begin{align*}
&= q_I (1 - q)x^*(L_1 + kT) - x^*qq_G\nu(L_2 - kT) \\
&\quad - (1 - q)q_I L_1 - q(1 - q_G - \nu)L_2 \\
&= - [(1 - q)q_I L_1 + q(1 - q_G - \nu)L_2];
\end{align*}

where that last equality in (6) comes from the fact that the first two terms after the second equality sum to 0. Note that (6) is an increasing function of \( \nu \). Moreover, \( \Delta W < 0 \) when \( \nu = 0 \) and \( \Delta W > 0 \) when \( \nu = 1 \). Therefore, there exist a \( \nu^0 \equiv (1 - q)q_I L_1 q(L_2 - kT) \) such that \( \Delta W > 0 \) if and only if \( \nu > \nu^0 \). However, since \( \nu = \frac{q_G (1 - q)(L_1 + kT)}{q(L_2 - kT)} \), we know that \( \Delta W > 0 \) if and only if \( \frac{(1 - q)(L_1 + kT)}{q(L_2 - kT)} \geq \nu^0 \frac{q_G}{q_I} \).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.pdf}
\caption{Welfare comparison between two systems.}
\end{figure}

We can now summarize all our comparison results in Figure 2. The curve \( \frac{(1 - q)(L_1 + kT)}{q(L_2 - kT)} = \frac{q_G}{q_I} \nu^0 \) partitions region \( B \) in Figure 1 into two parts \( (B_1 \text{ and } B_2) \).\footnote{Since \( \nu^0 \) is a function of \( q, q_I, q_G, L_1 \) and \( L_2 \) (by (6)), this curve is not necessarily a straight line.}

\[\begin{array}{c}
\nu^0 q_G
\\
\end{array}\]
Configurations lying in $B_1$ ($B_2$) will result in equilibria having lower (higher) social welfare loss under a torture-based system. Thus among all possible equilibria, the torture-based system is superior only when the configurations of the parameters lie in region $B_1$, which have lower values of both $P/(P+T)$ and $\frac{(1-q)(L_1+kT)}{q(L_2-kT)}$. That is, a torture-based system is superior only when relative disutility of legal penalty to torture is low, and expected cost of type I error relative to that of type II error is low. This is an intuitive result. Recall that torture incurs two kinds of cost to the society; one is the pains of torture and the other the increase in type I error. Thus only when type II error incurs serious cost in the judge’s eyes, and legal penalty is not great enough to be deterring, can the judge be justified in resorting to torture. However, if one of the conditions is not met, then torture is more costly, and a system based solely on evidence will be better.

5.2 Improvement in Informativeness of Investigations

In this subsection, we investigate how the behavior of the suspects and judge will change in response to change in the information revealed during investigation. As can be seen in Proposition 1, an important parameter that influences the behavior of both the judge and the suspect is $q_G/q_I$, which is a measure of the precision of the judge’s investigation. The value of $q_G/q_I$, however, is influenced by two independent factors. The first is the probability that the judge draws a guilty signal for the guilty suspect, $q_G$. The other is the probability that the judge draws an innocent signal for the innocent suspect, $1-q_I$. Either an increase in $q_G$ or a decrease in $q_I$ implies that more precise information is released during investigation and, therefore, an increase in $q_G/q_I$. We will thus say investigation becomes more informative if $q_G$ increases and $q_I$ decreases, with at least one strictly so.

When investigation becomes more informative, the outcomes in evidence-based
and torture-based systems will both be affected. In the evidence-based system, both inequalities in assumption A1 become more likely to hold. Consequently, the judge is more likely to release a suspect when he draws \( i \), and more likely to convict when he draws \( g \).\(^{13}\) Moreover, the expected social losses of type I error (when he convicts a suspect) and type II error (when he releases) both decrease. This can be easily seen from (3): As investigation becomes more informative, \( W^E \) will decrease. In particular, when \( q_G \) approaches 1 and \( q_I \) approaches 0, the social loss of an evidence system approaches 0.

In the torture-based system, both \( \hat{q}_G \) and \( \hat{q}_I \) will increase as investigation becomes more informative. Furthermore, the line \( \frac{(1-q)(L_1+kT)}{q(L_2-kT)} = \frac{q_G}{q_I} \) will shift to the right. Thus, region \( C \) will retract, and regions \( A \) and \( B \) will expand. That means the judge is now more likely to torture on the one hand, and the guilty suspect becomes more likely to confess on the other. Note that as \( q_G \) approaches 1 and \( q_I \) approaches 0, \( \nu \) will approach 0 and \( x^* \) will approach \( P/(P+T) \). This resembles a separating equilibrium. But unlike the evidence-based system, what separates the innocent and guilty is not evidence, but the fact that the innocent always denies and the guilty always confesses. This relies on the judge’s commitment that his probability to torture is bounded away from 0.

It can be shown that \( \frac{q_G \nu^0}{q_I} \) is a decreasing function of \( q_I \), meaning that as \( q_I \) becomes smaller, the dividing curve between \( B_1 \) and \( B_2 \) will shift to the right. In other words, the torture-based system will gain its advantage as \( q_I \) decreases. Moreover, \( \frac{q_G \nu^0}{q_I} \) is increasing (decreasing) in \( q_G \) if

\[
q_G < (>)(1-q)q^{-1}q_1L_1L_2^{-1} + 1)/2.
\]

That is, as the value of \( q_G \) increases from a small value, the value of \( \frac{q_G \nu^0}{q_I} \) will first

\(^{13}\) That is, the configuration of \( q, L_1, L_2 \) that make the judge release the suspect when he draws \( i \) and to convict when he draws \( g \) will increase.
increase, and then decreases as the value of $q_G$ passes the value on the right-hand side of (7). This implies that the dividing curve between $B_1$ and $B_2$ will first shift right then left. Consequently, an increase in $q_G$ will favor the torture-based system when $q_G$ is small, and favor the evidence-based system when it is large. Note that the smaller the value of $q_I$, the more likely $q_G$ is greater than the right-hand side of (7). Thus at the same time the judge becomes more competent in identifying the innocent suspect, an increase in $q_G$ is more likely to lend advantage to the evidence-based system. In particular, when $q_G$ is sufficiently large, then $\Delta W \geq 0$ regardless of the value of $q_I$. This can be seen from the following fact: if $q_G$ is sufficiently close to 1, then $\nu^0$ is close to $\frac{(1-q)qL_1}{qL_2}$, so that it is always the case that $\frac{(1-q)(L_1+kT)}{q(L_2-kT)} \geq \frac{\nu^0}{q_I}$. In other words, as the judge’s ability to identify a guilty suspect through investigation passes a certain threshold, there will be no more room for torture: A system based on evidence always yields greater social welfare.

The picture that emerges from the theoretical results is that when little information is revealed during investigation, then as the investigation becomes more informative, torture will be used more frequently. However, as investigation becomes sufficiently informative (i.e., $q_I$ is small enough, or $q_G$ large enough), then further improvement in the information it reveals will start to lend advantage to the evidence-based system, and eventually guarantee its domination. This theoretical result implies that the historical development of torture will first go hand in hand with the advancement of investigation technology, then — as investigation becomes sufficiently sophisticated — yield to a system based on evidence. In the following section, drawing from a wide variety of historical records, we will show that this is indeed the case.
6 Historical Development of Torture

The decline of judicial torture in modern Europe has been the subject of a number of scholarly works in the past few decades. The conventional historical account of the abolition movement placed great weight on philosophic goals of the Enlightenment, which is considered as the driving forces behind the abolition of torture.\(^{14}\) We will call this the humanist theory. J. Langbein, as a law specialist, proposes a more specific albeit controversial explanation based on changes within the penological history itself. He suggests that judicial torture became increasingly useless when a new law of proof emerged around the 16th century that no longer required a strict standard of proof, accompanied by less rigorous punishments, a development he calls “evidentiary revolution” (Langbein 1977, 1983). We will call this the legal theory. Langbein’s thesis was criticized by another law specialist M. Damaška, especially on the significance of the “evidentiary revolution” in the 16th century. For Damaška, judicial torture still had a raison d’être in Europe during and after the 16th century (Damaška 1978).

More recently, historian L. Silverman proposes another explanation for the final abolition of judicial torture in 18th-century France, partially rehabilitating the humanist theory rejected by Langbein. For Silverman, there was a “dramatic paradigm shift” in the way abolitionists of the Enlightenment understood the relation to judicial torture (pain), truth and the body. Bodily pain was no longer believed to be able to produce truth. (Silverman 2001.) One reason Silverman does not find satisfaction in Langbein’s theory is that long after the “evidentiary revolution”, judicial torture was still practiced in many parts of Europe, a point also raised in Damaška’s review.

Indeed one difficulty in explaining the historical decline of judicial torture was

\(^{14}\) See discussion in Silverman (2001), and the references therein.
the synchronization of the judicial and cultural changes with the rise, decline, and eventual abolition of judicial torture itself. Despite the tremendous efforts made by historians and legal scholars, a time gap seems to exist in most explanations. As Silverman herself confesses, “In delineating the decline in both the use of torture and in the rate of confession, scholars have failed to account for the continued reliance on torture” (Silverman 2001, pp.18-19).

European countries do not have exactly the same history of judicial torture, as each had a different judicial tradition and religious culture that greatly affects legal application. However, from the specific cases of England and France, one may construct a picture reflecting the general history of judicial torture in Europe: the application of judicial torture began to appear in significant numbers from around the 13th century15 onward and peaked in the 16th century. It declined thereafter and was nominally abolished in most European countries towards the end of the 18th and early 19th centuries. In England, “references to its use (use of judicial torture) in the earlier years are remarkably scanty, but during the fifteenth, sixteenth and seventeenth centuries, the evidence is abundant...(J)udicial torture reached its greatest ecumenicity in the reign of Elizabeth (1533-1603).” (Scott 2003, pp. 88-89.) In France, in a southern town such as Toulouse where the tradition of Roman law was strong, judicial torture “was first employed... during the 13th century by the town consuls”, it “reflected a new methodology and a new epistemology” (Silverman 2001, pp.5-7). The decline of the actual use of judicial torture in France in general seemed to begin from the mid-16th century to the 17th century, depending on the region, until its nominal abolition in 1788. (Silverman 2001, p.18.) The pattern corresponds roughly to the one observed in England. By 1850 the movement for the abolition of torture swept most parts of Europe. And “by the closing decades

15 "Torture was introduced for the express purpose of extracting confession," being authorized by Pope Innocent in a Bull issued in 1252 (Scott 2003, 66).
of the 19th century it was widely thought that torture was a barbaric practice that belonged to history.” (Evans and Morgan 1998, pp. 12-13). Although recent scholarship still casts doubt on the real disappearance of judicial torture in Europe by the end of the 19th century, as evidence suggests the persistence of illegal use of judicial torture in a democratic country like the US even in the 1920s (Evans and Morgan 1998, p.14), such evidence, however, shows that torture was no longer the norm. The decline of judicial torture was therefore a continuous process beginning from the 16th century up to the early 20th century. It was legally and ideologically condemned by the end of the 18th century, and in practice, it may persist up to the turn of the 20th century.

This historical evolution of judicial torture in Europe corresponds meaningfully with our model if we look at another factor that might explain in a more satisfactory way the rise and decline of judicial torture: the employment of medical science in law. There are indications of the application of medical expertise in judicial cases in northern Europe from the 10th century onward, and was explicitly mentioned in Norman law from the early 13th century onward. Between the 13th and the 16th centuries the trend was on a stable increase until the publication in 1562 of the first judicial postmortem in France by Ambroise Paré, although the conventional wisdom is that the first comprehensive work on forensic medicine, De relationibus medicorum libri quatuor was published by Fortunato Fedele (1550-1630) in 1602. An increasing number of works of the kind were published throughout the 17th and 18th centuries with the first serial devoted to forensic medicine published in 1755. These early European publications on forensic medicine indicate that the practice of judicial forensic medicine was already common throughout the 16th century, if not earlier. Later developments in the science of fingerprinting (from the late 17th century), physical matching (late 18th century), anthropometry (late 19th century), and so
on were the intensification of the same trend. More significantly, the systematic application of modern scientific technology in criminal investigations began in the last decades of the 19th century. Most notable was the new classification system created by E. Henry in 1896 enabling fingerprints to be easily filed, searched and traced. It was quickly used worldwide and is still used today.\textsuperscript{16} Henry himself was appointed assistant administrator of Scotland Yard in 1901 in charge of the Criminal Investigation Department, and forced the adoption of fingerprint identification.\textsuperscript{17} New York State in the US followed in 1903.\textsuperscript{18} Most European countries established crime detection laboratories as government or university units in the first quarter of the 20th century. (Parker 1983, p.432.) In other words, there were two critical points in the development of forensic medicine in criminal investigation: first, the 16th century marked the beginning of modern forensic medicine for legal uses (and the “evidentiary revolution”). Second, the end of the 19th century marked a paradigm shift in such technology and its systematic application in criminal investigation. (Nemec 1968, pp.5-15; Saferstein 2001, p.3; Parker 1983, p.430.) At the same time, the 16th century represents the peak of the legal employment of torture in most parts of Europe. The decline, abolition, and the ultimate disappearance of judicial torture in democratic countries would take more than three hundred years.

To explain this development in term of our theoretical results, first note that although medical expertise had started to be used in judicial cases since the 13th century (so that informativeness in investigation had improved), there was actually an increase in the use of torture. This is consistent with our prediction that improvement in technology of investigation lends advantage to torture, when investigation


is relatively uninformative. Only after the 16th century (the beginning of the use of modern forensic science) did the practice of torture start to decline. This is also consistent with our prediction that as technology of investigation is sufficiently advanced, increase in $q_G$ will give advantage to an evidence-based system. Finally, at the end of 19th century, when more advanced testing methods (like fingerprints and blood type identification) started to be accepted in court and systematic use of scientific investigation for legal use became common (this corresponds to $q_G$ passing the threshold in our theory), a system based on evidence dominated thereafter.

7 Torture as a Screening Device

In this section, we consider the potential informational advantage in torturing the plaintiff or the witness. Since the basic logic behind our argument is the same for torturing the plaintiff and the witness, we will consider only the former. The model is modified in the following way. Suppose before the judge accepts a case, he can torture the plaintiff, ostensibly to “verify” or “test” whether the plaintiff is telling the truth. Once tortured, the disutility for the plaintiff is $T_p$, and it also yields a social loss, $k_p T_p$, where $k_p < 1$. The events then follow as described in the original model.

By torturing the plaintiff, the incentive of the plaintiff to come to the court is reduced. Moreover, it is those plaintiffs who are less sure of the suspects’ guilt that are deterred. Consequently, the average quality of cases brought to the court is increased in the sense that the probability that suspect being guilty is higher. This reduces type II error. More importantly, when the equilibrium involves positive probability of torturing the suspect, raising the average quality of cases will also decrease the loss associated with type I error. The judge thus trades off the social
loss resulting from torturing the plaintiff with the benefit of reducing both type I and II errors. If the gain from screening (reduction in the loss of both types error) is greater than its cost (pain of the plaintiff), then torturing the plaintiff can improve social welfare. Consequently, torturing the plaintiff serves as a screening device.

In previous sections, we have assumed that the prior of the judge that the suspect is guilty is \( q \). Here in order discuss how torture changes the plaintiff’s incentives to come to the court, and thus the judge’s prior, we assume that \( q \) is determined by the cases that plaintiff brings into the court. Specifically, suppose that every plaintiff has a belief regarding how likely the suspect is guilty, represented by a probability \( \rho \). That is, he believes that the suspect is guilty with probability \( \rho \). The distribution of plaintiffs’ beliefs is given by the density function \( l(\rho) \). Hence the prior of the judge, if all plaintiffs go to court, will be \( \int_0^1 \rho l(\rho) d\rho \). However, not all plaintiffs go to court. Thus what we call the prior of the judge in the previous sections is actually the mean value of the belief of the plaintiff which enters the court.

Assume that the plaintiff obtains a utility \( R > 0 \) when a guilty suspect is convicted. If the guilty suspect is wrongly released, or the innocent suspect is wrongly convicted, the plaintiff is assumed to obtain a negative utility \(-r, r > 0\). The plaintiff’s utility is 0 if an innocent suspect is released. Therefore, given the probability that the judge tortures the plaintiff, \( \varphi \), the expected payoff for the plaintiff with belief \( \rho \), when he comes to the court, is

\[
\rho[(1 - \nu_G^* + \nu_G^* q_G x^*) R - \rho \nu_G^*(1 - q_G x^*) r - (1 - \rho)(1 - \nu_I^*) + \nu_I^* q_I x^*) r - \varphi T_p]. \tag{8}
\]

If, on the other hand, the plaintiff decides not to come into the court, his expected payoff is simply \(-\rho r\).

To see how torturing the plaintiff affects the equilibrium of the torture game, we again need to discuss two types of equilibrium in Proposition 1 separately. In the pooling equilibrium in which \( \nu_I^* = \nu_G^* = 0 \) and \( x^* = 1 \), (8) becomes \( \rho R - (1 - \rho) r - \varphi T_p \).
Therefore, a plaintiff with belief \( \rho \) will come into the court if and only if \( \rho R - (1 - \rho) r - \varphi T_p \geq -\rho r \). Note that a plaintiff with \( \rho = 0 \) will never go to court, as it implies a lower utility regardless of the value of \( \varphi \). For a plaintiff with \( \rho = 1 \), whether he will go to court depends on the value of \( R - \varphi T_p \) versus \(-r\). We will make the assumption that \( R + r > T_p \). In that case, there exists a \( \rho(\varphi) \in [0, 1] \) so that a plaintiff with belief \( \rho \geq (\varphi) \rho(\varphi) \) will (will not) come into the court; where

\[
\rho(\varphi) = \frac{\varphi T_p + r}{R + 2r}. \tag{9}
\]

Given that only those plaintiffs whose beliefs are greater than \( \rho(\varphi) \) will go to court, the belief of the judge that the suspect is guilty is then

\[
q(\varphi) = \int_{\rho(\varphi)}^{1} \frac{\rho l(\rho)}{1 - L(\rho(\varphi))} d\rho; \tag{10}
\]

where \( L(\cdot) \) is the distribution function of \( \rho \). Note that in our previous setup, the event starts when a case has already entered into the court. We will interpret \( q(\varphi) \) as the prior of the judge. Since in the previous setup the judge does not torture the plaintiff, it corresponds to the case \( \varphi = 0 \). Consequently, \( q(0) = q \), that is, the prior \( q \) in the previous sections is actually \( q(0) \) in the current setup.

After the plaintiff brings the case into court, the judge chooses an optimal probability of torturing the plaintiff, \( \varphi^* \), to minimize the expected social loss:

\[
\min_{\varphi} \ W^T(\varphi) = (1 - q(\varphi)) L_1 + \varphi k_p T_p. \tag{11}
\]

A full characterization of \( \varphi^* \) is difficult except that it must satisfy the first-order condition. This is because a change in \( \varphi \) will affect the judge’s and suspect’s behavior during the investigation in a complicated way that is hard to capture simply. However, we can find a simple sufficient condition under which the judge will torture the plaintiff with positive probability. Differentiating \( W^T \) with respect
to $\varphi$, evaluated at the point where $\varphi = 0$, we have
\[
\frac{\partial W_T}{\partial \varphi} \bigg|_{\varphi=0} = k_p T_p - L_1 \frac{l(\rho(0))(q - \rho(0)) T_p}{(R + 2r)(1 - L(\rho(0)))},
\]
which is negative if
\[
L_1 > \frac{k_p(R + 2r)(1 - L(\rho(0)))}{l(\rho(0))(q - \rho(0))}.
\]
This means that if $L_1$ is sufficiently large or $k_p$ is sufficiently small, then increasing the value of $\varphi$ slightly from 0 will decrease the value of social loss in the torture-based system. Consequently, at the optimum the judge will torture the plaintiff with strictly positive probability.

Inequality (12) has a very simple intuition. Recall that in the pooling equilibrium all types of suspect confess, so that only type I error is committed. If the cost of type I error is large enough, then the benefit in raising the quality of cases entering into court (by torturing plaintiff) outweighs its cost.

Also note that (12) is less likely to hold when either $R$ or $r$ is large. This is again an intuitive result. If the plaintiff or witness has strong sense of “justice” in that he derives (suffers) great utility (disutility) in having a guilty (innocent) suspect convicted, then torture will be less likely to deter him from coming into court. The screening function of torture is thus small, and will be less likely to be applied on the plaintiff.

The second case to consider is the non-pooling equilibrium. Here the expected payoff for the plaintiff if he comes into the court is
\[
\rho[(1 - \nu_G^*) + \nu_G^* q_G x^*] R - \rho \nu_G^* (1 - q_G x^*) r - (1 - \rho) q_I x^* r - \varphi T_p.
\]
Thus, he will come to the court only if $\rho \geq \rho(\varphi)$, where
\[
\rho(\varphi) = \frac{T_p + q_I x^* r}{1 - \nu_G^* + \nu_G^* q_G x^*}(R + r) - \nu_G^* r.
\]
Similarly, we can define the judge’s prior \( q(\varphi) \). Again, we assume that \( q(0) = q \) to facilitate comparison with case when the plaintiff is not tortured. The judge chooses an optimal probability of torture, \( \varphi^* \), to minimize the expected social loss in this equilibrium:

\[
\min_{\varphi} W^c(\varphi) = (1 - q(\varphi))q_t x^* L_1 + q(\varphi)\nu^*_G (1 - x^* q_G) L_2 \\
+ \{(1 - q(\varphi))q_t + q(\varphi)\nu^*_G q_G\} x^* kT + \varphi k_p T_p. \tag{15}
\]

In the special case when \( \nu^*_t = \nu^*_G = 1 \) and \( x^* = 0 \) (i.e. region C in Figure 1), the expected payoff of the plaintiff when he comes to the court is (by equation (13)) \(-\rho r - \varphi T_p\). As the expected payoff of a plaintiff is \( \rho r \) if he does not come to court, we know that once \( \varphi > 0 \), no plaintiff will be coming to the court. Assume that the judge’s prior is \( q \) in this case, then the expected social welfare loss is \( q L_2 + \varphi k_p T_p \).

Obviously, the value of \( \varphi \) that minimizes social welfare loss is 0. Consequently, \( \varphi^* = 0 \). We thus have

**Lemma 5.** *In the equilibrium in which the suspect never confesses and the judge never tortures, the judge will not torture the plaintiff either.*

In region A of Figure 1 where \( x^* = \nu^*_t = \nu^*_G = 1 \),

\[
\left. \frac{\partial W^T}{\partial \varphi} \right|_{\varphi=0} = k_p T_p + [(1 - q_G) L_2 - q_t L_1 + (q_G - q_t)kT] \frac{\partial q(\varphi)}{\partial \varphi} \bigg|_{\varphi=0}, \tag{16}
\]

which is negative if\(^{19}\)

\[
(L_1 + kT) > \left[ \frac{q_G}{q_t} (R + r) + r \right] H + \frac{L_2}{q_t} - \frac{q_G}{q_t} (L_2 - kT), \tag{17}
\]

where \( H \equiv k_p / [\rho(0) \frac{\partial q(\varphi)}{\partial \varphi} |_{\varphi=0}] \). Since \( q_G / q_t > (1 - q)(L_1 + kT) / q(L_2 - kT) \) in region A, a sufficient condition for (17) to hold is

\[
(L_1 + kT) > q \left[ \frac{q_G}{q_t} (R + r) + r \right] H + q L_2. \tag{18}
\]

\(^{19}\) We use the fact that \( \frac{\partial q(\varphi)}{\partial \varphi} |_{\varphi=0} = \frac{[c(0) - q_c(0)]T_p}{q_c(R + r) + qr}. \)
The important thing to note is that (18) is more likely to hold if \( q_G/q_I \) is small. That is, torturing the plaintiff is more likely to be welfare-improving when investigation is less informative.

In region B of Figure 1 where \( \nu_G^* < 1 \) and \( x^* < 1 \),

\[
\frac{\partial W^T}{\partial \varphi} \bigg|_{\varphi=0} = -k_p T_p + \left\{ \frac{\nu T}{P + T} L_2 - \frac{q_I P}{q_G(P + T)} (L_1 + kT) + \frac{P q_G}{P + T} k_T \right\} \frac{\partial q(\varphi)}{\partial \varphi} \bigg|_{\varphi=0} \\
+ q(\varphi) \left( (1 - x^* q_G) L_2 + q_G x^* kT \right) \frac{\partial \nu_G^*}{\partial q} \frac{\partial q(\varphi)}{\partial \varphi} \bigg|_{\varphi=0}. \quad (19)
\]

Since \( \frac{\partial \nu_G^*}{\partial q} \leq 0 \), we know that \( \frac{\partial W^T}{\partial \varphi} \bigg|_{\varphi=0} < 0 \) if the first two terms in (19) is negative. Substituting for \( \nu = \frac{q_I (1-q) (L_1 + kT)}{q_G q (L_2 - kT)} \) and \( \frac{\partial q(\varphi)}{\partial \varphi} \bigg|_{\varphi=0} \) into (19), we know that (19) is negative if

\[
(L_1 + kT) > \frac{q_G P + T k_p [\rho(0) - \rho(0)]}{1 - L_1(\rho(0))} \rho'(0). \quad (20)
\]

Condition (20) is more likely to hold if \( k_p \) is small, or \( \frac{\partial q}{\partial \rho(\varphi)} \bigg|_{\varphi=0} \) is large. That \( k_p \) needs to be small is easy to understand. A large \( \frac{\partial q}{\partial \rho(\varphi)} \bigg|_{\varphi=0} \) means that the plaintiff’s belief are more concentrated on small \( \rho \) (i.e., in most cases the plaintiffs are not quite sure that the suspect is guilty), so that the deterring effect of torture is strong. Condition (20) is also more likely to hold when \( R + r \) (which enters into (20) via \( \rho(0) \)) is relatively small. The reason is similar to the pooling equilibrium case. If the plaintiff does derive much utility from convicting the guilty, or does care much about type II error, then they will bring in cases during facing torture. In that case the function of torture as a screening device will be small.

Again, (20) is less likely to hold when \( q_G/q_I \) is large. If information revealed during investigation is relatively precise, then the chance of committing either type of error is low. In that case there is no need to torture the plaintiff to reduce their costs.
In summary, our result shows that when investigation is less informative, torturing the plaintiff or the witness can be welfare-improving. This means that there is a negative relation between the informativeness of investigation and the tendency to torture the witness or the plaintiff. Moreover, since it is also optimal to torture the suspect when \( q_G/q_I \) is low, the two practices (torturing the suspect and torturing the witness) generally go hand in hand.

8 Conclusion

This paper applies the economic theory of information to analyze the nature and consequences of judicial torture. It is shown that if the judge aims to balance type I and type II errors in judgment, and if during investigation little information is revealed, then torture can maximize social welfare by forcing the guilty suspect to confess with higher probability. However, as the information revealed during investigation passes a certain threshold, the advantage of torture decreases relative to a system based on evidence. This may explain an increase of the practice of torture before 16th-century, on the one hand, and a transition from a torture-based system to an evidence-based system starting from the mid-18th century in Europe on the other. Furthermore, torturing the witnesses might also be welfare-improving, as it helps to screen the cases so that only those with greater merits enter the court. Again, when investigation reveals little information (so that the judge needs to rely on torture), then this increase in case quality will improve social welfare.

We also use the results implied by our framework to explain the rise and fall of torture in Europe. There have been two major theories to explain the abolition of torture, one based on humanist concern and the other based on legal change. In this paper we propose an alternative theory based on the progress of technology (one may
call it the technological theory). We hasten to add that our theory does not mean to replace, or even compete with, the existing explanations. It might very possibly be the case that all of the three factors contributed to the decline of torture, and in this sense they are complementary to each other. In fact the humanist theory and the legal theory can both comfortably fit into our framework. The reason why there had been a change in rule of proof might be precisely because the improvement in the technology of evidence collection has been such that a system based on solely evidence was found to reach more precise verdict. On the other hand, change in the attitude of the general public towards torture is equivalent to either a raise in $k$ or $L_1$ in our model, both of which shift the curve dividing $B_1$ and $B_2$ to the left, and therefore reduces the advantage of torture. From this viewpoint, the contribution of this paper might be seen as providing a unifying framework in which the humanist, legal, and technological theories are all important components in explaining the rise the fall of judicial torture.

**Appendix: Proof of Proposition 1**

First note that an innocent suspect will confess with positive probability only if $P \leq q_I (P + T)x$, i.e., the loss of confessing is smaller than that of denying. The similar condition for a guilty suspect is $P \leq q_G (p + T)x$. Also, given the posterior $\hat{q}_G$, the judge has positive probability to torture only if $\hat{q}_G L_2 \geq (1 - \hat{q}_G) L_2 + kT$, or, equivalently, $\frac{q_G}{q_I} \geq \frac{(1-q)\nu_I (kT + L_2)}{qG(kT - L_2)}$.

We first focus on the equilibrium in which $\nu_I^* > 0$.

[1]. Suppose $\frac{qG}{qI} > \frac{(1-q)(L_1 + kT)}{q(L_2 - kT)}$.

First of all, we only have to consider the case $x > 0$ in equilibrium. The reason is as follows. If $x = 0$, then $\nu_I = \nu_G = 1$, because it is always the case where
\[ P > q_I (P + T)x \text{ and } P > q_G (P + T)x. \]
However, if this is the case, then \( x = 1 \) due to the assumption \( \frac{q_G}{q_I} > \frac{(1-q)(L_1+kT)}{q(L_2-kT)} \), which is a contradiction. Secondly, if in equilibrium \( x \in (0,1) \), then it must be the case that \( \nu_I = 1 \) and \( \nu_G \in (0,1) \). To see this, recall that we have already shown that \( \nu_I \geq \nu_G \). If \( \nu_I = \nu_G = 1 \), then again it must be that \( x = 1 \), which violates the assumption \( x \in (0,1) \). If \( \nu_I = 1 \), or \( \nu_I \in (0,1) \) and \( \nu_G = 0 \), then \( x = 0 \), also a contradiction.

There are three further cases to consider: (1) If \( \frac{P}{P+T} < q_G \), it is not possible that \( x = 1 \) in equilibrium. The reason is the following. If \( x = 1 \), since \( P < q_G (P + T)x \), \( \nu_G = 0 \). It follows that \( x = 0 \) as long as \( \nu_I > 0 \), a contradiction. Thus, \( x \in (0,1) \). Since the judge is using a mixed strategy, it must be the case that he is indifferent between torture and releasing the suspect. That is, it must be that \( \frac{q_G}{q_I} = \frac{(1-q)(L_1+kT)}{q(L_2-kT)} \). As we have already shown, that fact that \( x \in (0,1) \) also implies \( \nu_G \in (0,1) \). Again, the guilty suspect is indifferent between confessing and denial, so \( P = q_G (P + T)x \). Thus, in equilibrium, \( \nu_I^* = 1 \), \( \nu_G^* = \frac{q_I (1-q)(L_1+kT)}{q_G q (L_2-kT)} = \nu \), and \( x^* = \frac{P}{q_G (P + T)} \). (2) If \( q_G = \frac{P}{P+T} > q_I \), then \( P = q_G (P + T) \) and \( P > q_I (P + T) \). The inequality above implies that \( \nu_I = 1 \). If \( x = 1 \), then \( \nu_G \in (0,1) \). Since by assumption \( \frac{q_G}{q_I} > \frac{(1-q)(L_1+kT)}{q_G q (L_2-kT)} \), it must be that \( \nu_G > \nu \). If \( x \in (0,1) \), then \( \nu_G = 1 \), which will result in \( x = 1 \), a contradiction. Therefore, the equilibrium is \( \nu_I^* = 1 \), \( \nu_G^* \in (\nu,1) \), and \( x^* = 1 \). (3) If \( \frac{P}{P+T} > q_G \), then both types of suspect prefer to deny for any \( x \), so \( \nu_I = \nu_G = 1 \). It follows that \( x^* = 1 \).

[2]. Suppose \( \frac{q_G}{q_I} = \frac{(1-q)(L_1+kT)}{q(L_2-kT)} \).

First, if \( x = 0 \), then it must be that \( \nu_G = \nu_I = 1 \). The fact that \( \frac{q_G}{q_I} = \frac{(1-q)(L_1+kT)}{q(L_2-kT)} \) thus implies that the judge is indifferent between torture and release. That is, \( x^* = 0 \), \( \nu_I^* = \nu_G^* = 1 \) constitute an equilibrium. Next, if \( x > 0 \), then it must be the case that to torture is at least as good as to release for the judge, which means \( \frac{q_G}{q_I} \geq \frac{(1-q)(L_1+kT)\nu_I}{q(L_2-kT)\nu_G} \). Since \( \nu_I \geq \nu_G \), this is possible only if \( \nu_I = \nu_G \). But if
\( \nu_I = \nu_G < 1 \), then it must be that \( P \leq q_I(P + T)x \) and \( P < q_G(P + T)x \) (note that \( q_G > q_I \)). This implies \( \nu_G = 0 \) and thus, \( \nu_I = 0 \). This violates the case we set out to consider \((\nu_I^* > 0)\). That means if \( x > 0 \), then \( \nu_G = \nu_I = 1 \). There are thus two cases to consider. (1) Let \( \frac{P}{P+T} \geq q_G \). Then \( P \geq q_G(P + T)x \) and \( P > q_I(P + T)x \) for any \( x \in [0,1] \). Consequently, \( x^* \in [0,1] \) and \( \nu_I^* = \nu_G^* = 1 \) constitutes an equilibrium. (2) Let \( \frac{P}{P+T} < q_G \), then in order for \( \nu_I^* = \nu_G^* = 1 \) to hold, it must be that \( P \geq q_G(P + T)x \), i.e., \( x \in [0, \frac{P}{q_G(P+T)}] \). In other words, \( \nu_I^* = \nu_G^* = 1 \) and \( x^* \in [0, \frac{P}{q_G(P+T)}] \) constitutes an equilibrium.

[3] Suppose \( \frac{q_G}{q_I} < \frac{(1-q)(L_1+kT)}{q(L_2-kT)} \):

In this case, since \( \nu_I \geq \nu_G \), it must be that \( \frac{q_G}{q_I} < \frac{(1-q)\nu_I(L_1+kT)}{q
u_G(L_2-kT)} \). Therefore, \( x^* = 0 \), which implies \( \nu_I^* = \nu_G^* = 1 \).

So far we have assumed that \( \nu_I > 0 \). Suppose now \( \nu_I = 0 \), then since \( \nu_I \geq \nu_G \), it must be the case that \( \nu_I = \nu_G = 0 \). In order that neither type of suspect deviates, requires \( P \leq q_I(P + T)x \) and \( P \leq q_G(P + T)x \). Since \( x \leq 1 \), this is possible, however, only if \( q_I \geq \frac{P}{P+T} \). Moreover, it must be that \( x \in [\frac{P}{q_I(P+T)}, 1] \). If any suspect denies, then (since \( \nu_I = \nu_G = 0 \)) the judge must form an out-of-equilibrium belief \( q^* \). The value of \( q^* \) can be arbitrary under the sequential equilibrium requirement. However, since the social cost of torture under belief \( q^* \) is \((1 - q^*)L_1 + kT\), and the social cost of acquittal is \( q^*L_2 \), to support \( x \in [\frac{P}{q_I(P+T)}, 1] \) as equilibrium, it must be \((1 - q^*)L_1 + kT \leq q^*L_2 \). In other words, \( q^* \in [\frac{L_1+kT}{L_1+L_2}, 1] \).

References


