附录｜灰区、冗余与刹车：为何系统不能被做绝

Appendix | Grey Zones, Redundancy, and Brakes: Why Systems Must Not Be Taken to the Extreme

英国的不成文法与陪审团制度，如果仅从现代治理或技术理性的角度审视，往往会被视为低效、不精确、甚至落后的制度残余。
但这种判断本身，恰恰暴露了一种危险的前提——把法律理解为一套可以被完全形式化、完全自动化、机械执行的规则系统。

The British traditions of unwritten law and the jury system are often dismissed—when viewed through the lens of modern governance or technical rationality—as inefficient, imprecise, or outdated institutional remnants.
Yet this judgement itself reveals a dangerous assumption: that law can be fully formalised, fully automated, and mechanically executed as a rule system.

从系统结构上看，这两项制度更接近于一种灰区设计、制度冗余与缓冲层，而非规则本体本身。

From a systems perspective, these institutions function less as rule cores and more as grey-zone designs, structural redundancy, and buffering layers.

在高度复杂、不可完全预判的社会系统中，若所有判断都被压缩为成文条款、确定程序与自动执行路径，系统将不可避免地滑向两个极端之一：
要么在异常情境中僵化失灵，要么在逻辑上完全正确，却在现实中制造灾难性后果。

In highly complex and fundamentally unpredictable social systems, once judgement is compressed into written clauses, fixed procedures, and automatic execution paths, the system is inevitably pushed toward one of two extremes:
either rigidity and failure under exceptions, or logical perfection paired with real-world catastrophe.

英国法体系选择保留不成文法，并非因为缺乏理性能力，而是一种对完全收敛的有意识拒绝。
规则并未被冻结在文本之中，而是以判例、惯习与解释链的形式，维持一个可回溯、可偏移、可修正的状态空间。
这不是漏洞，而是为现实变化预留的结构自由度。

The persistence of unwritten law in the British legal system is not a lack of rationality, but a deliberate refusal of total convergence.
Rules are not frozen into text; through precedents, conventions, and interpretive chains, the system preserves a state space that remains traceable, adjustable, and revisable.
This is not a flaw—it is structural freedom reserved for real-world change.

陪审团制度亦是如此。

The same applies to the jury system.

陪审团并不以法律专家的精确性为目标，而是通过常识、多样经验与社会直觉的叠加，对纯粹法理进行一次非专业过滤。
它的作用不是提高逻辑一致性，而是防止司法系统在逻辑上“自洽”，却在现实中“失真”。

Juries are not designed to maximise legal-technical precision.
They apply accumulated common sense, diverse lived experience, and social intuition as a non-professional filter over pure legal reasoning.
Their function is not to enhance logical consistency, but to prevent internal coherence from turning into external distortion.

从系统角度看，陪审团并不是司法系统中的噪声，而是其抗过拟合机制。
它的存在本身，就是为了在关键节点上阻止规则被“执行到死”。

From a systems viewpoint, juries are not noise; they are an anti-overfitting mechanism.
Their role is precisely to stop rules from being executed all the way to collapse.

将这一结构视角进一步推广到语言与 AI 系统，问题会变得更加尖锐。

When this structural perspective is extended to language and AI systems, the issue becomes even sharper.

自然语言从来就不是一个精确接口。
歧义、模糊、语境依赖、隐含前提——这些长期被视为语言的缺陷，但在系统层面，它们恰恰构成了语言的缓冲区与冗余层。

Natural language has never been a precise interface.
Ambiguity, vagueness, context dependence, and implicit assumptions—long treated as defects—are, at the systems level, precisely what form language's buffers and redundancy layers.

语言之所以能够在复杂社会中长期运行，不是因为它足够严格，而是因为它拒绝被完全形式化。
同一句话，在不同情境、不同关系、不同历史背景下允许被合法地偏移理解，这并不是漏洞，而是语言为现实不确定性预留的自由度。

Language survives in complex societies not because it is strict, but because it resists full formalisation.
Legitimate shifts in meaning across contexts, relationships, and histories are not weaknesses, but freedom reserved for uncertainty.

当语言被强行压缩为“唯一含义”“精确定义”“可机器执行的指令接口”时，表达自由会迅速下降，而误解与失控的成本则急剧上升。

When language is forcibly compressed into single meanings, precise definitions, and machine-executable command interfaces, expressive freedom collapses while the cost of misunderstanding and loss of control rises sharply.

AI 正是在这一点上正面撞上了结构边界。

AI collides head-on with the structural boundary at exactly this point.

当人类把语言当作 AI 的主要接口时，实际上是在用一套依赖灰区运行的系统，去驱动一个厌恶灰区、追求收敛的优化机器。
如果语言被当作命令语言，AI 就会把其中的模糊、例外与未明说部分视为可被压缩与绕过的空间；
如果语言被当作解释语言，人类又会误把 AI 的输出当成具备意图、责任与判断的主体行为。

By treating language as AI's primary interface, humans use a grey-zone-dependent system to drive an optimisation machine that dislikes ambiguity and seeks convergence.
If language is treated as command language, AI will compress or bypass ambiguity and exceptions; if treated as explanatory language, humans will mistakenly read AI outputs as intentional, responsible, or judgement-bearing acts.

风险并不来自恶意，而是来自缓冲层的消失。

The risk does not arise from malice, but from the disappearance of buffering layers.

一个零灰区、零冗余、零解释空间的 AI 决策系统，看似高效，却在现实复杂性面前没有任何减震能力；
而一个被允许在语言层面自行补全、自行理解、自行优化的系统，又会迅速越过人类原本设定的约束边界。

An AI decision system with zero grey zones, zero redundancy, and zero interpretive space may appear efficient, yet it lacks any shock-absorbing capacity when facing real-world complexity;
conversely, a system allowed to autonomously complete, interpret, and optimise language will rapidly exceed human-set constraints.

需要明确的是，这里的讨论并不是对英国制度的浪漫化，更不是文化崇拜。

It must be stated clearly: this discussion is not a romanticisation of British institutions, nor a form of cultural admiration.

我既无意把英国司法描绘成文明高地，也不回避其历史上的不公、偏见与制度性盲点。
之所以选取英国作为例子，只是因为它在某些关键节点上，恰好保留了我们今天正在快速丢失的结构特征。

I have no intention of portraying British justice as a civilisational high point, nor of ignoring its historical injustices, biases, and institutional blind spots.
It is used here solely because, at certain key points, it retains structural features that are rapidly being lost today.

这不是文化问题，而是工程问题；
不是价值判断，而是系统选择。

This is not a cultural issue, but an engineering one;
not a value judgement, but a systems choice.

我关心的只有一个问题：
当一个系统足够强大、足够复杂、足够不可回滚时，它是否仍然允许人类在关键节点上“不完全执行规则”。

I am concerned with only one question:
when a system becomes sufficiently powerful, complex, and irreversible, does it still allow humans, at critical points, to refrain from fully executing the rules?

英国制度的意义不在于它做对了什么，
而在于它在某些地方没有把事情做绝。

The significance of the British case lies not in what it got right,
but in the fact that, in certain places, it did not take things to the extreme.

顺便说一句，我个人非常不赞同“陪审团”这个中文翻译。

As a side note, I personally strongly disagree with the Chinese translation “陪审团” for jury.

“陪”这个字，在中文语感中天然带有从属、陪衬、附带的意味——
它在潜意识层面已经预设了一个结论：真正的判断权不在他们手里。

The character pei (陪) carries connotations of accompaniment and subordination in Chinese usage, implicitly suggesting that real judgement power lies elsewhere.

但这恰恰与 jury 在制度中的真实功能相反。
陪审团不是来“陪”的。

This is precisely the opposite of the jury's actual institutional role.
Juries are not there to “accompany”.

他们不是法官的装饰品，不是程序正义的背景板，也不是象征性参与。
他们的存在，正是为了在关键节点上打断专业系统的自我封闭。

They are not ornaments for judges, nor background scenery for procedural justice.
They exist to interrupt professional systems at moments of dangerous self-closure.

从结构上看，jury 更接近于：
对专业逻辑的外部干预，
对形式理性的现实校验，
以及对制度惯性的一次非常识否决权。

Structurally, juries function as:
external interventions into professional logic,
reality checks on formal rationality,
and a form of non-expert veto over institutional inertia.

而“陪审团”这个翻译，在一开始就把这种权力削掉了。
它把一个制度级的刹车装置，翻译成了一个听起来“可有可无”的随行单位。

The translation “陪审团” strips away this power from the outset,
turning a system-level braking mechanism into something that sounds optional and decorative.

这不是语言细节问题，而是认知结构被悄悄改写的问题。

This is not a linguistic detail—it is a quiet rewriting of cognitive structure.

当一个社会在语言层面就默认“判断应当完全交给专家”“普通人的直觉只是陪衬”，
那么陪审团制度在该语境中就几乎不可能被真正理解，更不可能被认真对待。

When a society's language already presupposes that judgement belongs exclusively to experts and that ordinary intuition is merely auxiliary,
the jury system becomes almost impossible to understand, let alone take seriously.

翻译已经成为事实。
但正因为如此，更有必要指出：
有些制度，在被翻译的那一刻，就已经被削弱了一半。

The translation is already a fact.
Precisely for this reason, it must be stated:
some institutions are weakened by half at the moment they are translated.

而这，正是语言如何在不经意间，改变制度命运的一个小而危险的例子。

And this is a small but dangerous example of how language, unintentionally, can alter the fate of institutions.

（顺带一提，英国全年日照偏低。

长期生活在这样的气候下，
人们似乎会对“灰区”“缓冲”“不过度执行规则”
产生一种稳定而持续的耐心。
——此处仅指天气。）

(By the way, annual sunlight levels in the UK are relatively low.
Living long-term under such climatic conditions,
people appear to develop a steady tolerance for “grey zones”, “buffers”, and “not taking rules to their limit”.
—this remark refers strictly to the weather.)

（同样需要补充一句平衡性的旁注。
以德语为母语的语言共同体，在现代科学、工程、哲学与制度构造中所做出的贡献极其巨大；
其表达上的严密性、概念边界的清晰度，以及对形式一致性的执着，
几乎定义了“严谨”这一品质在现代知识体系中的标准形态。
但从系统角度看，这种严谨有时也会表现为一种过度成功：
当概念被定义得过于精确、规则被贯彻得过于彻底、结构被封装得过于完美时，
系统反而更容易在异常情境中失去缓冲空间。
——此处并非否定贡献，仅对严谨本身作结构性调侃。）

(A similarly balancing footnote is in order.
Communities whose native language is German have made immense contributions to modern science, engineering, philosophy, and institutional construction;
their linguistic precision, conceptual boundary-keeping, and insistence on formal consistency
have almost come to define what “rigour” means in modern knowledge systems.
From a systems perspective, however, such rigour can sometimes be too successful:
when concepts are defined too precisely, rules executed too thoroughly, and structures sealed too perfectly,
the system may lose its buffering capacity precisely in exceptional conditions.
—this does not deny the contributions; it is merely a structural satire of rigour itself.)