Quaestio facti. Revista Internacional sobre Razonamiento Probatorio
Quaestio facti. International Journal on Evidential Legal Reasoning
Sección: Conjeturas y refutaciones
2026 I 11 pp. 225-232
Madrid, 2026
DOI: 10.33115/udg_bib/qf.i11.23292
Marcial Pons Ediciones Jurídicas y Sociales

ISSN: 2604-6202
Recibido: 28/05/2026 | Aceptado: 12/06/20266 | Publicado *online*: 30/06/2026

WHAT GOES AROUND COMES AROUND: WHY THE CRITIQUE OF EXPERT-LED INTEGRAL BAYESIAN MODELLING ALSO UNDERMINES JUDGE-LED BAYESIAN MODELLING*

Bojan Spaić
University of Belgrade Faculty of Law

ABSTRACT: This article offers an internal refutation of Anne Ruth Mackor's defence of court-led integral Bayesian modelling in criminal adjudication. Mackor usefully distinguishes between the now well-established Bayesian treatment of individual items of forensic evidence and the more controversial Bayesian modelling of criminal cases as a whole. She also convincingly argues that expert-led whole-case modelling is objectionable because the expert must select hypotheses, evidence, dependencies, and probabilistic inputs in a manner that effectively places the expert in the judge's chair. Yet she maintains that this objection applies with less force when whole-case Bayesian modelling is performed by judges themselves, especially as an internal "means of inspection" or "sharpening" device and with the support of a forensic adviser. This article argues that the distinction fails. The decisive objection to integral Bayesian modelling does not arise merely from the formal identity of the modeller. It lies in the modelling choices themselves being adjudicatively decisive. Those choices remain no less decisive when made by judges, and the involvement of the forensic adviser reintroduces, in a less visible form, the very institutional concern Mackor's critique seeks to avoid.

KEYWORDS: evidential reasoning; legal probabilism; legal abductivism; Anne Ruth Mackor.

* This paper is a contribution to the 2026 annual strategic project of the Faculty of Law University of Belgrade "Problems of creation, interpretation and application of law."

1. INTRODUCTION

Anne Ruth Mackor's article on Bayesian modelling of criminal cases as a whole is a careful intervention in a debate that is often marred by needless disagreement and conceptual confusion. Its merit primarily lies in its insistence that one must distinguish between different uses of Bayesian reasoning in criminal adjudication. Mackor emphasises that Bayesian methods are widely accepted for the likelihood-ratio assessment of individual items of forensic evidence, and that the genuinely controversial question concerns the integral modelling of the criminal case as a whole. She therefore rightly criticises judicial reactions that blur the difference between objections to one particular use of Bayes and objections to Bayesian reasoning as such (Vlek *et al.*, 2016, p. 285).

The article's central claim, however, is unstable. Mackor argues that Dutch lower courts were right to reject expert-led integral Bayesian modelling because such modelling requires the expert to make a series of choices concerning hypotheses, evidence, interdependencies, and probabilistic inputs that are not merely technical but adjudicatively decisive. On her account, the problem with the expert-led model is not that the Bayesian method is itself unsound. The problem is that in constructing the model, the expert effectively serves as the judge. (Mackor, 2026, pp. 364-365) The expert selects, structures, and evaluates. That, Mackor claims, is institutionally objectionable from a rule-of-law perspective.

Mackor then takes a judicial turn. She contends that the same objection does not hold, or does not hold in the same way, where judges themselves undertake integral Bayesian modelling, especially in the deliberation chamber, and especially where such modelling functions as an internal "means of inspection" after a more traditional evidential assessment has already been formed. She goes on to develop this possibility in more concrete terms by proposing that judges may, with the support of a forensic adviser, construct Bayesian networks to compare their traditional judgment with the outcome of their own Bayesian analysis, thereby sharpening their evidential decision (Mackor, 2026, pp. 371-372).

This article argues that Mackor's move from rejecting expert-led integral Bayesian modelling to defending judge-led integral Bayesian modelling does not succeed. The objections she raises against expert-led modelling cannot be limited to the expert-led version alone, because the very choices that make expert-led modelling objectionable remain present and decisive in judicial modelling as well. If the vice of expert-led modelling is that the evidential architecture of the case is structured through contest-

able and outcome-sensitive choices, then changing the identity of the modeller does not by itself neutralise that vice. The problem lies not simply in who models the case, but in what it means to model a criminal case as a whole.

The argument proceeds in four steps. (1) I reconstruct the relevant points of Mackor's position. (2) I show that the objection she raises against expert-led integral modelling applies, mutatis mutandis, to judge-led modelling as well. (3) I then argue that the role she reserves for the forensic adviser reintroduces, in internalised form, the very problem of expert displacement that grounds her critique of the expert-led model. (4) Finally, I argue that her proposal to treat judicial Bayesian modelling as a "means of inspection" or "double-check" is methodologically unstable. My claim is modest. I do not argue that Bayesian reasoning is always inappropriate in law, nor that integral Bayesian modelling could never, under any circumstances, prove superior to its competitors. I argue only that Mackor's own premises do not justify the distinction on which her judicial alternative depends (Di Bello, 2019).

2. STRENGTHS AND WEAKNESSES OF JUDICIAL BAYESIANISM

An important aspect of Mackor's paper is to be found in her refusal to attack Bayesian methods indiscriminately. She repeatedly stresses that the Bayesian model is widely accepted as a method for estimating the likelihood ratio or probative force of forensic evidence, and that Dutch case law eventually clarified that the real controversy begins only when Bayesian reasoning is extended from individual forensic evidence to the modelling of a criminal case as a whole. This distinction is correct and valuable. It prevents a familiar overreaction in which scepticism about whole-case probabilistic modelling is transformed into scepticism about all Bayesian reasoning in adjudication (Vlek *et al.*, 2016, p. 288).

Even more important is Mackor's diagnosis of what is actually wrong with expert-led integral modelling. She does not say, at least in her own voice, that Bayes is too mathematical, too foreign to adjudication, or too controversial simply because probabilities are involved. On the contrary, she explicitly argues that "the method itself is not disputed among forensic experts" and that the real difficulty arises in "the selection and evaluation of the hypotheses, the evidence and their interdependencies." (Mackor, 2026, p. 371) On that basis, she explains why an expert who undertakes an integral analysis of a criminal case as a whole risks taking the place of the court: because the expert's selection and evaluation will not necessarily correspond to that of the judge and may therefore amount to placing "the expert on the court's chair."

This is one of the main points in the article. Whole-case Bayesian modelling is an exercise in constructing the evidential architecture of the case. Someone must formulate the hypotheses, determine which evidence matters, decide how pieces of evidence are related, and assign or approximate the relevant probabilistic force (Hunt & Mostyn, 2020). Those decisions are not neutral. They shape the path to the result.

Mackor is therefore right to see the expert-led model not as a harmless aid but as a possible displacement of adjudicative responsibility.

3. WHY THE SAME OBJECTION APPLIES TO JUDGE-LED MODELLING

When the strongest points of Macor's paper are taken seriously, the weakness of her alternative becomes apparent. The objection to expert-led modelling is that whole-case Bayesian analysis requires the making of adjudicatively decisive choices. But those choices do not disappear when judges themselves perform the modelling. They remain exactly the same choices: the formulation of hypotheses, the inclusion and exclusion of evidential items, the characterisation of dependence or independence, and the estimation or positing of priors and likelihoods where statistical grounding is weak or absent.[1] The only thing that changes is the institutional identity of the person who makes them.

This difficulty is central. Either those modelling choices are adjudicatively decisive, or they are not. If they are not, then Mackor's institutional objection to expert-led modelling is significantly weakened. If they are, judicial ownership cannot, by itself, answer the objection. The modelling process still structures the evidential field in a way that bears on the outcome. A change in authorship does not alter the character of the act (Allen, 2019, p. 15).

Mackor's own account makes this plain. She acknowledges that judges are not experts in Bayesian modelling and notes empirical work suggesting that judges often misunderstand Bayesian reports. She also emphasises that, outside well-supported forensic contexts, priors and many non-forensic evidential estimates cannot be grounded in reliable statistics and must therefore be estimated by the court. In other words, the judicial version retains the modelling choices that troubled Mackor in the expert-led context, and assigns them to actors who, by her own account, are often technically ill-equipped to manage them without assistance and must frequently do so on the basis of contestable, non-statistical judgments (Mackor, 2026, pp. 369-370).

The problem is therefore structural. Whole-case Bayesian modelling is a constructive activity. The case must be represented under descriptions. Some propositions must be elevated to the status of hypotheses, others to evidence, and others may drop out of the model entirely. Dependencies must be posited at one level rather than another. Some uncertainty must be stabilised into qualitative or numerical ranges (Pardo & Allen, 2008). None of this becomes less contestable merely because the modeller is now a judge rather than an external expert.

---

[1] Authors who argue in favor of Bayesian networks modeling emphasize the characteric dificulty of the procedure: "Representing a scenario is nonetheless still a difficult task due to all the probabilities that need to be elicited" (Vlek *et al.*, 2016, p. 318)

Indeed, one might say that Mackor's own reasoning generates a dilemma. If the decisive concern about expert-led modelling is that it structures the adjudicative field through contestable choices, then judge-led modelling inherits that concern because it performs the same structuring operation. If, on the other hand, judicial authorship is said to cure the problem simply because judges are entitled to make adjudicative choices, then Mackor's critique of expert-led modelling becomes oddly incomplete, since the mere existence of contestable modelling choices can no longer serve as the decisive reason for resistance. In that case, the problem would have to be redescribed more narrowly as one of institutional authority alone. But Mackor's own article says more than that. It emphasises not only who decides but also the fragility, discretionary character, and outcome sensitivity of what must be decided.

For that reason, the judicial alternative cannot be vindicated simply by insisting that judges, unlike experts, are entitled to decide cases. Of course they are. The question is whether modelling the case as a whole in Bayesian form introduces a set of structurally problematic operations into adjudication that judicial entitlement alone cannot purify.

4. THE FORENSIC ADVISER AS HIDDEN CO-AUTHOR

The instability of Mackor's position becomes clearer once the forensic adviser enters the scene. Mackor expressly envisages the possibility of judicial Bayesian modelling with the assistance of a forensic adviser. At one point, she notes that judges may be assisted in Bayesian modelling by the court's forensic adviser, while conceding that even then, "there is a risk that the forensic adviser's modelling proposals will exert undue influence on the judge's evidential decision" {Mackor, 2026 #36748@369}. Later, when describing her group's research agenda, she states that the relevant analysis should involve judges setting up a Bayesian network "with the help of a forensic advisor" {Mackor, 2026 #36748@378}.

This concession is more serious than the article allows. The forensic adviser is introduced because judges are not Bayesian experts and because modelling a case as a whole is technically and conceptually demanding. But the more necessary the adviser becomes, the more the original institutional objection returns. If expert-led modelling was objectionable because the expert shapes the evidential architecture of the case (Tribe, 1971, pp. 1381-1386), it is hard to see why the forensic adviser's analogous shaping role becomes innocuous merely because it is exercised inside the judicial process and under the label of assistance.

The adviser may not be the formal author of the final judgment, but he may still be a co-author of the model that structures the case. He may influence which nodes are created, how they are related, how alternative hypotheses are formulated, whether dependency is modelled at one level or another, and how verbal estimates are stabilised into ranges or numbers. Those are not peripheral matters. They are precisely

the matters Mackor had identified as decisive in explaining why expert-led integral Bayesian modelling threatens to put the expert in the judge's chair.

Nor is it enough to say that the judge remains free to reject the adviser's proposals. The same can be said of external expert reports in general. A court is always formally free to reject them. Mackor's objection to the expert-led model, however, was not defeated by that formal freedom. It was that the expert's modelling choices shape the evidential field itself. That point applies equally here. If the adviser is involved in building the network, the displacement concern has not disappeared but has been internalised.

There is, in fact, a further difficulty. In the expert-led model, the court and the parties at least encounter the model as the work of a separate actor and can criticise it as such. In the judicial version, by contrast, the adviser's technical influence risks becoming less visible as it is absorbed into the court's own deliberative process. The danger is not that the adviser openly replaces the judge. It is that expert influence becomes harder to identify because it is folded into what now appears as judicial self-reflection. The expert is no longer visibly in the judge's chair. He is standing behind it, arranging the model through which the court comes to understand its own reasoning.

5. THE INSTABILITY OF THE "MEANS OF INSPECTION"

Mackor's most sophisticated response to these worries is her suggestion that judge-led whole-case Bayesian modelling need not function as the primary basis of adjudication. She proposes, instead, that judges may first reach a preliminary evidential assessment in a "traditional" manner and then use integral Bayesian analysis as a "means of inspection," an internal check, or a "sharpening function." {Mackor, 2026 #36748@371}. Later, she develops this further by describing her group's investigation into judges who compare their traditional judgments with the results of their own Bayesian analyses {Mackor, 2026 #36748@377-381}.

If the Bayesian model merely confirms the traditional assessment, then its independent normative significance is limited. It adds discipline, perhaps, or rhetorical reassurance, but not much more. Yet Mackor wants it to reveal hidden assumptions, expose errors in reasoning, force attention to dependence relations and evidential levels, and improve the quality of judicial evidence evaluation {Mackor, 2026 #36748@367}.

But if the model can do that meaningfully, then it is not merely a check. It becomes capable of revising, unsettling, or competing with the traditional assessment. Once the Bayesian model acquires that capacity, the relation between the first-stage assessment and the second-stage model becomes methodologically unclear. If the two converge, the second risks redundancy. If they diverge, which governs? Mackor suggests that divergence should provoke critical comparison, but that is not yet a cri-

terion of resolution. A tool that is serious enough to change the court's view cannot be described as merely auxiliary without further explanation.

The difficulty becomes even sharper when Mackor turns to Bayesian networks. She stresses that the primary value of constructing a network lies not in the final numerical output but in the joint activity of structuring, discussing, and refining assumptions {Mackor, 2026 #36748@378}. At the same time, however, she notes that software can perform highly complex calculations once priors and evidential values are entered, and she defends quantification on the grounds that it can help judges grasp the cumulative effect of multiple likelihood ratios and priors. She also argues that quantification may encourage greater uniformity in the use of verbal probability judgments {Mackor, 2026 #36748@381}.

These are not negligible considerations. But they do not resolve the deeper instability. The more operational the model becomes, the more it threatens to become a rival decision framework rather than a mere reflective aid (Tribe, 1971, p. 1971). The sharper the "double-check," the less plausible it is to insist that it is only a check. Mackor's proposal thus oscillates between two poles. Either judicial Bayesian modelling is weak enough to remain subordinate, in which case its added value is modest (Hunt & Mostyn, 2020, p. 16). Or it is strong enough to alter the court's evidential outlook, in which case it assumes a decisional significance that her institutional framework does not adequately contain.

6. CONCLUSION

Mackor's article is at its best when it insists that the principal difficulty of whole-case Bayesian modelling lies not in Bayes as such, but in the construction of a probabilistic representation of the criminal case as a whole. Her most persuasive point is that expert-led integral modelling is objectionable because the modeller must select hypotheses, evidence, dependencies, and probabilistic inputs in a manner that risks displacing judicial responsibility. That insight is both important and correct.

But the insight cannot be confined to the expert-led model alone. The same modelling choices remain necessary when judges themselves undertake integral Bayesian modelling. Those choices do not become less contestable, less theory-laden, or less outcome-sensitive merely because their author is now the court. Nor does the involvement of the forensic adviser neutralise the problem. It reintroduces, in a less visible form, the very concern about expert influence that grounds Mackor's original critique. Finally, the attempt to stabilise the judicial alternative by redescribing it as a "means of inspection" or "double-check" fails to dispel the difficulty. The model is either too weak to matter greatly or strong enough to become a rival framework of adjudication.

The present argument is deliberately limited. It does not deny that Bayesian methods may have an important place in criminal adjudication (Di Bello, 2019, p. 6). It

does not deny that integral Bayesian modelling might, under some conditions, prove superior to its alternatives. It argues only that Mackor has not shown that judge-led integral Bayesian modelling escapes the very objection on which her critique of expert-led modelling depends. On her own premises, the objection returns. The same objection arises twice.

LITERATURE

Allen, R. J. (2019). Standards of Proof and The Limits of Legal Analysis. *Diritto & questioni pubbliche*, *XIX*, 7–25.

Di Bello, M. (2019). Plausibility and probability in juridical proof. *The International Journal of Evidence & Proof*, *23*(1-2), 161–167. *https://doi.org/10.1177/1365712718815355*

Hunt, I., & Mostyn, J. (2020). Probability reasoning in judicial fact-finding. *The International Journal of Evidence & Proof*, *24*(1), 75–94. *https://doi.org/10.1177/1365712719875753*

Mackor, A. R. (2026). Bayesian Modelling of Criminal Cases as a Whole. *Quaestio facti. Revista Internacional sobre Razonamiento Probatorio*, *10 Quaestio facti. International Journal on Evidential Legal Reasoning*, 363–384. *https://doi.org/10.33115/udg_bib/qf.i10.23299*

Pardo, M. S., & Allen, R. J. (2008). Juridical Proof and the Best Explanation. *Law and Philosophy*, *27*(3), 223–268.

Tribe, L. H. (1971). Trial by Mathematics: Precision and Ritual in the Legal Process. *Harvard Law Review*, *84*(6), 1329. *https://doi.org/10.2307/1339610*

Vlek, C. S., Prakken, H., Renooij, S., & Verheij, B. (2016). A method for explaining Bayesian networks for legal evidence with scenarios. *Artificial Intelligence and Law*, *24*(3), 285–324. *https://doi.org/10.1007/s10506-016-9183-4*