Relative Plausibility and a Prescriptive Theory of Evidence Assessment

Eivind Kolflaath*

I. Preliminaries: Background and motivation

This commentary on Allen and Pardo's extensive article will concentrate on the authors' presentation (pp. 13–35) of their own theory. Unlike them, I will discuss the theory of relative plausibility mainly as a possible starting point for a prescriptive theory of evidence assessment. Generally, I tend to think of descriptive theories of evidence not as the end but rather as a step toward a theory that provides explicit recommendations on fact-finding. My perspective is that of an evidence scholar who regularly participates in organized discussions with Norwegian judges,[1] both as part of the mandatory introductory program for newly appointed judges and in seminars organized by the courts.[2] Engaging in these practically oriented settings has made me interested in how evidence theory could make a difference in practice.

As will be argued in section II, the theory of relative plausibility has great potential in this respect. Some additional background may be needed, however, to justify the more general supposition underlying this commentary that recommendations on evidence assessment could actually influence courts' practices. While Allen and Pardo's article reflects the legal system (or various legal systems) in the United States, my point of view is informed by the Scandinavian legal context. Today, none of the Scandinavian countries

* Professor of jurisprudence and philosophy at the University of Bergen, Norway. I want to thank Ron Allen and Mike Pardo for inviting me to comment on their article, and Magne Strandberg for helpful comments on a draft.

[1] In this commentary the noun "judge" always means professional judge except when I use the term "lay judge." The word "fact-finder" is used when I talk about both professional judges and lay judges.

[2] The introduction program focuses mostly on practical skills and includes group sessions where participants perform deliberations by role playing after watching a video showing a condensed presentation of evidence in a (fictitious) criminal trial. Immediately after the role playing, we discuss what happened during the deliberation and, more generally, advantages and disadvantages of different ways of conducting deliberations.

has a jury system.[3] In serious criminal cases, the defendant's guilt or innocence is decided by a mixed court of professional judges and lay judges, deliberating together. Consequently, there is nothing similar to jury instructions in open court. Instead, the professional judges are supposed to provide the lay judges with the necessary legal guidance (on legal elements, standard of proof, sentencing practice, etc.) during the deliberation. The judge who chairs the trial is expected to facilitate the deliberation, but there are no legal rules that dictate how this should be done. Nor is there any broad consensus among judges on this point, as opinions differ about what are the main purposes of the deliberation, beyond reaching a verdict. My view (which is held by many Norwegian judges) is that the deliberation stage should be an arena for collective and systematic evidence evaluation, as a kind of quality assurance of individual assessments.[4] Accordingly, this part of the trial should primarily be *investigative* rather than an arena where the members of the court spend a lot of time defending their own opinions and trying to convince those who may disagree. The purpose of such a collective evaluation is not to reach unanimity (which is not required in Scandinavian trials), but to ensure that the evidence and its possible interpretations have been critically examined by the court. To fulfil this purpose the facilitator needs some method for structuring the assessment process. Obviously, such a method is one of the things a prescriptive theory of evidence assessment may offer. Thus, if judges are receptive to proposals on how to structure the deliberation, a prescriptive theory could actually make a difference in practice.

Allen and Pardo emphasize the *explanatory* and *comparative* dimensions of the theory of relative plausibility as "two fundamental ways in which the theory differs from the conventional probabilistic account" (p. 13). I will comment on both, starting with the explanatory dimension, which I take to be the most basic.

II. The explanatory dimension

A prescriptive theory of evidence assessment should contribute to factual accuracy in real trials. Such a contribution can happen only if fact-finders are able to apply the method and evaluation criteria recommended by the theory. Hence, the theory ought to be calibrated to real fact-finders – to what they are and what they do.

Fact-finders perform their task as they do because it comes naturally to them. The more a prescriptive theory strays from what is natural to fact-finders, the more likely it is they will never adopt the method or that after a while they will revert to old habits. A method that presupposes knowledge of a kind that fact-finders are unlikely to have, or a mode of thinking alien to most fact-finders, will make little difference to the courts' practices. So, rather than trying to replace current practices with something completely different, a

[3] Denmark and Norway abolished their jury systems in 2008 and 2018, respectively. In Sweden there is the marginal exception that a jury is used in some trials related to freedom of speech (approx. 200 jury trials over a period of 10 years, according to NOU 2011: 13, Når sant skal skrives: Juryutvalget, 55).
[4] I do not know whether this is also *the majority view* among Norwegian judges, but undoubtedly judges have become increasingly aware about the importance of their role as facilitators of deliberations.

prescriptive theory should build upon and aim to refine them. Adjustments are the best we can hope for. Thus, a prescriptive theory should take as its point of departure a plausible descriptive theory that tells us what fact-finders' habits are, and then build upon some of these – unless all their habits can be dismissed as mere obstacles to factual accuracy. Another point in the same vein is that although recommendations to fact-finders may be derived from theoretically advanced considerations, the recommendations themselves must be kept simple if they are to have any chance of being adopted.

I take *naturalness* and *simplicity*, therefore, to be necessary for the success of a prescriptive theory of evidence assessment.[5] Now, one of things that comes naturally to fact-finders is to search for and evaluate explanations of the evidence. This explanatory dimension of evidence assessment, which I take to be the core of Allen and Pardo's theory, was clearly demonstrated to me during my observation of the deliberations in 105 criminal cases in five of the six Norwegian courts of appeal around ten years ago. As a conceptual tool for analyzing discussions that occurred behind closed doors in the deliberation room, the notion of explanation was definitely more useful than a Bayesian framework could be. But that is just to be expected. It would be miraculous if people unfamiliar with Bayes' rule used it in evaluating legal evidence. The real battle between explanatory and probabilistic approaches in evidence theory is not over the empirical question about what fact-finders do. It relates instead to interpretative and normative questions about which theory provides the most suitable approach to evidence assessment, where "suitability" includes several criteria such as rationality and feasibility, as well as compliance with legal rules, goals, and policies.

Anyway, I take the explanatory dimension to be indispensable in legal fact-finding. When drawing inferences from pieces of evidence to facts of the past, there is typically an explanatory relation in the opposite direction (i.e., past facts explaining items of evidence), and probably most such inferences are made exactly *because* of this explanatory relation. Moreover, stories are expected to be not mere chronologies but causal chains, so that one element of the story explains the next. And at the global level, it is pertinent to think of the parties' versions of the facts as competing overall explanations of the total body of evidence, somewhat similar to our conception of scientific hypotheses as explanations of empirical data. Even if we would prefer to think of evidence assessment in Bayesian terms, explanatory considerations are often needed if the estimations of the various likelihood ratios are expected to be more than just wild guesses. Hence, the notion of explanation should be central to any prescriptive theory of evidence assessment.

Certainly, the theory of relative plausibility is not the only descriptive theory to consider in developing a prescriptive theory. Pennington and Hastie's story model is obviously a candidate. Allen and Pardo comment on the differences between the two theories (p. 17, n. 86; p. 35). Unlike the first versions of the theory of relative plausibility, the authors

[5] Similarly Floris J. Bex, Arguments, Stories and Criminal Evidence: A Formal Hybrid Theory (2011) 8 regarding naturalness ("the theory should be natural in that it employs concepts that are natural to an everyday reasoner such as a crime analyst or a judge").

now talk about "explanations" instead of "stories" because stories "are only one form of explanations pertinent to judicial proof" (p. 35). They mention several examples of civil litigation (anti-trust litigation, no-fault divorce, and "much contract litigation") in which "[c]ompeting explanations may be advanced […] but not necessarily in the form of the normal meaning of 'stories'" (p. 35). But the notion of stories is too limiting for another reason, one relevant to criminal cases and tort cases as well. In situations with conflicting evidence, we should not expect that stories explain every part of the evidence. Still, a conscientious review of the evidence requires that no part of it be left unexplained – all evidence should have been explained by fact-finders in ways consistent with their conclusions. While explanations of the various parts of the evidence must of course be mutually consistent, we should not demand a *unitary* explanation of the evidence as a whole. For example, a possible explanation of a witness's testimony may simply be that the witness is mistaken. Such an explanation does not fit into a story, at least not if the word "story" means what it usually means in the theory of evidence.[6] For this reason, too, the prescriptive theory should be built on the notion of explanation rather than on the notion of stories.

So, the notion of explanation is more flexible than that of stories, by being applicable in all kinds of trials (as indicated by Allen and Pardo), and by allowing for a broader range of explanations. Even if assessments must be holistic in some ways, a prescriptive theory should encourage fact-finders to dwell upon individual items of evidence and to contemplate on how they can be explained. More generally, the idea that evidence assessment is about explaining the evidence puts the evidence more into the foreground than if stories are taken as the conceptual starting point.

In addition to the story model, one could consider the theory of explanatory coherence developed by Paul Thagard. Like Allen and Pardo's theory, the theory of explanatory coherence is oriented more directly toward the explanatory dimension than is the story model. But even if we grant that the theory provides a fairly accurate description of what fact-finders do, I find the concept of explanatory coherence too technical and complex to be used in recommendations to fact-finders. Generally speaking, recommendations on fact-finding should be about something that real fact-finders can *choose* to do. To the extent that fact-finders actually reason in accordance with the seven principles describing the symmetrical relation of explanatory coherence,[7] they do so intuitively and not as a matter of choice. And if they do not, trying to inculcate them with those principles would probably be a waste of time.

Compared to Thagard's notion of explanatory coherence, the notion of explanation relied upon by Allen and Pardo is more straightforward. This is the conception we are all familiar with – that is, the non-symmetrical relation between something explaining and

[6] I am not sure, though, whether such an explanation of the testimony fits Allen and Pardo's notion of explanations as "versions of the facts," either. If it does not, I think their notion will not bring out the full potential of an explanatory approach to evidence assessment, either.

[7] See, for example, Paul Thagard, Causal Inference in Legal Decision Making: Explanatory Coherence vs. Bayesian Networks, 18 Applied Artificial Intelligence 231 (2004) 234–235.

something being explained. In short, to explain is to answer a "why" question.[8] Why did the witness testify as she did? Maybe because what she reported is what actually happened. Or maybe because she was mistaken. If so, *why* was she mistaken? And so on. Or at the global level: Why do we have this body of evidence? Our intuitive notion of explanation seems well suited for exploring evidence and hypotheses in legal fact-finding.

Admittedly, the term "explanation" lacks theoretical precision. It is undefined in the theory of relative plausibility, as in Thagard's theory. And its meaning may vary somewhat: When we say that some particular piece of evidence is "explained" by some possible fact of the past, or that one element of a story "explains" the next, we are usually talking about causal explanations, albeit in a wide sense that allows intentions, among other things, to be causes. But at the global level, as when Allen and Pardo say that "the central fact-finding task […] is to determine whether potential explanations of the evidence and events satisfy the applicable standard of proof" (p. 13), the term is apparently used in a looser sense. While some evidence theorists will surely regard this vagueness or ambiguity as a weakness or defect, I trust that fact-finders possess an understanding of what explanations are, and that this understanding is sufficiently clear to direct their assessments in ways that could improve factual accuracy. Again: To explain is to answer a "why" question.

Allen and Pardo say that "both the explanatory and probabilistic accounts focus on the same end: the likelihood of disputed facts" (p. 13). One important implication of this is that evaluation criteria that are not genuinely epistemic (i.e., truth conductive) should not carry over to the prescriptive theory. Generally, explanatory criteria includes "consistency, coherence, fit with background knowledge, simplicity, absence of gaps, and the number of unlikely assumptions that need to be made" (p. 15, see also p. 13, n. 3). While some of these criteria (e.g., consistency and fit with background knowledge) are obviously epistemic in legal fact-finding, I am skeptical about taking simplicity as an indication of truth in this context. A rather trivial example is that at the outset, a "disjunctive explanation" (pp. 25–29) having explanations A and B as its disjuncts is more likely than either A and B on their own, despite the disjunction being more complex than any of its disjuncts. Moreover, it is necessary to clarify what "coherence" and "absence of gaps" mean in this particular context in order to assess the epistemic significance of these criteria in legal fact-finding. It is also necessary to discuss possible tensions between various criteria, for instance whether coherence typically requires a more detailed explanation at the expense of simplicity.

Peter Lipton has argued that in science, inference to the best explanation should be inference to the "loveliest" explanation rather than to the likeliest explanation. In short, the difference between the two is that "[l]ikeliness speaks of truth; loveliness of potential understanding."[9] To be sure, in science the most likely explanation is sometimes too

[8] Although in general, the word "explain" is also used in a different way, as when someone "explains" the meaning of a word.
[9] Peter Lipton, Inference to the Best Explanation (2d ed. 2004) 59.

trivial to increase understanding. In legal fact-finding, however, versions of the facts are never too trivial if they relate to elements of the relevant substantive rule, either by instantiating some or all the elements or by implying that at least one element is not met. Fact-finders should not be seduced by loveliness, and a prescriptive theory should fight such inclinations. Certainly, discussions within the philosophy of science may be a fruitful source of inspiration for a theory of evidence assessment, yet science and legal fact-finding have different goals. While in science there are arguably "explanatory virtues" that are non-epistemic,[10] in legal fact-finding non-epistemic evaluation criteria (other than legal relevance) should be discarded. In the legal context, the best explanation is the most likely one, although it may be insufficient if the standard of proof is higher than preponderance of evidence, as Allen and Pardo have pointed out (p. 22). If there is to be a prescriptive theory derived from the theory of relative plausibility, it is necessary to discuss which explanatory criteria are genuinely epistemic in the particular context of legal fact-finding. In such a discussion one should not take for granted that every criterion that is epistemic in the scientific context is also epistemic in legal fact-finding.

III. The comparative dimension

According to Allen and Pardo, the second distinguishing feature of the theory of relative plausibility is that "the explanatory account is inherently comparative" (p. 13). The comparative dimension is manifest in their interpretation of the standards of proof in civil trials. Here, fact-finders must compare the respective degrees of plausibility they ascribe to the competing versions of the facts, to decide whether the standard is met. The comparative dimension is evident not only in the authors' version of the preponderance standard but also in their account of the "clear and convincing" standard (pp. 31–32).[11] In both, the word "plausibility" denotes a gradual notion, and it is required that the fact-finder decide whether the explanation supporting the party having the burden of proof is either at least slightly better (the preponderance standard) than or considerably better (the "clear and convincing" standard) than the explanation supporting the other party. Such decisions obviously necessitate a comparison.

In criminal trials, however, it is less clear what the comparative dimension amounts to. According to Allen and Pardo, "[t]he BARD standard is met when there is a plausible explanation consistent with guilt and no plausible explanation consistent with innocence" (p. 30). Like their accounts of the standards in civil trials, and unlike some of Allen's

[10] See, e.g., Larry Laudan, The Epistemic, the Cognitive, and the Social, in Science, Values, and Objectivity 14 (Peter Machamer and Gereon Walters ed., 2004) and Frank Cabrera, Can there be a Bayesian explanationlism? On the prospects of a productive partnership, 194 Synthese 1245 (2017).

[11] Confusion on this point may have arisen because the authors sometimes use the phrase "inference to the best explanation." No doubt, the proof process outlined in the theory of relative plausibility has much in common with abduction, but talk about "inference to the best explanation" is suitable only in connection with the preponderance standard, and is therefore better avoided in a general presentation of the theory.

early formulations of the "beyond a reasonable doubt" standard,[12] the quoted statement is two-sided in that it involves both explanations supporting one party and explanations supporting the other. But this two-sidedness does not in itself imply a comparison. And at least a comparison of the kind involved in the authors' accounts of civil trial standards is absent in their interpretation of the BARD standard. The latter boils down to questions of whether an explanation is *plausible or not*, where plausibility serve as a "yes" / "no" criterion. I take the authors' remarks on the BARD standard (p. 7 in particular) to suggest that fact-finders should decide first whether there is a plausible explanation consistent with guilt, and then – if they take this to be the case – decide whether there is also a plausible explanation consistent with innocence. Thus, Allen and Pardo seem to presuppose that the first decision can be made independently of the second. And although the *need* for the second decision depends upon the first decision, whether or not there *is* a plausible explanation consistent with innocence would not, as far as I can tell, depend upon whether there is a plausible explanation consistent with guilt.[13] Thus, both decisions could seemingly be made independently. If so, recommendations to fact-finders in criminal trials should not talk about a comparison, as this would just create confusion about what they should do.

Allen and Pardo maintain, however, that "before BARD is met there must be some comparison with possible defense explanation (in the sense of checking whether there is any plausible explanation consistent with innocence […])" (p. 32). Here it seems that the word "comparison" is used to describe something essentially different from the comparison involved in the other standards. To avoid confusion regarding the comparative dimension, we should distinguish carefully between

(1) whether the fact-finders need to compare *the degrees of overall plausibility* of competing explanations (as implied by the authors' interpretation of the standards of proof in civil trials),
(2) whether fact-finders would be unable to decide whether particular explanations are *plausible or not* without looking at the *content* of explanations supporting the other party, and
(3) whether competing explanations, when taken together, need not be *exhaustive* (i.e., cover all scenarios).

Maybe the alleged comparison in criminal trials has something to do with (2) or (3), but in any case, the comparative dimension seems less significant here, if relevant at all. While I agree with Allen and Pardo that it is not limited to the preponderance standard, I think the authors have yet to show how it is relevant in criminal trials, given their interpretation of the BARD standard.[14] Pointing out this divergence or ambiguity is not to

[12] See Ronald J. Allen, The Nature of Juridical Proof, 13 Cardozo L. Rev. 373, 382 and 413 (1991) and Ronald J. Allen, Factual Ambiguity and a Theory of Evidence, 88 Nw. U. L. Rev. 604, 609 (1994).
[13] Two competing versions of the facts could both be plausible, or they could both be implausible, or one could be plausible while the other is implausible. Any combination is possible at the outset.
[14] See Dale A. Nance, The Burdens of Proof: Discriminatory Power, Weight of Evidence, and the Tenacity of Belief (2016) 82–83, n. 219, where the author comments that "once [Allen] turns to the criminal standard, he abandons the comparative emphasis entirely, but without admitting it" and (quoted by Allen

express “dissatisfaction with a label rather than the underlying reasoning” (p. 32), as it goes to the heart of the authors’ characterization of their own theory. The theory of relative plausibility, which was originally a theory of civil trials,[15] needs further clarification regarding what it means in the context of criminal trials that the fact-finder must make “a comparisons of the explanations in light of the applicable standard of proof” (p. 14).

IV. Conclusion

The theory of relative plausibility provides a broad descriptive account of the proof process. To be empirically plausible, an account of what fact-finders generally do cannot be very specific, as the fact-finding occurring in courts is hardly uniform at the more detailed level. A prescriptive theory, on the other hand, must be more concrete, so that its recommendations are clearly understood. In this commentary on Allen and Pardo’s article, I explained why I think the theory of relative plausibility is a very promising starting point for a prescriptive theory of evidence assessment, but I also pointed out two aspects of their theory that need to be discussed further if one wants to give advice on evidence assessment using its conceptual framework. First, it is necessary to examine the epistemological status of the various explanatory criteria regarding their potential as guides to truth or likelihood – not in science but in the particular context of legal fact-finding. Second, recommendations to fact-finders regarding criminal trials should either be clear about what must be compared, or – if it is both possible and sufficient that fact-finders consider separately explanations consistent with guilt and explanations consistent with innocence – not mention a comparison at all. In my view, these are among the most pressing issues to be dealt with if a prescriptive theory of evidence assessment is to be derived from the theory of relative plausibility.

---

and Pardo on p. 29) that “[h]is suggested test for criminal cases is simply not comparative in any meaningful sense.”

[15] See Ronald J. Allen, A Reconceptualization of Civil Trials, 66 B.U.L. Rev. 401 (1986), in particular 436–437.