Structuring inferential reasoning in criminal fact finding: an analogical theory

FEDERICO PICINALI

London School of Economics and Political Science, Houghton St., London, WC2A 2AE, UK

[Received on 2 August 2011; revised on 10 February 2012; accepted on 19 March 2012]

The article proposes a normative theory of inferential reasoning for criminal fact finding, centred on the concept of ‘analogy’. While evidence law scholars have devoted little attention to the topic, the article maintains that analogy deserves more consideration. In particular, it argues that an analogical theory of inferential reasoning has three main advantages. First, the theory makes it possible to incorporate within a single coherent framework the important insights of different approaches to ‘reasoning under uncertainty’; indeed, it welcomes both the Pascalian notion of ‘relevance’ based on the Bayesian likelihood ratio and the Baconian concept of ‘weight’. Secondly, it helps advance the conventional understanding of the reference class problem, an evidential conundrum widely discussed in the recent legal scholarship. Finally, the theory allows for a functional taxonomy of reasonable doubts.

Keywords: inference; analogy; fact finding; criminal.

Without some degree of resemblance, as well as union, ‘tis impossible there can be any reasoning: but as this resemblance admits of many different degrees, the reasoning becomes proportionably more or less firm and certain. An experiment loses of its force, when transferr’d to instances, which are not exactly resembling; tho’ ‘tis evident it may still retain as much as may be the foundation of probability, as long as there is any resemblance remaining

David Hume


1. Introduction

This article proposes a normative theory of inferential reasoning for criminal fact finding. The straightforward approach of the work consists in structuring factual inference, highlighting the characteristics of its components, and exploring its essential operations. While doing so particular attention will be paid to the influence played by two crucial requirements of evidence law: that of relevance and the standard of proof.

The central claim is that inferential reasoning is to be conceived of as a type of analogical reasoning. More particularly, it is argued that the classical concept of analogy, enriched with certain important
specifications, captures an essential mechanism of inference: i.e., the translating of a general statement of fact into an individual one.

While analogical reasoning has been widely studied in connection with legal reasoning and legal adjudication, evidence law scholars have devoted little attention to it. This work aims to show that analogy deserves greater consideration. In particular, it claims that viewing criminal fact finding through the lens of an analogical theory of inferential reasoning has three main advantages.

First, the theory makes it possible to incorporate within a single framework the important insights of different approaches to ‘reasoning under uncertainty’ in a way that is consistent with the demands of evidence law. In the last few decades a notorious debate has taken place in the evidence law scholarship between those arguing for the normative attractiveness of a Pascalian theory of fact finding and those opposing it in favour of a Baconian approach. While the former theory has been well supported especially because it provides a convincing account of the notion of relevance together with a system to update prior probabilities in light of new evidence, i.e., Bayes’ theorem, the latter has been strongly endorsed in particular because it emphasizes the importance of the evidential weight on which a probability rests. The article counters the extreme stances often taken by the debaters with respect to the two theories: indeed, it views such theories not as offering alternative rules to perform the same epistemic task; rather, it considers that each illuminates separate operations which are all essential to inferential reasoning.


3 Attention will be concentrated on these particular aspects of the debate. However there is more to the question of the comparative adequacy of the two theories: consider, for instance, the disputes on the ‘difficulty about conjunction’ and on the ‘costs of precision’. See, in particular, ALLEN, Rationality, Algorithms and Juridical Proof: a Preliminary Inquiry, supra note 2, COHEN, The Probable and the Provable, supra note 2, at 49–120, TRIBE, Trial by Mathematics: Precision and Ritual in the Legal Process, supra note 2 and RICHARD O. LEMPERT, The New Evidence Scholarship: Analyzing the Process of Proof, in TILLERS AND GREEN (eds.), Probability and Inference in the Law of Evidence: The Limits and Uses of Bayesianism 61, supra note 2.

4 Other scholars have already stressed that both the Pascalian and the Baconian theory tell us something valuable about inferential reasoning, although none of the two may be considered exhaustive. See ANDERSON ET AL., Analysis of Evidence, supra note 2, at 261 and SCHUM, The Evidential Foundations of Probabilistic Reasoning, supra note 2, at 41, 201.
Secondly, the theory of inference presented here helps to advance the conventional understanding of the reference class problem, an evidential conundrum widely discussed by the legal scholarship in recent years. In the light of the proposed inferential structure, the article attempts to clarify the nature of the problem, distinguishing between the ‘logical’ and the ‘empirical’ faces thereof and underlining the context-dependence of its solution. This last feature will be expressed in terms of the normative influence played by the relevant standard of proof, namely, the reasonable doubt standard.

Finally, the proposed conceptualization allows for a functional taxonomy of reasonable doubts, a taxonomy that is automatically derived from the inferential structure outlined in the article. Differentiating between distinct types of reasonable doubt will allow for interesting considerations, ranging in subject from the choice of the standard for the admission of expert evidence at trial to the demystification of the hypotheticals concerning naked statistics.

Following an initial exploration of the structure of inferential reasoning, elaborating on its components and essential operations, the article will discuss the reference class problem in the light of such structure and then put forward a taxonomy of reasonable doubts.

2. Inferential reasoning in criminal fact finding

Inferential reasoning is defined here as the making of a judgement concerning the existence of a fact based on the judgement concerning the existence of another fact, where the term ‘fact’ refers to a knowable state of affairs of variable complexity.

The definition identifies a category of arguments that, moving from descriptive premises, aim at establishing the truth or falsity of an assertion. These arguments may be of various types. In particular, they may be either inductive or deductive, depending on whether, assuming the truth of their premises, it is logically possible or logically impossible for their conclusion to be false. Moreover, the facts stated in their premises and in their conclusion may involve either a plurality of instances or a single one.

In criminal fact finding, inferential reasoning is typically inductive and its conclusion generally refers to an individual instance. It is against this backdrop that the proposal is advanced to structure inferential reasoning as a type of analogical reasoning, in particular, one defined by two integral components: a generalization, and a probability statement. The question as to whether this same structure is normatively required in all sorts of inferential reasoning is irrelevant for the purposes of this article and will not be addressed here.

To explore the analogical structure of inferential reasoning the work will start by introducing its inner components—a necessary step towards understanding the subsequent explanation of its overall layout.

3. What sticks facts together

The word ‘generalization’ here refers to a statement expressing a synthetic relationship between the occurrences of facts belonging to distinct classes or of prototypes thereof, where a ‘class’ is a set of instances sharing characteristic properties and a ‘prototype’ is an abstract case adopted to exemplify such instances. Whether implicitly or explicitly, a generalization takes the form ‘if A then B’—a conditional that is to be understood in probabilistic terms, as the following will clarify.\(^5\)

---

Examples of generalizations are: ‘child testimony is not trustworthy’, ‘most climbers are not afraid of heights’, or ‘there is a 0.2 chance that a first-time sex offender will commit a further sex crime’. As the first two examples show, it may happen that one term of the generalization is nothing more than an attribute of the other term. To be sure, if such an attribute is a necessary element of the other term’s definition—as in the well-known sentence ‘bachelors are unmarried men’—the statement is analytic, and therefore fails to be a generalization. In other words, since generalizations are synthetic, they state what might be called a ‘relational predicate’—a phrase adopted here to refer generally to predicates expressing any relation between entities that goes beyond the mere definition of either of them. In the particular case of generalizations, such entities are represented as classes or prototypes and the relation between them is evidential in that it concerns their relative occurrences. Precisely for this ‘evidentiary relational predicate’, conveyed by the conditional through the expression ‘then B’, it is appropriate metaphorically to describe generalizations as ‘the “glue” that holds our arguments together’. The potency of this metaphor will become more evident as the discussion proceeds.

Experience is crucial in testing the reliability of a generalization. Certainly, generalizations might be created on the basis of a single observation or of no observation at all: it is the form of the statement, and not the underlying information, that makes it a generalization. However, as far as reliability is concerned, it is primarily the observation of multiple similar instances—whether in statistical research or in our everyday life—that matters. In other words, experience is the main factor that allows for the production of reliable statements to be applied generally, to wit, beyond the observed sample.

However, our experience both as humankind and as individuals is limited, to the extent that new instances are often encountered that cannot be satisfactorily classified under the generalizations we have come to recognize. A particularly problematic case is that of an instance consisting of a rare or unique intersection between classes of familiar instances. Imagine a hiker camping at an altitude of 2500 m. At night she spots a tiny steady light glowing from a fixed point on a rocky spire 20 m away from her tent. She has never before seen such a phenomenon. Personal experience has taught her that fireflies produce a light of similar colour and intensity. On the other hand, her knowledge of firefly behaviour is limited to observations of fireflies emitting an intermittent light, glowing while flying and living at a much lower altitude: therefore, a steady light glowing from a fixed point at high altitude is not for the hiker a clear sign that a firefly is out there. What might she expect to be producing that light? Is it a firefly or is it something else? Her common sense would probably suggest...
merging the available generalizations, thus creating a more specific one that better fits her epistemic needs.\textsuperscript{10}

Inferential reasoning very often resorts to generalizations resulting from similar mergers. This is a sign of its malleability, of our capacity to extract maximum benefit from our experience. However, as will be pointed out, combining generalizations is no easy task.

4. Probabilities as specifications of relational predicates

Each generalization necessarily includes a probability statement specifying, more or less precisely, the connective ‘then’ which constitutes the fulcrum of the relational predicate. Thus, a probability expresses the extent of the association between the terms of a generalization, i.e., what will be referred to as the\textit{ evidential strength} of A with regard to the existence of B.

Probability is conceived of here as abiding by Kolmogorov’s axioms and their corollaries, being therefore ‘Pascalian’ in nature.\textsuperscript{11} Thus, it may be conceptualized as a number between 0 and 1, where 0 stands for impossibility and 1 for certainty, subject to the familiar rules of negation, multiplication, and addition.\textsuperscript{12} Probability does not need to be expressed numerically, however; it may be conveyed verbally without compromising the axiomatization, in particular through those elastic terms usually referred to as ‘fuzzy’, due to their resistance to a precise definition\textsuperscript{13}, e.g., ‘probable’, ‘very probable’, ‘likely’, ‘unlikely’, and ‘possible’. There is no incoherence in maintaining that a Pascalian probability statement may be conveyed in a non-numerical fashion; as scholars have shown, terms of the above sort may be converted into a numerical scale consistent with the Pascalian tenets, particularly by referring them to intervals instead of to individual values.\textsuperscript{14}

The theoretical and normative importance of both including a probability statement as a component of inferential reasoning, and distinguishing it from the generalization where it resides, will become clearer as soon as the analogical layout of the structure is introduced and the pivotal concept of ‘weight’ subsequently addressed.

A clarification is worth making. A probability statement always underlies also the conjunction ‘if’ that introduces each generalization. This is evident in the statement ‘if there is a 0.8 chance of rain for the next day, only 10% of cyclists in London will go to work by tube’. However, it also applies to the examples of generalizations given in the preceding Section. Indeed generalizations are themselves instances of inferential reasoning, in particular a reasoning involving classes or prototypes. Therefore,

\begin{itemize}
  \item \textsuperscript{10} Back home, the hiker would discover that the tiny light was probably produced by a female \textit{Lampyris noctiluca}, a glow-worm species of the \textit{Lampyridae}—commonly called firefly—family.
  \item \textsuperscript{11} On this terminology see COHEN, The Probable and the Provable, supra note 2, at 2–4 and SCHUM, The Evidential Foundations of Probabilistic Reasoning, supra note 2, at 40, 41. Especially after Cohen’s distinction between Pascalian and non-Pascalian concepts of ‘probability’ (all referable to an overarching conception of probability as ‘gradation of inferability’), it is far from superfluous to specify in what sense the term is adopted here. See LAURENCE J. COHEN, An Introduction to the Philosophy of Induction and Probability, Clarendon Press, Oxford, 1989, at 13–27, 109–114. In any case, the opportunity and the adequacy of Cohen’s terminological move are not discussed here.
  \item \textsuperscript{12} The negation rule states that $P(\overline{A}) = 1 - P(A)$; the multiplication rule states that $P(A \cap B) = P(A/B)P(B)$, which becomes $P(A \cap B) = P(A)P(B)$ if the two events are independent; finally, the addition rule states that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, which becomes $P(A \cup B) = P(A) + P(B)$ if the two events are disjoint.
  \item \textsuperscript{13} See SCHUM, The Evidential Foundations of Probabilistic Reasoning, supra note 2, at 263–265.
  \item \textsuperscript{14} Id. at 266–269. N.B. Here no descriptive statement is being made as to how these terms are ordinarily used; I am claiming only that their use is compatible with a Pascalian axiomatization.
\end{itemize}
the final judgement of fact that they state—that concerning B—necessarily proceeds from another such judgement—that concerning A. For the sake of simplicity in the following part of the article it will be assumed that the facts constituting the basis for an inference have already been proven according to the controlling standard. Thus, only generalizations where ‘if’ implies satisfactory probabilities will be adopted. This allows the concentration of attention on the structure and dynamics of a single inference, without being concerned with the lower-level ones. Sequences of inferences are not addressed.

5. Inferential reasoning as analogical reasoning

Analogy is conceived of here in its traditional sense of ‘resemblance of relations’ between two or more cases.15 The structure of analogy is captured by the formula ‘c is to d as a is to b’, where a is called the source or base and c is called the target;16 a is to b is referred to as the phoros, a familiar relation, and ‘c is to d’ as the theme,17 the object of the discourse. The conjunction ‘as’ plays the essential role of extending information to the theme that originates from the phoros, by connecting the two.

‘A lit light bulb is to my desk as the sun is to the solar system’ is an example of analogy. As in this case, analogies are often expressed elliptically in that the reasons for the resemblance are to be understood from contextual clues. The lit light bulb might be considered analogous to the sun because it warms my desk as the latter warms the solar system and/or because they both irradiate their surroundings.

The important point indicated by the formula and the example is that analogy is not mere similarity. For a similarity to occur it is necessary and sufficient that the source and the target share at least one ‘attribute’, i.e., a property defining them.18 Thus, the sun and the lit light bulb are similar in that they both have an approximately spherical shape. While the sharing of a property between source and target is necessary for analogy, in itself it is not sufficient. The further necessary condition, one that distinguishes analogy from similarity, is that the source and the target share a relational predicate. Thus, the analogy between the sun and the lit light bulb rests not merely on their having a particular shape, colour or heat, but on their being in a like relation to other objects.

15 See CHAIM PERELMAN, LUCIE OLBRECHTS-TYTECA, The New Rhetoric. A Treatise on Argumentation (JOHN WILKINSON AND PURCELL WEAVER trans.) University of Notre Dame Press, Notre Dame, 1969, at 372. To be sure, in ancient Greek the term ‘ἀνάλογος’ originally referred to the arithmetic concept of ‘proportion’. Plato and Aristotle extended its meaning to include relations other than the purely arithmetic ones. See SCOTT BREWER, Exemplary Reasoning: Semantics, Pragmatics, and the Rational Force of Legal Argument by Analogy, 109 Harv. L. Rev. 923, 1995–1996, at 949, 950 and GEORGE POLYA, Induction and Analogy in Mathematics, Princeton University Press, Princeton, 1954, at 14. Aristotle, for instance, used the term to explain the concept of ‘justice’. See ARISTOTLE, Nicomachean Ethics (JOE SACHS trans.) Focus Publishing, Newburyport, 2002, 1131a, at 84. The concept of ‘resemblance’ is here conceived as encompassing also instances of identity of relations, such as the case of mathematical proportions—e.g. 15 is to 30 as 1 is to 2. To be sure these cases are of scarce importance for criminal fact finding.


18 Cf. GENTNER, Structure-Mapping: A Theoretical Framework for Analogy, supra note 7, at 157—defining ‘attributes’ as monadic predicates, that is, ‘predicates taking one argument’, in opposition to ‘relations’, defined as ‘predicates taking two or more arguments’. The above notion of ‘similarity’ is wider than the one proposed in id., at 159 and in DEIDRE GENTNER, ARTHUR B. MARKMAN, Structure Mapping in Analogy and Similarity, 52 Am. Psychol. 45, 1997, at 48.
This feature of analogy is often evident in metaphors. John Donne, e.g., famously wrote in reference to his own soul and that of his beloved:

If they be two, they are two so/ As stiff twin compasses are two/ Thy soul, the fix’d foot, makes no show/ To move, but doth, if th’ other do./ And though it in the centre sit,/ Yet, when the other far doth roam,/ It leans, and hearkens after it,/ And grows erect, as that comes home.

Consider the soul of Donne’s beloved. Although the similarity is disguised, this soul may be viewed as sharing with the fixed arm of the compass the attributes of being firm, unwavering, and in control. However, the poet goes beyond this, portraying a complete picture of the relationship between the lovers’ souls in the light of the compass’s dynamics. Thus, the soul of Donne’s beloved and the fixed arm of the compass are depicted as sharing a relational predicate based on the above common properties: the former behaves towards the poet’s soul as the latter behaves towards the moving arm. It is precisely drawing over a nontrivial similarity the commonality of a relational predicate that renders this analogy—and successful metaphors in general—vivid, evocative, and poetic.

Analogical reasoning may be generally defined as ‘any type of thinking that relies upon an analogy’. It has been studied with reference to a large variety of contexts and disciplines. Interestingly, in the legal area it seems that this sort of reasoning has been considered mainly, if not exclusively, in relation to legal reasoning and legal adjudication. Especially in light of the attention that analogical reasoning has received in the domain of probability theory and inductive logic, starting, most notably,

with David Hume and, later, John Stuart Mill, it is curious to notice that even the most theoretical approaches to juridical fact finding have failed to appreciate both its explanatory and its normative power. Perhaps, the application of analogy to inferential reasoning has been viewed by legal scholars as a truism: it is generally agreed that factual inference is based on the information derived from analogous cases. Once this trivial point has been established—the sceptic might contend—the investigation of analogical reasoning is unable to cast new light on the dynamics of factual inference.

It would seem that such scepticism is misplaced and unconvincing. The aim of this article is indeed to demonstrate the theoretical importance of adopting the lenses of analogical reasoning to approaching the issues presented by juridical—and, in particular, criminal—fact finding.

To this end, in the following pages inferential reasoning will be structured in terms of an analogical argument, i.e., an ‘explicit form of analogical reasoning’ where the premises consist in the phoros and in a similarity between source and target, while the conclusion is the analogy itself. The fundamental question underlying analogical arguments is to determine what warrants the conclusive analogy. Answering this question is necessarily a context-dependent endeavour: it hinges on the requirements that the analogical argument is asked to satisfy within a particular context.

6. Evidentiary analogy

In inductive logic an analogical argument may be defined as the process of inferring how $d$ is evidentially related to a particular instance $c$ from the fact that $c$ shares a relevant property or a set of relevant properties with another instance $a$ to which $b$, similar to $d$, is evidentially related. It will come as no surprise that this definition features the type of argument proposed here as the structure of the sort of inferential reasoning involved in criminal fact finding. Call ‘evidentiary analogy’ the conclusion of such argument: ‘$c$ is to $d$ as $a$ is to $b$’.

As the foregoing shows, evidentiary analogy is concerned with the proof of facts. Here relational predicates do not express an act or a function that the source and the target perform with regard to other objects, as in the examples of the lit light bulb and of the compass, nor do they refer to spatial or temporal relations. Instead, they communicate the evidential connection between the existence of source and target, and the existence of other facts.

From the preceding general introduction of analogy, it is derived that evidentiary analogy is necessarily based on similarity: a resemblance of evidential relations between source and target may only be established if the two are similar to each other in some respect. Consider the following examples. (1) A tourist travelling from New York City is looking for a cab in nearby New Haven. She knows that in New York City yellow cars are likely to be cabs. Upon seeing a yellow car approaching she might wave to the driver, believing the car to be a cab; (2) A scientist experimenting on mice has observed that a new drug is effective on a particular liver disease. Given the high degree of homology between mice and humans, she expects the drug to have a beneficial effect also on a human patient; (3)
A climber knows that intense whistles often precede rocks falling from above. Upon hearing an intense whistle, she might cling to the wall believing that rocks are falling towards her.

These examples show another peculiarity of evidentiary analogy that brings the discussion back to generalizations. In evidentiary analogy the phoros ‘a is to b’ takes the form of a generalization, ‘if A then B’. Indeed, the relational predicate within the phoros expresses here a particular relation, i.e., the connection between the occurrences of facts belonging to distinct classes. To be more precise, when included in an evidentiary analogy, the generalization is to be formulated as referring to prototypes of the facts falling into each class.27 Indeed, the individual instance constituting the target cannot be meaningfully analogized to a plurality of facts belonging to a class, only to another instance—which for the purposes of this investigation is that exemplifying these facts. The use of lowercase letters to express not only the theme—regarding the individual instance at stake—but also the phoros indicates that evidentiary analogy involves prototypes.

The preceding considerations suggest how evidentiary analogy is derived through argument. It is essentially a matter of classification. A similarity is perceived between an individual instance and the prototype of a class. This warrants applying to the instance a generalization attaching to the class, i.e., ‘transferring’ to the theme the relational predicate within the phoros, thus establishing a resemblance of relations. Following on from this logic, it is correct to affirm that the analogical argument leading to evidentiary analogy presupposes generalizing and involves individualizing.28 The choice of letters used to express the phoros and the theme indicates that each term of the latter is a member of the class exemplified by the similarly positioned term of the former.

Below is a preliminary visualization of the structure of evidentiary analogy, highlighting its essential components.

![Fig. 1. The structure of evidentiary analogy.](https://academic.oup.com/lpr/article-abstract/11/2-3/197/916298 by guest on 16 December 2018)

Before proceeding with the study of evidentiary analogy in the criminal fact finding context, two possible early objections to the proposed theory need to be addressed, both contesting the choice of understanding the phoros in terms of a generalization.

First, it might be argued that this choice simply turns evidential analogical argumentation into ‘enumerative’ induction or ‘single-case’ induction (depending on whether the generalization is built

---

27 It would be imprecise to qualify evidentiary analogy as a ‘multiple analogy’, that is, an analogy having several sources—on this terminology see ILKKA NIINILUOTO, Analogy and Similarity in Scientific Reasoning, in HELMAN (ed.), Analogical Reasoning, Perspective of Artificial Intelligence, Cognitive Science, and Philosophy 271, supra note 16, at 275 and BARTHA, By Parallel Reasoning. The Construction and Evaluation of Analogical Arguments, supra note 21, at 140–148. Indeed the source in evidentiary analogy is only one: the number of instances that have been observed in order to create it has no bearing on this.

on multiple instances or on a single one), thus making it an uninteresting and superfluous form of reasoning, or merely a different label for familiar types of inferential argument. This objection, however, would miss the point. The theory that is being proposed is not aimed at offering an understanding of analogical argument by means of already digested structure and content requirements of inductive reasoning. Its goal, instead, is somewhat the opposite: it uses analogical argument to highlight important and often ignored analogical aspects underlying induction. Moreover, the theory does not tackle the question of whether analogical argument is reducible to any familiar form of inferential reasoning and to inductive reasoning in particular; instead, it exclusively implies the significant claim that the latter—as featured in criminal fact finding—should better be understood in terms of an analogical argument leading to what is referred to here as evidentiary analogy.

The second possible objection states that ‘an account of analogical arguments that presupposes an underlying generalization . . . can trivialize an analogical argument if it makes information about the source domain unnecessary’, thus rendering the analogy between source and target immaterial to justification. By saying that in evidentiary analogy the phoros takes the form of a generalization the theory presented here does not render the analogical argument trivial. As the following will show, the fundamental operations underlying this type of analogy all problematize the information coming from the phoros. Moreover, they have a bearing on the accuracy of the analogical conclusion, an ingredient that the presence of a generalization alone does not guarantee.

Two fundamental questions regarding evidentiary analogy as the conclusion of the analogical argument are yet to be answered. First, how is the similarity between source and target connected with the resemblance of their relational predicates? Secondly, when is the analogy warranted? Both questions will be addressed in the remainder of the article. Their answers being context-dependent, they will hinge for the purposes of this enquiry on the requirements that evidence law poses to criminal fact finding.

7. Similarities of the appropriate kind: the problem of relevance

As noted above, the similarity between the source and the target is a necessary condition for analogy and evidentiary analogy in particular. It is the sharing of properties between source and target that allows theme and phoros to express like evidential relationships. However, not every similarity warrants the evidentiary analogy. There is a fundamental rationale that singles out the similarities of the appropriate kind, i.e., those having a bearing on the essential ingredient that characterizes analogies, the relational predicate. Within the vast array of similarities between the source and the target—possibly broad enough to warrant referring to them both with the same term—this rationale picks out those influencing the evidential relationship expressed by that predicate, so that the analogy is built exclusively upon them. Such is the task of ‘relevance’.

29 Cf. Bartha, By Parallel Reasoning. The Construction and Evaluation of Analogical Arguments, supra note 21, at 49, 50 and Gianformaggio, Analogia, supra note 23, at 322. This objection might extend to deduction. However, given the premises made in Section 2 regarding inferential reasoning in criminal fact finding, its scope is limited here.


31 See id. at 92.

32 Here lies the fundamental difference between the present theory and Mill’s theory of ‘analogical induction over properties’. According to the latter, ‘if we discover, for example, an unknown animal or plant, resembling closely some known one in the greater number of the properties we observe in it, but differing in some few, we may reasonably expect to find in the unobserved remainder of its properties a general agreement with those of the former, but also a difference corresponding proportionately to the amount of observed diversity’—Mills, A System of Logic, supra note 22, at 335. Therefore, in Mill’s view the proportion of
From the role played by relevance in the construction of the relational predicates involved in evidentiary analogy and from the requirement to view the predicates included in—or ‘transferred’ from—generalizations as Pascalian statements, it follows that the notion of relevance adopted here has to be Pascalian as well. More precisely, relevance is to be conceived in terms of the Bayesian likelihood ratio: 33 a property \( x \) of fact \( a_1 \) is relevant to fact \( b_1 \) if and only if \( \frac{P(x/b_1)}{P(x/b_1') \neq 1} \), i.e., the probability that the property \( x \) obtains given \( b_1 \) and the probability that the property \( x \) obtains given non-\( b_1 \) are not the same. To put it in more basic, and possibly more commonsensical terms, \( x \) is relevant to \( b_1 \) if and only if the conditional probability of \( b_1 \) given \( x \) differs from the probability of \( b_1 \) not conditioned on \( x \). 34 thus, a property is relevant to the factum probandum as long as it has some evidential strength—no matter how much and how directed—with regard to its existence. As this second definition implies, relevance should be assessed in relation to context rather than ‘in a vacuum’. It may be predicated of property \( x \) only based on the prior probability of \( b_1 \), i.e., on the information coming from the relevant properties already processed, so that the statement ‘\( x \) is relevant to \( b_1 \)’ is always short for ‘\( x \) is relevant to \( b_1 \) given the assessment of properties \( s, t \) etc.’

There is yet another way of defining this very concept of relevance, one that seems to be more suitable for the present discussion on evidentiary analogy. Consider \( a \), the prototype of the facts falling into class \( A \), to be the source in the phoros ‘if \( a \) then \( b \)’. Assume that property \( x \) partitions class \( A \) into two classes, one having the property and one not having it. If this is the case, \( x \) is relevant to \( b \) if and only if the probability of \( b \) given \( a \) and \( x \) is different from the probability of \( b \) given \( a \). More loosely, \( x \) is relevant to \( b \) if it detects a subclass of \( A \) that provides a different probability of \( b \) than the one provided by \( A \) itself. 35 The notation used in this definition refers to the phoros instead of the theme of the analogy. This has the purpose of showing that it is the generalization that provides information as to the relevance of a property. It does so by conveying through the relational predicate the evidential strength of such property with regard to the factum probandum, i.e., the probability of the latter given the former. However—to re-state a point made in the previous paragraph—it is fundamental to keep in mind that for an assessment of relevance a single generalization is not sufficient: it is necessary to compare two generalizations differing only in that one encompasses the property at issue and the other does not.

For a relational predicate to be accurately ‘transferred’ to the theme it is necessary that the target shares all the relevant properties that underlie the generalization expressed in the phoros. Indeed, the phoros ‘if \( a \) then \( b \)’ stands for a more complex statement describing \( a \) in terms of those properties that

33 This approach to the legal notion of relevance is inspired by Lempert, Modeling Relevance, supra note 2. See also Dan M. Kahan, The Economics—Conventional, Behavioral, and Political—of ‘Subsequent Remedial Measures’ Evidence, 110 Colum. L. Rev. 1616, 2010, at 1631–1641.

34 To see why the two definitions are equivalent, consider the following. Applying Bayes’ theorem \( P(b_1/x) = P(b_1) \cdot \frac{P(x/b_1)}{P(x/b_1') \cdot P(x/b_1) + P(x/b_1')} \), which may also be written as \( P(b_1) \cdot \frac{P(x/b_1)}{P(b_1)/P(b_1') + P(b_1')/P(b_1)} \). If the likelihood \( P(x/b_1) \) is equal to the likelihood \( P(x/b_1') \) (and thus their ratio is 1), then \( P(b_1/x) = P(b_1) \cdot \frac{P(x/b_1)}{P(x/b_1') + P(x/b_1')} \). But since \( P(b_1') + P(b_1) = 1 \), then \( P(b_1/x) = P(b_1) \cdot \frac{P(x/b_1)}{P(x/b_1') + P(x/b_1')} = P(b_1) \).

are relevant to the occurrence of \( b \), and the probability specifying the connective ‘then’ results from the consideration of all such properties. Therefore, only when each of these properties characterizes the target is the ‘transfer’ of the relational predicate warranted. To see why this is so, imagine that in order to determine the identity of the perpetrator a fact finder is considering whether to apply the generalization stating that somebody fleeing from the scene of the crime is likely to have committed it. Obviously, if the defendant were found simply to be loitering in the area it would be inaccurate to apply such generalization to her case. The case would be lacking a relevant property characterizing the relational predicate of the phoros.

It is important to note that relevant properties may also consist in the absence, rather than the presence, of an attribute. Thus, for instance, the absence of a hearing impairment, of an alibi, or of a mental disorder are relevant properties in inferences concerning, respectively, the accuracy of a conversational testimony, the identity of the perpetrator, and the awareness of wrongdoing.

It is yet to be explained how the proposed conception of relevance is appropriate for the purposes of criminal fact finding.

The interpretation of the legal notion of relevance in terms of the Bayesian likelihood ratio was first advanced in a seminal article by Richard Lempert\(^{36}\) with reference to rule 401 of the American Federal Rules of Evidence. In a recent debate scholars discussed whether this notion is compatible with the behaviour of—American—judges regarding the admission of evidence that is not committed exclusively to any one of the accounts offered by the parties.\(^{37}\) Although the debate was primarily stimulated by descriptive considerations, the issue it raises is useful to approach the normative question regarding the concept of relevance to be adopted in criminal fact finding. The gist of the discussion is expressed by the following passage.

Consider a simple case: a person accused of murder in a small town was seen driving to the small town at a time prior to the murder. The prosecution’s theory is that he was driving there to commit the murder. The defense theory is an alibi: he was driving to the town because his mother lives there to visit her. The probability of this evidence if he is guilty equals that if he is innocent, and thus the likelihood ratio is 1, and under what is suggested as the ‘Bayesian’ analysis, it is therefore irrelevant. Yet, every judge in every trial courtroom of the country would admit it... And so we have the puzzle.\(^ {38}\)

The issue with this passage is that it rests on a doubtful premise; nothing in the example mandates that the likelihood ratio in question equals 1. To see why this is so is it necessary to introduce the two concepts of ‘frame of discernment’ and of ‘uncommitted evidence’, both derived from the mathematical theory of evidence elaborated by Glenn Shafer.\(^{39}\) The former is a set of possible hypotheses relating to the occurrence of a fact. This set not only comprises the two general and overarching hypotheses asserting the existence and non-existence of that fact respectively (e.g. \((H_1) \ ‘X \text{ killed } Y’\) and \((H_2) \ ‘X \text{ did not kill } Y’\)), but, ‘depending on what we know or think we know’,\(^ {40}\) it contains more refined hypotheses as to how the fact did or did not come about (e.g. \((H_3) \ ‘X \text{ killed } Y \text{ with a knife}’\) and

\(^{36}\) See LEMPERT, Modeling Relevance, supra note 2, at 1025, 1026.

\(^{37}\) See PARK ET AL. Bayes Wars Reditivus—An Exchange, supra note 2.

\(^{38}\) Id. at 10.


\(^{40}\) SHAFER, A Mathematical Theory of Evidence, supra note 39, at 36.
A piece of evidence is ‘uncommitted’ when it equally supports more than one of the hypotheses within the frame (the fact that X had previously threatened Y is not committed to either (H₃) or (H₄)).

Now, relevance is relative: it depends on the task that the evidence is employed for. Thus, if the goal is to determine whether a fact took place or not, an uncommitted piece of evidence need not be irrelevant in the proposed Bayesian terms. By itself it would surely be of no use in the choice between the various hypotheses that it supports; however, unless it equally buttresses the general hypothesis that the fact took place and its negation, the piece of evidence cannot be said to be irrelevant to the question of whether the fact occurred or not. This is because—always assuming that the aim is to answer this last question—the relevance of that evidence cannot be predicated merely upon a comparison of its effects on hypotheses comprising only a subset of the frame of discernment. The crucial contest involves the two general hypotheses exhausting together the whole frame: only if the evidence is uncommitted to any of the two, is it then irrelevant to the task. Therefore, to say that a piece of evidence is not committed to any of the possible hypotheses within the frame is not yet the same as saying that it is irrelevant to the existence of the central fact that these hypotheses concern.

In the above passage the piece of uncommitted evidence—the testimony concerning the whereabouts of the defendant at a time prior to the crime—has no utility by itself for discriminating the probabilities of the prosecution’s and of the defence’s stories. However, it is not necessarily irrelevant to the question of whether the fact—in this case, that the defendant is the perpetrator—took place or not. Indeed it might help the fact finder in discarding several other possible hypotheses included in the frame, so as to alter the overall probability concerning the existence of the fact and its negation. Significantly, these are exculpatory hypotheses that the defence might put forward and try to confirm: in particular all the possible assertions that the defendant was not in town at the time of the crime. If this is so—to put it in a Bayesian tongue-twister—even if the conditional probability of the testimony given that the defendant is the perpetrator equals the conditional probability of the testimony given that the defendant was in town visiting his mother, the conditional probability of the testimony given that the defendant is not the perpetrator might turn out to be significantly lower than this value, thus rendering the evidence relevant to the fact that the defendant is the perpetrator.

In light of the above, in order to define the concept of relevance to be employed in criminal fact finding we need to determine what is the final goal of this epistemic enterprise. If we conclude that criminal fact finding aims to prove whether certain facts took place or not, not merely to choose between two competing accounts advanced respectively by the prosecution and by the defence, a piece of evidence cannot be deemed irrelevant just because it is uncommitted to either of the accounts.

Two major normative considerations suggest that this conclusion regarding the goal of criminal fact finding is warranted. First, the reasonable doubt standard is normatively referred to the existence of facts, in particular, those constituting guilt. Thus it requires that the fact finder confronts not just the parties’ cases, but the hypothesis of guilt—comprising all of the inculpatory hypotheses within the

---

41 See ibidem. See also id. at 285—pointing out that the construction of a frame of discernment is a ‘creative act’.
42 See id. at 37–56. To be sure, Shafer calls ‘not committed’ the belief related to the evidence, not the evidence itself. However, the two concepts are here considered to be cognate: belief is uncommitted when the evidence is uncommitted in the sense expressed above.
43 This is true of the American system—see In re Winship, 397 U. S. 358, 1970, stating at 364: ‘We explicitly hold that the Due Process Clause protects the accused against conviction except upon proof beyond a reasonable doubt of every fact necessary to constitute the crime with which he is charged’—and of the Italian one—see articles 187 paragraph 1 and 533 paragraph 1 of the Italian ‘Codice di Procedura Penale’.
frame of discernment—with the hypothesis of innocence—comprising all of the exculpatory hypotheses in the frame—so that facts are found only when no reasonable case for the defence stands. Moreover, the presumption of innocence allows the defence to remain passive throughout the whole trial so that it does not have to offer any exculpatory account at all, possibly leaving it to the fact finder alone to elaborate and assess accounts of this sort.

The Bayesian notion of relevance defended here is functional to the task of criminal fact finding just identified. Indeed it is built on the association between a property—a piece of evidence—and the fact tout court, not merely between that property and some of the possible particularized accounts concerning that fact. Therefore this notion of relevance is normatively demanded, lest the process of proof be utterly irrational.

It may be argued that, besides having a different standard of proof, civil and criminal fact findings also differ in a more essential respect. Indeed, the former might be conceptualized as requiring the trier of fact only to choose which party offered the most probable account, instead of determining whether it is more probable than not that the disputed facts occurred.44 However, if civil fact finding were to perform this characteristic task, then the notion of relevance to be employed therein would differ from the one that has been defended so far. To be sure it may still be defined in terms of the inequality between 1 and a Bayesian likelihood ratio, but with the peculiarity that the numerator and the denominator of such ratio would condition the probability of the piece of evidence, respectively, on one of the accounts formulated by the parties and on the other, not merely on the existence and the non-existence of the disputed fact.45 It remains to be seen whether FRE 401, and similarly formulated provisions designed for civil trials, could be construed as to make room for this particular notion of relevance.

8. Analogy: one dimension of the weight of a probability

In the previous Section it was pointed out that the phoros ‘if a then b’ stands for a more complex statement describing the prototype a in terms of those properties that are relevant to b. This complexity has great importance for analogical reasoning because it allows for a gradation of competing analogies and for the consequent selection of the most appropriate relational predicate to be applied to the theme.

Consider this hypothetical situation.46 A man dies of salmonella after eating a handful of contaminated mussels he purchased at dockside in a port. The port is composed of two large areas, North and South, which comprise three wharves each. From the statement of a reliable witness, we know that the man purchased his mussels at Wharf 3 North. We also know from a dependable agency that the local company Fresh Seafood sells 15 per cent of the seafood sold in the state each day, 50 per cent of the seafood sold in the port each day, 54 per cent of the seafood sold in the North area each day, 74 per cent of the seafood sold on Wharf 3 North each day, and 65 per cent of the molluscs sold on Wharf 3 North each day. Being the fact finder in an American federal trial, we want to determine whether Fresh
Seafood sold the deadly mussels, in order to assess the allegation that the company violated the American Federal Food, Drug, and Cosmetic Act.

With some simplification, the target—i.e., the purchase of seafood by the victim—may be described as having five properties: (1) the purchase of seafood occurred ‘in the port’; (2) the purchase of seafood occurred ‘in the North area of the port’; (3) the purchase of seafood occurred ‘on the third wharf of the North area’; (4) the purchase involved a type of seafood, i.e., ‘molluscs’; (5) the purchase involved a particular subclass of molluscs, i.e., ‘mussels’. Properties 2 and 3 are nothing but progressive specifications of property 1. The same is true for property 5 with regard to property 4. From the generalizations provided by the agency concerning the daily transactions of Fresh Seafood, we derive that properties 1, 2, 3, and 4 are relevant. On the other hand no available generalization suggests the relevance of property 5. Therefore, we leave this property aside while developing the analogy.

The target may be said to be analogous to four distinct prototypes, each exemplifying a separate class of purchases to which one of the above generalizations is respectively attached: prototype $a^1$, having property 1; prototype $a^2$ having properties 1 and 2; prototype $a^3$, having properties 1, 2, and 3; and prototype $a^4$ having properties 1 to 4. However, it is maintained here that the target is more analogous to $a^2$ than to $a^1$, to $a^3$ more than to $a^2$ and to $a^4$ more than to $a^3$. This is because the relational predicate that it shares with prototype $a^4$ is the richest.

As the example shows, evidentiary analogy admits of different degrees. Under the proposed theory, an increase in analogical degree does not depend on whether a shared property produces an increase in the probability statement. In fact, it has been stated that the hypothetical target is more analogous to prototype $a^4$ than to prototype $a^3$, notwithstanding that the evidential strength expressed by the relational predicate of $a^4$ is lower than the one expressed by the relational predicate of $a^3$. Analogical degree merely depends on the number of relevant properties shared by the source and the target. More precisely, it is conceived of as a ratio between the relevant properties shared by the target and the source and the relevant properties characterizing the target, such that the highest degree of analogy equals 1.

There may indeed exist cases in which the prototype and the target have the highest degree of analogy, yet—leaving aside for the moment the issue of the generalization’s reliability—relying on such analogy for the ‘transfer’ of the relational predicate would lead to inaccuracies. This occurs when the prototype has some relevant property absent from the target, such as in the above example of the defendant loitering near the scene of a crime. To clarify, add to the hypothetical concerning contaminated mussels a prototype $a^5$ having, in addition to properties 1 to 4, a further property that is not shared with the target (e.g., the mussels are sold by a vendor ‘wearing an apron with the brand name of the company at issue written on it’) and that is relevant based on an available generalization (e.g., ‘vendors wearing an apron with the brand name of a company written on it are likely to sell the products of such company’). Although the degree of analogy between $a^5$ and the target is

47 Consider the passage from Hume used as an epigraph as well as the following passage: ‘The exact similarity of the cases gives us a perfect assurance of a similar event; and a stronger evidence is never desired nor sought after. But wherever you depart, in the least, from the similarity of the cases, you diminish proportionably the evidence; and may at last bring it to a very weak analogy, which is confessedly liable to error and uncertainty. After having experienced the circulation of the blood in human creatures, we make no doubt that it takes place in Titius and Maevius: but from its circulation in frogs and fishes, it is only a presumption, though a strong one, from analogy, that it takes place in men and other animals. The analogical reasoning is much weaker, when we infer the circulation of the sap in vegetables from our experience, that the blood circulates in animals; and those, who hastily followed that imperfect analogy, are found, by more accurate experiments, to have been mistaken’. See DAVID HUME, Dialogues Concerning Natural Religion, William Blackwood and Sons, Edinburgh-London, 1907, at 32, 33.
the highest, the relational predicate resulting from the consideration of this further property should not be ‘transferred’ to the theme, lest the inference be utterly detached from the reality of the case at stake, which is missing such property. This is an obvious, although fundamental, additional rule to bear in mind.

It should also be remembered that sharing properties that are not relevant does not alter the degree of analogy; this has a bearing exclusively on the issue of the similarity between source and target. Indeed only relevant properties add to the relational predicate, i.e., the essential ingredient distinguishing analogy from similarity.

The importance of analogical degrees has yet to be clarified. The claim is that they express one dimension of what is usually referred to as the ‘weight of the evidence’ or, as seems more appropriate, the ‘weight of a probability’. The other dimension thereof will be examined below.

The concept of ‘weight’, first introduced by John Maynard Keynes, has been widely discussed in the literature on probability and induction and, especially, on legal fact finding. Keynes claimed that ‘the weight, to speak metaphorically, measures the sum of the favourable and unfavourable evidence, [where] the probability measures the difference’. In other words, according to his view, the weight refers to the amount of evidence relevant to an inferential task, while the probability synthesizes and expresses the information that this evidence conveys with regard to such a task.

Given that an increase in weight might determine an increase as well as a decrease of the probability, the measurements of the former and of the latter are independent. On this point Keynes affirmed that a measure of weight cannot be incorporated into the probability statement itself. The claim was later underlined by L. Jonathan Cohen in his famous attacks against the attempts to conceptualize juridical fact finding in Pascalian terms. In particular, Cohen argued that given the pivotal role played by the concept of weight in criminal fact finding, a normative theory of inferential reasoning centred exclusively on the application of Bayes’ theorem would fail, especially in this context, to accurately direct the fact finder in the performance of her task. Indeed, in light of Kolmogorov’s axioms and of the rule

52 See Keynes, A Treatise on Probability, supra note 48, at 76—stating that ‘weight cannot (…) be explained in terms of probability. An argument of high weight is not ‘more likely to be right’ than one of low weight; for the probabilities of these arguments only state relations between premiss and conclusion, and these relations are stated with equal [precision] in either case. Nor is an argument of high weight one in which the probable error is small; for a small probable error only means that the magnitudes in the neighbourhood of the most probable magnitude have relatively high probability, and an increase of evidence does not necessarily involve an increase in these probabilities’.
53 See the works from this author cited supra in notes 49 and 50.
54 Cohen, The Role of Evidential Weight in Criminal Proof, supra note 50, at 116, 117. These considerations would be less compelling in civil trials if the ‘reconceptualization’ briefly discussed in the former Section was adopted.
of negation, a probability statement completely exhausts the space between 0 (impossibility) and 1 (certainty), no matter how much relevant evidence it is based on: without information concerning the amount of such evidence there is no way for the fact finder to determine the extent to which her belief in the statement is warranted. A means of assessing and expressing this amount is therefore needed.

Cohen is clearly correct in asserting that a measurement of weight is necessary for criminal fact finding—where maximizing the quantity of evidence on which an inference is based is particularly important—and that Pascalian probabilities alone cannot provide it. However, his conclusion that a ‘Pascalian judgment is otiose’ goes too far.

As outlined in the previous section, probability theory provides a convincing account of the notion of relevance to be applied in criminal fact finding. The power of such an account resides not only in its rigour, but also in its functionality and adherence to both logic and common sense. Prior and related to this is the consideration that probabilities—the likelihood ratio in particular—are an efficient instrument for measuring and conveying the ‘algebraic sign’ and, more importantly, the evidential strength of each piece of evidence; i.e., for expressing whether such evidence speaks for the existence or the non-existence of the probandum, and the extent to which it is associated with it. Moreover, by doing so probabilities allow—especially through a process of updating—for the creation of statements of evidential strength that take into account multiple relevant properties. As Keynes’ intuitions have already shown, these are functions that a mere measurement of weight cannot perform.

Given that Pascalian probabilities and assessments of weights both play roles essential to inferential reasoning, there is no reason why they could not be made to co-exist as complementary ingredients of the same normative theory. In the analogical structure proposed here, a Pascalian probability statement specifies the relational predicate shared by the theme and the phoros. It results from the consideration of all those properties that have been found to be relevant based on a Bayesian-likelihood-ratio approach. The degree of analogy between the source and the target, instead, stands as a measurement of weight, so that in the example offered at the beginning of this Section the weightiest analogy—which is to be preferred in order to maximize the accuracy of the relational predicate ‘transferred’ to the theme—is the one between the target and prototype $a^4$.

9. Reliability: the other dimension of weight

Up to this point the conventional understanding of ‘weight’ as the ‘amount’ or ‘sum’ of relevant evidence that the inference is based upon has been accepted without critical evaluation. However, it seems that this Keynesian notion is deficient, since it fails to properly account for a factor which is essential to inferential reasoning but, differently from the analogical degree, cannot be expressed merely in quantitative terms. This criticism is based on the consideration that it would be analytically correct—and coherent with the Keynesian dichotomy between weight and probability—to make room for this factor within the notion of ‘weight’; indeed, since it regards the information underlying the probability statement, such factor cannot be captured by the statement itself, just as with the analogical degree.

55 Id. at 117.
56 The much more complex notion of ‘relevance’ proposed by Cohen in his theory of induction does not appear to be as applicable and intuitive, to say the least. See COHEN, The Probable and the Provable, supra note 2, at 138.
57 Cf. LEMPERT, Modeling Relevance, supra note 2, at 1025–1027.
What is referred to here is the reliability of the generalization that we plan to individualize to the target—a basic ingredient that influences our belief concerning both the relevance and the evidential strength of a property, i.e., the information conveyed by the relational predicate. In Section 3 it was suggested that the quantity of instances observed to produce a generalization is an important criterion in the assessment of its reliability; as statistical theory teaches, the size of the sample matters for the accuracy of the result. However, the number of observed instances is not everything: the techniques of the observation are equally fundamental. Consider, for instance, the positive influence that a proper randomization or experimental design has on the trustworthiness of a generalization resulting from statistical research.58 Thus, an assessment of reliability is not reducible to a matter of quantifying the evidence: reliability also hinges on how such evidence is discovered, handled, and interpreted.59

This said, a more comprehensive notion of weight is needed—one that is faithful to the Keynesian dichotomy, but that captures both the dimensions of analogical degree and of reliability. It is proposed that weight be defined as ‘the totality of factors which bear on the accuracy of a relational predicate’, whether such predicate is part of a generalization or is individualized to the target of the evidentiary analogy. In the first case weight is exclusively a matter of the reliability of the generalization; in the latter, instead, a measurement of weight comprises the assessment of both reliability and analogical degree—it goes without saying that if either of the two is low the relational predicate is not appropriate for the target.

Once weight is understood in its two dimensions—i.e., analogical degree and reliability—the importance it has for the analogical argument leading to evidentiary analogy may be better appreciated. In this and the previous Section a distinction is being suggested between three essential operations that characterize this peculiar form of analogical reasoning: (1) measuring the reliability of a generalization; (2) measuring the probability expressed therein; (3) measuring the analogical degree. While the second operation is necessary to determine the extent of the association expressed by the relational predicate, the first and the third operations determine whether the ‘transfer’ of such a predicate to the theme is warranted or not. This last is the fundamental function performed by the measurement of weight in the analogical argument.

As the following Sections will show, distinguishing between the two dimensions of weight plays an important analytical role in the understanding of the reference class problem, while detecting the three essential operations underlying the evidentiary analogy helps in the elaboration of a taxonomy of reasonable doubts.

See DAVID S. MOORE, GEORGE P. MCCABE, BRUCE A. CRAIG, Introduction to the Practice of Statistics, W. H. Freeman and Co., New York, 2009, at 163–210 and JOHN A. RICE, Mathematical Statistics and Data Analysis, Thomson, Belmont, 2007, 3rd ed., at 199–200. It is important to remember that randomization warrants the application of the laws of probability to the collected data. This allows for tests of statistical significance which express, through the significance level α, the reliability of an assessment of relevance. In any case, the above is far from saying that the absence of randomization necessarily jeopardizes the trustworthiness of generalizations. On this last issue see JOHN WORRALL, Why There’s No Cause to Randomize, 58 Brit. J. Phil. Sci. 451, 2007.

Figure 2 below provides a visual summary of the structure of evidentiary analogy with reference to the fundamental operations constituting the analogical argument and to the components to which such operations pertain.

10. The choice of a reference class: two faces of the problem

Speaking of a choice between competing evidentiary analogies is just another way of conceptualizing the so-called ‘reference class problem’. This problem, which has been widely discussed in philosophical and legal literature, refers to the selection of a class—or of the respective prototype—from which to derive a generalization to be individualized to the single case studied.

For our purposes, the reference class problem may be further characterized as follows. As the proposed structure has already shown, it is ‘ubiquitous’, to the extent that it regards every inferential
task of the type discussed here. Moreover, the problem is ‘epistemological’ as opposed to ‘metaphysical’. Indeed it makes sense only in the domain of conditional probabilities: were ‘absolute probabilities’—if this expression ever means anything—available, inferential reasoning itself would be unnecessary, for we would be able to assert a factual statement without having to condition it on any piece of evidence. Finally, precisely because it is not ‘metaphysical’, the problem is also necessarily ‘relative’, in that its solution depends on the evidential requirements of the context in which the inference must be drawn, most notably, on the standard of proof applied therein.

As the analogical structure clarifies, the reference class problem has two faces, the ‘logical’ and the ‘empirical’. The more familiar is the former, referring to the selection of the most suitable class or prototype among a series of classes or prototypes, relating to all of which reliable generalizations—expressing the relevance and inferential strength of the respective properties—are known. In other words, given that reliability is assumed, the logical face of the problem exclusively consists in a question of weight understood in its first dimension, that of analogical degree.

Contrary to what some legal scholars seem to claim, the solution to the logical aspect of the problem is straightforward. Once we assume the reliability of the generalizations, the class to be chosen for the inferential task is simply the most specific one among those available, where specificity depends exclusively on the consideration of the relevant properties characterizing the class and the instance at stake. Viewed in terms of evidentiary analogy, this is nothing but a restatement of a point made in the previous section: analogical reasoning demands that we pick the prototype with the richest relational predicate; i.e., with the highest degree of analogy to the target.

---


63 Cf. HAJEK, The Reference Class Problem is Your Problem Too, supra note 60, at 565, 582–584.

64 Cf. id. at 583, 584.

65 This distinction, which is now going to be discussed, is made evident in FRANKLIN, Feature Selection Methods for Solving the Reference Class Problem: Comment on Edward K. Cheng ‘A Practical Solution to the Reference Class Problem’, supra note 61, at 17 and in FRANKLIN, The Objective Bayesian Conceptualization of Proof and Reference Class Problems, supra note 5, at 559–561.

66 Cf. ALLEN AND PARDO, The Problematic Value of Mathematical Models of Evidence, supra note 2, at 112, 113. The authors—apparently without problematizing the reliability of the generalizations attached to each class—seem to be sceptical about the specificity requirement—which will be introduced shortly—and state that ‘nothing in the natural world privileges or picks out one of the classes as the right one’.

67 This conclusion seems to be obvious. Besides, provided the assumption of reliability, it may be equally derived by the rules formulated by Reichenbach, Hempel, and Salmon for the solution of the problem. Reichenbach proposed the requirement of the ‘narrowest class for which reliable statistics can be compiled’—REICHENBACH, The Theory of Probability, supra note 35, at 374 (italics removed); Hempel the ‘requirement of maximal specificity’ of the reference class—HEMPEL, Aspects of Scientific Explanation, supra note 22, at 399 (italics removed); Salmon the rule of the ‘broadest homogeneous reference class’—SALMON, Statistical Explanation and Statistical Relevance, supra note 35, at 43. In formulating their rules, Reichenbach and Salmon were also concerned with the reliability of the available generalizations, a factor that we are now taking for granted. This explains Reichenbach’s reference to the availability of reliable statistics and Salmon’s requirement of breadth. Hempel’s requirement, instead, was criticized for emphasizing excessively the canon of specificity, thus paying insufficient attention to the issue of reliability—see id. at 49. Moreover, Fetzer claimed that, differently from Salmon’s one, Hempel’s account is ambiguous, in that it is able to provide a ‘unique value solution’ to the problem, but not a ‘unique description solution’—to understand why, see FETZER, Reichenbach, Reference Classes, and Single Case Probabilities, supra note 60, at 208–210 (italics removed). This last distinction is irrelevant to our discussion on inferential reasoning. Indeed since we are primarily concerned with the proof of facts, not with providing a story as to how the facts came about, a requirement leading to a unique value solution—i.e. a unique probability statement—is enough for our purposes. The solution to the logical aspect of the reference class problem based on ‘selecting relevant features’ is also argued for in FRANKLIN, The Objective Bayesian Conceptualization of Proof and Reference Class Problems, supra note 5, at 559–561.
This logical solution plainly applies to cases both of nested and of intersected classes. The former consist in matryoshka-like classes, as in the above hypothetical situation concerning contaminated food. The latter, instead, comprise two or more classes which intersect, thus creating one or more weightier classes; consider, for instance, the intersection between the class of nearsighted eye-witnesses and the one of eye-witnesses of a crime which occurred at night—an intersection that will be chosen for an inference concerning the trustworthiness of a nearsighted eye-witness of a crime committed at night.

Consider now the well-known scenario of the ‘Nixon diamond’, which may be formulated as follows: ‘Quakers are usually pacifist, while Republicans are usually non-pacifist. Richard Nixon is both a Quaker and a Republican. Is he pacifist or not?’ The situation is one of intersected classes where no generalization is given regarding the intersection—into which the case of Nixon falls. Here, the canon of specificity alone can no longer control the choice: indeed, it would suggest the selection of a class related to which we apparently lack information.

Fact finders often find themselves before similar epistemic predicaments. The way in which these are usually tackled has been mentioned in Section 3: when information regarding the most specific class is not provided by their experience, either direct or indirect, fact finders employ generalizations which result from combining the information obtained from other classes. Fact finders do not content themselves with making a choice between the available generalizations; they create instead a maximally specific one by pasting together the relevant properties—or clusters thereof—about which they have satisfactory information. Thus, a generalization regarding the credibility of confessions ‘obtained through suggestive techniques from a minor who is mentally impaired and in a state of deprivation’ is not necessarily derived from the direct or indirect experience of cases of this particular sort; it may, and usually does, derive from a combination of the experiences referring to less specific cases.

It is evident that these considerations bring us beyond the discussion of mere analogical degree. They introduce a new difficulty: that of assessing the reliability of the generalization created by merging the ones available. We are thus led into the tricky area of what is here referred to as the ‘empirical face’ of the reference class problem, i.e., the issue of selecting the most suitable reference class—or prototype—tout court. To clarify, in addition to the logical task, the empirical face also incorporates the problem of determining the reliability of the generalizations attached to each class: it combines the issues of measuring weight in each of its dimensions, so that none of the two measurements alone may dictate the final choice of class. For example, imagine that we want to determine whether Guido, a male white-collar worker will live to see the age of 40. There are two classes available, the one of ‘males’ and the one of ‘male white-collar workers’. The first class consists of 1000 cases and suggests that males are very likely to live to forty. The second class comprises only 10 cases showing exactly the opposite. Even though the prototype of the latter class has a higher degree of analogy with Guido, it may be inaccurate to ‘transfer’ to him the respective relational predicate. Due to the smaller sample size, and all other factors being equal, the generalization provided by the most specific class is less reliable.

There is more to the empirical face of the reference class problem. Since it encompasses the issue of the reliability of the generalizations, it has a bearing on the very assessment of the relevance of those properties that might be considered when measuring the analogical degree. Thus, the accuracy of this
last operation is greatly imperilled by a lack of reliability, to the extent that the whole inferential task might be sent into a tailspin.\(^7\)

As the foregoing suggests, the reference class problem in its empirical face is the problem of inferential reasoning, since the reliability of generalizations, whether they come from common knowledge or from scientific theories, is an ingredient that no fact finder could ever take for granted. This is especially true in the context of criminal fact finding. Serving here as the controlling parameter for the solution of the problem, at least in those inferences concerning the existence of the final *probanda*, i.e., the elements of a crime,\(^7\) the reasonable doubt standard requires that the reliability of the information regarding the relevance and the strength of the evidence be carefully tested.

The issue of reliability becomes particularly problematic in those very cases discussed above in which the lack of direct or indirect experience forces the fact finder to create more specific generalizations by merging the information coming from the known ones. How are we to check the reliability of these new generalizations? On the one hand, mathematical methods accurately to combine the numerical probabilities that might be featured in cases similar to the Nixon diamond are yet to be found,\(^7\) whereas on the other, it is unclear how commonsensical reasoning copes with the merger problem. It seems, however, that it often deals with such problem successfully.

11. A taxonomy of reasonable doubts

Structuring inferential reasoning is of particular importance when it comes to the discussion of the pivotal contextual requirement for the inferential task, namely, the standard of proof. An automatic consequence of detecting the components of inferential reasoning, as well as exploring the dynamics of its structure, is that of uncovering a normative feature of the standard’s application: if such reasoning comprises several necessary operations, then the standard should be applied to each one of them.

The foregoing discussion on evidentiary analogy highlighted three essential operations: measuring the reliability of a generalization, measuring the probability expressed therein, and measuring the analogical degree. Given a particular standard of proof for the inferential task, a reliable generalization and a high probability will not suffice if the degree of analogy is unsatisfactory. On the other hand, a high degree of analogy and a high probability will not suffice if the generalization is unreliable according to the standard. Finally, a reliable generalization and a high degree of analogy will not suffice if the magnitude of the probability is not accepted as satisfactory.

In the criminal fact finding domain, at least when it comes to draw an inference concerning the existence of a final *probandum*, the reasonable doubt standard is the parameter against which each of the three operations should be assessed. It is this normative consideration that allows for the taxonomy of reasonable doubts.

\(^7\) Cf. id. at 5.

\(^7\) For a discussion on whether the inferences preceding the one concerning the existence of a final *probandum* must be governed by the reasonable doubt standard see Hamer, The Continuing Saga of the Chamberlain Direction: Untangling the Cables and Chains of Criminal Proof, supra note 2.

\(^7\) See Franklin, Feature Selection Methods for Solving the Reference Class Problem: Comment on Edward K. Cheng ‘A Practical Solution to the Reference Class Problem’, supra note 61, at 17–18. To be sure it would be accurate to use Bayes’ theorem in this case provided that reliable information as to the several probabilities necessary to applying it was available. However, often this may not be the case.
A first possible partition of reasonable doubts is the one between ‘negative’ and ‘positive’ doubts. ‘Negative’ doubts are those that consist in a lack of weight, either in the form of insufficient reliability or in the form of an unsatisfactory degree of analogy. If the prosecution resorts to an unreliable generalization (e.g., an individual who owns the car used in a robbery has surely committed such crime) or tries to individualize it to a target with a low degree of analogy to the source (e.g., an individual who owns that car, but reported its theft several months before the robbery took place), no matter what the magnitude of the probability, its case will leave room for one or more negative doubts. On the other hand, a doubt is ‘positive’ when it consists in the insufficiency of the probability’s magnitude with respect to the hypothesis to which it refers. Such sort of doubt stems not from an evidential void, but from the reckoning that the probability is insufficiently favourable to the hypothesis, i.e., that its negation is entirely reasonable. Thus, e.g., saying that human beings usually intend the results of their actions is accurate, yet ‘usually’ is not enough to prove beyond a reasonable doubt that a defendant intended an assault to result in the death of the victim.

A more specific partition of reasonable doubts is based on the individual operation upon which the doubt is cast. Thus, a reasonable doubt could concern the reliability of a generalization, the magnitude of the probability, or the degree of analogy. This tripartition leads to various considerations, the first regarding the admissibility of expert testimony.

In the particular case of expert testimony, the law usually subordinates admissibility to a test of dependability. This is merely a provisional assessment for the purpose of deciding whether the testimony may be introduced or not; a ultimate judgement on dependability will be possible only once the expert has testified, thus furnishing important information bearing on this decision. Generalizations are the fundamental ingredient of expert testimony: even when the expert is allowed to draw an inference himself concerning an issue of the case, the bulk of his expertise comes in the form of general statements concerning classes or prototypes of facts. Therefore, a check of the dependability of expert testimony for admissibility purposes is essentially an assessment of the reliability of expert generalizations. Given that, as claimed above, the reliability of a generalization adopted by the prosecution to infer a final probandum must be established beyond a reasonable doubt for conviction, then the expert testimony requested by the prosecutor with this very aim should be admitted only if the court—whether in a jury system or not—judges that it has the potential to satisfy this standard. This position is perfectly rational, since it simply affirms that the standard of reliability for admissibility purposes should be proportional to the aim that the expert testimony is introduced for, establishing a fact beyond reasonable doubt. However, if the foregoing is correct, then the defendant


Consider, for instance, Rule 702 of the American Federal Rules of Evidence.

Consider Rule 702 (3) and Rule 704 of the American Federal Rules of Evidence. However, see also the Advisory Committee’s Note to Rule 702 reported in MUELLER AND KIRKPATRICK (eds.), Federal Rules of Evidence. With Advisory Committee Notes and Legislative History, supra note 59, at 178, 182—stating: ‘it seems wise . . . to encourage the use of expert testimony in nonopinion form when counsel believes the trier can itself draw the requisite inference’.

An issue of admissibility is not posed by non-expert generalizations for the simple reason that, given that these generalizations are based on common knowledge, no particular device is required for their introduction at trial. They are immediately available to the fact finder—although often they need to be formulated and exposed by the parties. In any case they too must be found reliable beyond reasonable doubt for conviction.

This claim is similar to the one made by Michael Risinger in his discussion of the US Supreme Court decision in Kumho Tire v. Carmichael (526 U. S. 137, 1999). See D. MICHAEL RISINGER, Preliminary Thoughts on a Functional Taxonomy of Expertise for the Post-Kumho World, 31 Seton Hall L. Rev. 598, 2000–2001, at 533–536. In particular, referring to the question

74 Consider, for instance, Rule 702 of the American Federal Rules of Evidence.
75 Consider Rule 702 (3) and Rule 704 of the American Federal Rules of Evidence. However, see also the Advisory Committee’s Note to Rule 702 reported in MUELLER AND KIRKPATRICK (eds.), Federal Rules of Evidence. With Advisory Committee Notes and Legislative History, supra note 59, at 178, 182—stating: ‘it seems wise . . . to encourage the use of expert testimony in nonopinion form when counsel believes the trier can itself draw the requisite inference’.
76 An issue of admissibility is not posed by non-expert generalizations for the simple reason that, given that these generalizations are based on common knowledge, no particular device is required for their introduction at trial. They are immediately available to the fact finder—although often they need to be formulated and exposed by the parties. In any case they too must be found reliable beyond reasonable doubt for conviction.
77 This claim is similar to the one made by Michael Risinger in his discussion of the US Supreme Court decision in Kumho Tire v. Carmichael (526 U. S. 137, 1999). See D. MICHAEL RISINGER, Preliminary Thoughts on a Functional Taxonomy of Expertise for the Post-Kumho World, 31 Seton Hall L. Rev. 598, 2000–2001, at 533–536. In particular, referring to the question
should be allowed to introduce expert testimony to counter the inference of a final probandum on the basis of a lower standard of reliability, since such testimony only serves to raise a reasonable doubt. More precisely, a possible expert generalization proposed by the defendant with that very aim should be admitted as long as it promises to be at least reasonably reliable. Nothing more than this should be required, if we are to take the reasonable doubt standard seriously.

To clarify this point, consider the following situation. The prosecution proposes an evidentiary analogy resting on the expert generalization ‘if \( a^{1,2,3} \) then \( b \)’, where 1, 2, and 3 are relevant properties known to be shared by the source and the target. Imagine a child-abuse case in which the prosecution’s expert claims that ‘a child (1) having behavioural problems, (2) presenting regression in toilet training and (3) vomiting frequently is likely to have been the victim of abuse’. To counter the evidentiary analogy proposed by the prosecution, the defendant—for instance, the child’s father—may try to do one or more of the following three things: raise a reasonable doubt as to the degree of analogy, raise a reasonable doubt as to the reliability of the generalization, and raise a reasonable doubt as to the magnitude of the probability expressed therein.

The last defensive action would not require the defendant to introduce any further expert evidence; he would only have to claim that the probability expressed with ‘likely’ does not satisfy the reasonable doubt standard. The other two avenues, instead, necessitate increased effort from the defence. To contest reliability alone, the defendant could introduce another expert generalization based on exactly the same relevant properties, but showing that these properties have been poorly assessed by the prosecution’s expert because either of the insufficiency of cases studied or of a flaw in the methods of observation and interpretation thereof. For instance, a competing generalization might be proposed suggesting that given those properties alone it is inaccurate to deem the abuse ‘likely’, because—the generalization states—they are frequently observed in non-abused children too. Since the aim of doing so is to raise—at least—a reasonable doubt concerning the reliability of the prosecutor’s generalization, the admission of the defence’s expert evidence should be controlled by the standard suggested above.

Now consider the case in which the defendant contests the degree of the evidentiary analogy proposed by the prosecution. In order to do so the defence could introduce an expert generalization involving, this time, one or more other relevant properties that the target is known to possess, which were ignored by the prosecution and are able significantly to alter the probability of the accusatorial hypothesis. It must be noted that this argument would not raise a ‘positive’ doubt, one concerning the of admissibility of expert testimony Risinger states that ‘[a]ll things being equal, the higher the standard of proof applicable to the issue upon which the expertise is offered, the higher the required threshold of dependability should be’.


79 As clarified supra in Section 4, it is assumed as already established that the target has a given set of properties, since here the focus is exclusively on the individual piece of inferential reasoning based on such facts, not on a sequence of inferences. Of course, in a real case the defendant could also challenge the lower-level inference based on which one or more of these properties are believed by the prosecution to characterize the target. She might argue, for instance, that the case has no original problems, thus jeopardizing the above generalization proposed by the prosecution; to see why the prosecution’s inference would be endangered bear in mind the discussion on prototype \( a \) in the example concerning contaminated food considered supra in Section 8. In any case a direct challenge to an inference—that is, an attack that does not consist in unsettling a lower-level inference in the sequence—must necessarily take one of the three avenues indicated in the text.

80 Once again: for the present purposes these properties are assumed already to have been established. In the example the prosecution is ignoring them because of sloppiness or bad faith. Thus, they comprise not the object but the starting point of the expert testimony. See the preceding footnote, n. 79.
magnitude of the original probability, but a ‘negative’ doubt concerning, instead, a lack of weight in the prosecution's evidentiary analogy. The defence's expert generalization might be the one stating that when properties 1, 2, and 3 are present together with property 4 (e.g., the additional circumstance that the child has an autistic disorder, which might explain all the above symptoms) the fact b, i.e., the abuse, is improbable. Even in this case the introduction of expert evidence by the defence aims to raise—at least—a reasonable doubt, although now with regard to the analogical degree. Thus, it appears that its admissibility should depend once again on a standard proportional to this task.

Two final shorter points are worth making with reference to doubts concerning, respectively, the magnitude of a probability and the degree of analogy.

As to the former, it is interesting to note that once we assume that in evidentiary analogy the negative of a probability may by itself raise a reasonable doubt, then the longstanding debate about the quantification of the reasonable doubt standard in Pascalian terms is relevant to the proposed normative theory. To be sure, this quantification would be of no use for the assessment of reliability and of analogical degree which, as has been stressed before, are not subject to a Pascalian measurement: it would only regard the operation of measuring the probability statement. In any case it remains to be seen whether quantification would be able to maintain the flexibility that the concept of 'reasonableness' proves itself to have with regard to the evolution of common sense—a quality whose examination is left for future research.

The latter type of reasonable doubt, i.e., the one referring to the analogical degree, helps to clarify the doctrinal dispute on the so-called 'naked statistics'. Indeed, the uneasiness created by famous hypotheticals such as the 'prisoners in the yard' may be explained in terms of the insufficiency of the degree of analogy between source and target. Confronted with a reliable generalization and a high

---


83 Here are the facts of the hypothetical. 'In an enclosed yard are [a hundred] identically dressed prisoners and a prison guard. The sole witness is too far away to distinguish individual features. He sees the guard, recognizable by his uniform, trip and fall, apparently knocking himself down. The prisoners huddle and argue. One breaks away from the others and goes to a shed in the corner of the yard to hide. The other [ninety-nine] set upon the fallen guard and kill him. After the killing, the hidden prisoner emerges from the shed and mixes with the other prisoners. When the authorities later enter the yard, they find the dead guard and the [hundred] prisoners. Given these facts [ninety-nine] of the [hundred] are guilty of murder' NESSON, Reasonable Doubt and Permissive Inferences: the Value of Complexity, supra note 82, at 1192, 1193—the figures have been changed from the original to make the hypothetical more 'extreme'.
probability of guilt, we nonetheless feel that a defendant, identified among the persons featured in the hypothetical, should not be convicted solely on this evidence. The reason for this might be that one would expect many more relevant properties to be discussed in a similar—but real—situation, namely properties regarding the defendant’s peculiar case, which would help in revising the proposed generalization by making the source closer to the target.\(^{84}\) Only if one were persuaded that the hypothetical truly included all the possibly relevant properties likely to be found in similar situations would such uneasiness not be experienced.\(^{85}\) This, however, is not the case here.

Hypotheticals have to be simple if they are to be used as handy tools to buttress arguments about reality; however, the more an argument is framed in general terms, the more the very complexity of reality is important as a parameter for assessing its soundness. This is when the use of hypotheticals may begin to confuse rather than clarify the situation.

12. Conclusion

The aim of this work is mainly theoretical. When it was embarked upon, the principal goal was to clarify certain normative aspects of inferential reasoning as employed in criminal fact finding, which at times seemed to be either ignored or made obscure by the relevant legal literature. Still, although practicality was not the primary concern of this research, it is possible that some of its results are of immediate practical significance, particularly the contents of the last section: the discussion on the standards for the admission of expert evidence, as well as the taxonomy of reasonable doubts, which might prove to be a functional tool for the criminal fact finder to check the most problematic inferences she is asked to draw.

In any case, several questions relating to the proposed analogical approach are yet to be addressed. As fuzzy logic, semantics, and cognitive science suggest, classes are rarely internally homogeneous; moreover, the very membership of a class is most accurately conceptualized as a matter of degrees, referring to the ‘distance’ of an instance from the prototype—understood here as the central and clearest case.\(^{86}\) It is uncertain how this might affect the entailment to ‘transfer’ a relational predicate to the target. Does the degree of membership influence the question of relevance, so that a property might cease to be relevant when the target is sufficiently distant from the prototype? If not, may such degree, while leaving relevance untouched, at least influence the magnitude of the probability statement that we are entitled to ‘transfer’ to the target?\(^{87}\)

Another difficulty is that of providing a more precise conceptualization of the trade-off between maximizing the analogical degree and maximizing the reliability of generalizations. Should the

---

\(^{84}\) For instance: the defendant in a real case involving facts similar to those described in the hypothetical might not have had a criminal history of violent acts; it is possible that he was expecting to be released in the days immediately following the riot, thus he had a strong disincentive to misbehave; he might have been a friend of one of the guard’s brothers or of the guard himself, etc. If any of these relevant properties were present, the source within the numerical generalization featured in the hypothetical would have a low degree of analogy with the target.


\(^{87}\) The interaction between specificity and typicality is discussed in WANG, Reference Classes and Multiple Inheritances, supra note 49, at 4–5, 8.
balance be left to inscrutable commonsensical ‘information-processing mechanisms’ or is there any explicit conceptual tool that may somehow improve our ‘submerged’ reasoning?\(^{88}\)

It is debatable whether these questions are good only for pure philosophical speculation or whether they may lead to results that are significant for juridical fact finding. The hope is, of course, for the latter to be true.

\(^{88}\) Cf. Peter Tillers, The Structure and Logic of Proof in Trials, 10 L. Probability & Risk 1, 2011—stating at 3 that ‘the job of explicit inferential methods is, to the extent possible, to make the relatively implicit, the partially submerged, more explicit and less submerged’.