Artificial Intelligence and Law (2020) 28:91–109
https://doi.org/10.1007/s10506-018-9233-1

Assessment criteria or standards of proof? An effort in clarification

Giovanni Tuzet[1]

Published online: 10 October 2018
© Springer Nature B.V. 2018

Abstract
The paper provides a conceptual distinction between evidence assessment criteria and standards of proof. Evidence must be assessed in order to check whether it satisfies a relevant standard of proof, and the assessment is operated with some criterion; so both criteria and standards are necessary for fact-finding. In addition to this conceptual point, the article addresses three main questions: (1) Why do some scholars and decision-makers take assessment criteria as standards of proof and vice versa? (2) Why do systems differ as to criteria and standards? (3) How can a system work if it neglects one of these things? The answers to the first and second question come from the historical and procedural differences between the systems. The answer to the third focuses on the *functional connection* between criteria and standards.

Keywords Assessment of evidence · Beyond reasonable doubt · Fact-finding · *Intime conviction* · Judicial discretion · Standards of proof

1 Introduction

Legal argumentation concerning disputed facts deals with evidence that enables fact-finders to reach accurate verdicts concerning the facts. Evidence is necessary to support the factual claims made by the parties and the findings of fact made by the decision-makers. But evidence per se does not yield verdicts. Evidence must be assessed, and the fact-finders need to consider whether it satisfies a relevant standard of proof.

There has been a lot of work, recently, on the issue of legal standards of proof. Legal argumentation theorists, legal evidence scholars, civil and criminal procedure scholars among others, have extensively dealt with this issue. Some of them have made an analytical effort to clarify the idea of a probative standard; others have made some descriptive work to understand how standards actually operate; some

✉ Giovanni Tuzet
giovanni.tuzet@unibocconi.it

[1] Bocconi University, Milan, Italy

others have made a kind of normative work in the hope of suggesting better standards or at least better defined ones; and the best contributions to the debate do more than one of these things at the same time.[1]

That complex work on standards of proof has been mainly carried out in the common law world. In the civil law world there is less attention to that issue, but a greater concern for the criteria of evidence assessment. A very old Continental debate focuses on the question whether it is better to have evidence assessed according to fixed legislative or jurisprudential criteria (it was the idea of a "legal proof"), or to more relaxed ones accommodated by fact-finders themselves (it is the idea of a "free assessment" of evidence).[2] Sometimes the fixed criteria are understood as objective and the relaxed ones as subjective, in that the former constrain the fact-finders in a way that the latter do not. But various scholars have insisted on the possibility and importance of understanding also the latter criteria as objective: their objectivity would not consist in the fact that they are legislatively or jurisprudentially mandated, but in their epistemological grounds.[3] To give an example, in Italian civil cases judges are supposed to perform a "prudent assessment" of the evidence (more on this below). This is a rather relaxed criterion that confers significant discretion upon judges, but one can insist on the need to ground it in epistemic reasons, if "prudent" is read epistemically (if it is read practically, stressing the values at stake, one should add the relevant practical reasons).

Now, if we look at those issues and try to conceptualize them we see that *one thing is an evidence assessment criterion and quite another is a proof standard*. For evidence assessment is not sufficient to make a decision and because, in principle, the satisfaction of a proof standard can be reached through different criteria. Take the "beyond a reasonable doubt" standard: in principle it can be satisfied both through fixed and relaxed criteria, provided that in the first scenario the legislature or some other authority determines *ex ante* the fixed value of the relevant types of evidentiary items together with the relevant threshold, whereas in the second scenario fact-finders make appeal to some non-fixed criteria (being them epistemic, practical, or else) to assess the evidence at disposal and judge whether it satisfies the relevant standard.

Still, both in the common law and in the civil law world there is some tendency to overlap those issues and mix them up. For instance, the French *intime conviction* is often taken as a subjective criterion of assessment, while others take it as a subjective standard of proof: to quote a couple of Continental reputed scholars, Jordi Ferrer thinks it is an "entirely subjective" standard of proof, whereas Daniel González Lagier says it is a "system of evidence evaluation".[4] The worry about this state of affairs is not only theoretical, to my sense, but also practical in that such confusion makes decisions opaque as to what justifies them. Confusion results if the assessment criteria vocabulary is used to talk about standards of proof and vice versa.

[1] See among others Redmayne (1999), Lillquist (2002) and Picinali (2013). See also Vázquez (2013).
[2] On the criminal debate see Ferrajoli (1989: 112ff). See Taruffo (1992: 361ff) on civil aspects too.
[3] See Ferrer (2007: 147ff).
[4] See Ferrer (2007: 145) and González Lagier (2013: 52).

The present paper aims at making an effort in clarification, providing an adequate conceptual distinction between proof standards and evidence assessment criteria, to better understand and categorize them. Three things are to be noted in particular: (a) the differences between criteria and standards reflect some procedural and institutional differences; (b) the scope of legal argumentation is significantly reduced when fixed criteria of evidence assessment are mandated; (c) it is arguable that there is a functional connection between criteria and standards, which does not amount to say that they are the same thing or do the same job.

The most theoretically ambitious point of the paper is the claim about the *functional connection* between criteria and standards: some standards incorporate assessment criteria and vice versa. The functional connection account will be distinguished from a functional equivalence account, and some systems and jurisdictions will be referred to in order to support the functional connection claim.

The structure of the paper is as follows: Sect. 2 makes some analytical starting points to clear the discussion ground; Sect. 3 performs some comparative work considering different legal systems and the ways in which they deal with our problem; Sect. 4 tries to convey a sharper analysis and addresses some research questions, arguing in particular that criteria and standards have a functional connection; Sect. 5 concludes by recalling the research questions addressed and summing up the answers provided.

2 Some analytical starting points

Evidence enables fact-finders to reach accurate verdicts on litigated facts. But, as I said earlier, evidence per se does not yield verdicts. To this purpose there are at least four requirements on juridical evidence: first, it must be admissible according to the rules of the relevant legal system; second, it must be "inferentialized" by the parties and the fact-finders, since evidence doesn't speak for itself and the participants in the dispute have to construct evidentiary arguments based on the items presented; third, evidence must be assessed to determine its probative value, or, better, the evidentiary inferences and arguments at stake must be assessed to determine the evidentiary support, or warrant, or justification provided by the premises to the conclusions; fourth, fact-finders need to consider whether the evidence meets the relevant standard of proof, or whether the relevant burden of proof has been satisfied.[5]

[5] Many of the points discussed in this paper under the heading of "standards of proof" can be also discussed under the heading of "burdens of proof" (see Allen 2014). To make things simpler I won't use the latter terminology and the corresponding conceptual apparatus. Anyway, let me clarify this: it is true that, as one reviewer has observed, the burden of proof concerns the question as to which party has to prove what, while the standard of proof concerns the question as to "how much" evidence is necessary to discharge a given burden; but it is also true that the literature (especially in the US) focuses on one burden, namely the "burden of persuasion", which (a) is taken to indicate the relevant standard of proof and (b) is taken to be what another burden (the "burden of production") is a function of; see McNaughton (1955).

The points that concern us here are the third and the fourth. It is one thing to assess, or evaluate, the evidence presented. And quite another to find whether it satisfies the relevant standard of proof. The two operations are conceptually different and the latter requires the former as a necessary condition. No one can determine whether the evidence admitted and presented meets the relevant standard, unless the evidence has been assessed. To be more precise, the evidence must be "inferentialized", namely translated into evidentiary arguments, and the fact-finders have to assess such arguments in order to see whether the relevant standard is met or not. In fact, assessment per se doesn't determine a decision either, because a definite outcome can be justified only if a standard of proof is considered.

"Prudent assessment", for instance, is an assessment criterion in civil cases but, to make a decision, we need a standard of proof like the "preponderance of the evidence" (more on this below). No decision follows from assessment alone.

So, *both assessment criteria and standards of proof are necessary components of legal decision-making about facts*, and none of them is a sufficient condition of it.

This seems quite obvious from a conceptual and philosophical point of view. But, surprisingly enough, some courts and jurists seem to miss the point. One significant example is provided by the European Court of Human Rights (ECtHR) in *Gäfgen v. Germany* (2010), when the Court says that "in assessing the evidence on which to base a decision […] the Court adopts the standard of proof 'beyond reasonable doubt'" (§ 92). The understanding of "beyond reasonable doubt" as an assessment criterion appears as a recurrent theme in the decisions of the ECtHR.[6] In these decisions the standard of proof is taken to be an evidence assessment criterion, instead of being a decision rule which comes into play once the evidence has been assessed. The same confusion is not infrequent in scholarly work, when evidence assessment criteria are equated with standards of proof or vice versa standards are equated with criteria.[7] Now, why is it so? Why do educated scholars and skilled decision-makers fall into this confusion about evidence and proof? Why do they fail to see the difference between assessment criteria and standards of proof? This will be our first research question.

Before advancing some hypothesis on this surprising phenomenon, let us take a closer look at the rules of some contemporary systems.

[6] See e.g. *Ireland v. UK* (1978, § 161), *Labita v. Italy* (2000, § 121), *Fedorov v. Russia* (2011, § 57), *Rizvanov v. Azerbaijan* (2012, § 45) and *Najafli v. Azerbaijan* (2012, § 36). On the other hand, in these decisions the Court usually adds that such proof may follow from sufficiently "strong, clear and concordant inferences" or "similar unrebutted presumptions of fact", which in fact amount to ways of assessing the evidence (epistemically as to "strong, clear and concordant inferences", and legally as to presumptions).

[7] On the common law side see e.g. Clermont and Sherwin (2002) (treating *intime conviction* as a standard); on the civil law side see e.g. Iacoviello (2006) and Caprioli (2009) (equating "beyond a reasonable doubt" with an evidence evaluation method).

3 Some comparative work

I sketch in this section an overview of the rules of some actual legal systems about evidence and proof. Some systems have rules that clearly consist in standards of proof, whereas others have rules that clearly consist in evidence assessment criteria, while there are also rules which are difficult to categorize one way or the other. Given the theoretical purpose and the conceptual approach of the present paper I will overlook many positive details and make abstraction from many systemic complexities. What is important for our purposes is the grasping of the ways in which different systems handle the issue we are dealing with.

I shall start by making some remarks on the common law world and then move to some civil law countries.

3.1 Common law countries

In the common law world three basic standards of proof are distinguished. Two of them apply to civil cases, one to criminal cases.

In civil cases the usual standard is the *preponderance of the evidence* (as they call it in the US), or the *balance of probabilities* (as they call it in the UK). This means that, according to the received view, the claim of the burdened party (usually the plaintiff) must be more probable than not in light of the evidence presented. If the factual claim of the party with the burden of proof is considered to be more probable than not, given the evidence, that party is entitled to a verdict in its favor. If it is not considered to be so, the party has no such entitlement.

But in some civil cases the US system adopts a second standard, the *clear and convincing evidence* standard. It is used in "serious cases", namely cases in which what is at stake is more serious than in normal ones. Examples of such "seriousness" are punitive damages cases and immigration cases where what is at stake is the right to be in a certain country.[8] This second standard is more demanding than the first. In order to have a verdict in its favor, the burdened party must make a greater evidentiary effort, presenting evidence that is stronger, or has a higher probative value, than that which suffices in normal cases.

In criminal cases the well-known standard of guilt, especially entrenched in the US, is the proof *beyond a reasonable doubt* (BARD). This standard requires an amount of evidence that only leaves room for unreasonable doubts about the defendant's guilt. So, if the evidence presented makes it reasonable to believe that the defendant is guilty and makes it unreasonable to doubt it, then the prosecution is entitled to a verdict in its favor and the triers of facts are committed to decide against the defendant. Of course this criminal standard is even more demanding than the clear and convincing evidence standard used in serious civil cases: the evidence presented has to be even stronger than this, leaving room for unreasonable doubts only,

[8] See Redmayne (1999: 189) (where the adoption of this standard in the UK is suggested to face such serious cases).

which is to say doubts that lack specific epistemic reasons, being merely metaphysical (such as Cartesian doubts and "brains in a vat" doubts[9]) or essentially philosophical and cast in probabilistic terms (such as the idea that we cannot be 100% sure of anything).

There have been some attempts to translate those qualitative standards into quantitative thresholds. This is done with some probability theory and some assignment of a numerical probability value to each of the standards. It is not a difficult task if we consider the preponderance of evidence, which is easily translated into something such as $>.5$, while the clear and convincing evidence is usually considered to be around .7 (or .75) and the proof beyond a reasonable doubt is usually located around .9 (or .95).[10]

The rationale for having a higher standard in serious civil cases, compared to non-serious ones, is the protection of some more important value which is at stake in them. And the received rationale for having an even higher standard in criminal cases, compared to civil ones, is the protection of the individual from state coercion, that is, the protection of a fundamental value such as individual liberty (and even life in those systems where capital punishment is still in place). There are quite sophisticated pieces of literature that work out the details of the received view.[11] And there are critical pieces that address some deficiencies in the standards or in the ways they operate. One of the disputed things is the capacity of those standards to do the jobs they are designed for (minimizing errors, distributing errors in an appropriate way, or else), due to the lack of empirical information about particular cases or to conceptual confusion about some aspects of the standards. There is in fact an increasing amount of critical literature (on the criminal standard in particular[12]) motivated by the idea that such standards are not really helpful in legal decision-making, insofar as they do not provide genuine epistemic reasons to decide a case one way or the other. This is not the topic of the present paper, though. What I want to focus upon is the ultimate nature, so to say, of such standards: I join those who think that they are decision rules.

The standards of proof set the qualitative or quantitative threshold that must be reached to have a decision in favor of the burdened party. This party bears the risk of there being an amount of evidence which is insufficient with respect to the standard. If the evidence is insufficient, the burdened party has no entitlement to a verdict in its favor. If it is sufficient, the party has such an entitlement. So standards of proof

[9] Putnam (1981: 1ff, 1995: 17).

[10] Of course the .9 value is worrisome for criminal defendants, but the higher the standard is set (to avoid false convictions) the higher is the expected rate of false acquittals. I cannot expand on these topics here. See among others Kaplan (1968), Kaye (1986), Kaplow (2012), Cheng (2013), Allen and Stein (2013) and Verheij (2014).

[11] See especially Bell (1987).

[12] Laudan (2006: 52ff). Note that the UK has been recently abandoning BARD (out of troubles in making it understood?): now fact-finders are to convict only if they are sure (see Roberts and Zuckerman 2010: 253ff). But on BARD's interesting origins see Whitman (2008) (it was supposed to protect the juror's religious conscience facing the risk of wrongful convictions, it was not originally designed to protect the defendant).

are in the end decision rules. They justify decisions in favor of one party or the other, on their factual claims given the evidence.

Now, all of that clearly presupposes an assessment of the evidence. An evaluation of the evidence that was admitted, presented and discussed is necessary to the judgment about standard satisfaction. Standards of proof, *pace* the European Court of Human Rights, are not criteria of evidence assessment. Instead, they presuppose and require some criterion that determines the probative force, or probative value, of the evidence in play, so as to see whether it meets the relevant threshold.

To make a simple exemplification, it is not enough to assign the plaintiff's claim a .65 probability of being correct: in order to make a decision we need to know whether the standard is the simple preponderance of the evidence or the clear and convincing evidence standard. On the other hand, the standard itself doesn't tell you if the evidence presented satisfies it: you need to assess the evidence, either quantitatively or qualitatively. To take a qualitative example, the well-known decision in O.J. Simpson's criminal case can be understood like this: the assessment lead to the "he did it" hypothesis as the best explanation of the evidence but also to an acquittal given that, according to the jury, the evidence didn't match the BARD standard.

If this reconstruction is correct, where are the evidence assessment criteria in the common law world? Basically, they are in the scholarly disputes about evidentiary value, or probative force, or the weight of (combined) evidence.[13] They are not codified, as far as I know, in statutory rules nor in other legislative materials. Truly, there are some documents that give general guidance. For instance, the Crown Court Compendium (Part I, on "Jury and Trial Management and Summing Up") by the English Judicial College, where assessment criteria for the jury are given with reference to inferences from expert evidence, bad character evidence, hearsay, and the defendant's silence at trial, to mention some of the issues.[14] For the rest, the criteria are advanced in scholarly controversies and reflected in jury instructions and in judicial opinions insofar as these take a position on such controversies. Consider the literature about probability theories and their application to legal matters and cases; consider the controversy between the supporters of some subjective (Bayesian) conception of probability and the supporters of objective probabilities, or the dispute between Pascalian and Baconian probabilities, or the passionate controversy between the supporters of the IBE (Inference to the Best Explanation) account and the supporters of the mathematical probability account[15]; consider also the ways in which judges explicate to juries, with the so-called "instructions", how to assess whether a standard is met. All of this is about the criteria that fact-finders need to use to check whether the evidence presented in a case satisfies the relevant standard of proof.

[13] See, among others, Anderson et al. (2005: 224ff), Barzun (2008) and Haack (2014: 208ff).

[14] I thank one reviewer for drawing my attention to this.

[15] See e.g. Cohen (1977), Redmayne (2003) and Allen and Pardo (2007). See also Verheij et al. (2016).

3.2 Civil law countries

Civil law countries usually have in their respective codes some basic rules about evidence and proof. Most of the time these rules set evidence assessment criteria instead of standards of proof. And nowadays these criteria are rather relaxed, leaving room to judicial discretion. This has a definite historical explanation: in many Continental countries the codification effort pointed at getting rid of, *inter alia*, the prior complexities in the assessment of evidence. Those complexities amounted to the so-called system of "legal proof", which dominated Europe from the end of the XIII century to the French Revolution at the end of the XVIII.[16] In that system every sensible kind of evidence was assigned a predetermined and fixed probative value, on the basis of some legal authority or doctrinal opinion; then, once some evidence was presented at trial, the judge's task was basically that of calculating the value of each factual claim given the evidence at disposal. In other words, once the *types* of evidence had received a fixed value, the judge was to consider the *tokens* in the case at hand and to calculate the outcome accordingly, with no or little discretion on his or her part.[17] The best known example of that was the traditional two-witness rule according to which two independent witnesses testifying the same amounted to "full proof" of the fact so testified, leaving no room to the judicial appreciation of the witnesses' credibility, of the circumstances of testimony, etc. Then, to cut a long story short, Enlightenment legal thinkers and French Revolution authorities designed a system that, on the contrary, trusted fact-finders conferring some discretion upon them but, at the same time, stating guidelines of evidence assessment. Eventually the spirit of the Revolution spread throughout Europe and civil law countries adopted the new framework, which is still in place.

Let me start with Italy. The actual code of civil procedure states that fact-finders must evaluate the evidence according to their "prudent assessment" (art. 116), unless a criterion of "legal proof" is mandated. There are in fact some provisions attributing legal value to some kinds of evidence such as official documents (art. 2700). But these are the exceptions to the general rule of "prudent assessment". This rule clearly provides a general criterion of evidence assessment. Then the question is: What is the standard of proof in the Italian system of civil procedure? It is hard to say, because there is no explicit rule neither in the code nor in statutory law. Some courts have in the last years felt the need to specify that the standard of proof in matters like medical malpractice is the "more probable than not".[18] So, according to these judges, it is basically the standard used in common law countries. On the other hand, it is unclear whether the standard is conceived to apply only to such

[16] See Padoa-Schioppa (2003: 280–292).

[17] But see Damaška (2003: 129): "It is true that these lawyers developed a maze of rules about the quantity and quality of evidence needed for fact determinations. But it is a mistake to believe that these rules turned adjudicators into automatons, who made factual determination on the authority of rules, independently of their beliefs."

[18] See e.g. Cass. sez. III civile, decision n. 10285/2009, n. 10741/2009, n. 15991/2011, n. 23933/2013, n. 18392/2017.

tort matters as medical malpractice or more generally to any matter of civil law and procedure.

The situation is different if we take into consideration the Italian criminal procedure code: fact-finders shall convict only if the BARD standard is met (art. 533, as modified in 2006). The standard was introduced by Parliament some years ago with the purpose of making criminal convictions harder. And it was supported by the belief that Americans "do it better", that is, better protect criminal defendants. Some commentators replied that the idea was already implicit in the system, and that therefore that legislative intervention was superfluous.[19] But this is not our concern here. The interesting question is, given the now explicit standard, whether the code mandates any assessment criterion: in fact, as a general provision, there is just one laconic sentence according to which the judge assesses the evidence giving an account of the criteria adopted and the results obtained (art. 192, c. I). Literally construed, this provision makes any assessment criterion acceptable, provided that the judge specifies the method used and the results arrived at. Anything goes, apparently. But this would be an unsatisfying construal of the text; for the system purports to be rational and to have evidence rationally assessed. To take a somewhat extreme example, the judge could not say something like this: "I have consulted a clairvoyant and reached the conclusion that the defendant is guilty". Consulting a clairvoyant would be the method used, and the defendant's guilt would be the result arrived at. So, literally speaking, the judge would comply with the rule that requires making method and results explicit. But that would run counter to the goal of rationality. In addition, there are more definite rules concerning things like circumstantial evidence (that must be "serious, precise and consistent" to prove a fact, according to art. 192, c. II) and the declarations of the co-defendant (that shall be assessed considering the evidence that corroborates their reliability, according to art. 192, c. III). So it is interesting to point out that in Italy the criminal procedure system has an explicit standard of proof but, as a general matter, it has no explicit assessment criterion, whereas the civil procedure system has it the opposite way, namely an explicit assessment criterion and no explicit standard of proof.

Let's consider France now. According to the French code of civil procedure there is "free proof" of legal facts (see artt. 9–10, 198, 213) and "legal proof" of (certain) legal acts. No explicit standard is given, and the situation is similar to the Italian civil one.

The most interesting thing of the French system is the highly controversial notion of *intime conviction* used in criminal trials. According to the criminal procedure code, judges must decide on the basis of the evidence according to their *intime conviction* (art. 427); the same holds for mixed courts, made of judges and jurors, who are not bound by legislative rules of assessment and must ask themselves "in silence and meditation" if, given the evidence, they have an *intime conviction* about the defendant's guilt (art. 353). There has been a lot of talk about this idea. Several scholars have criticized it assuming it is nothing but an appeal to the fact-finder's

[19] See the reconstruction of the debate in Della Torre (2014: 9–11).

psychology.[20] If it were so, it would not provide any epistemic ground for the verdict and it would open the door to any form of bias, suggestion and emotive decision-making. Frankly, I take this to be an uncharitable reading of the rule. The relevant texts underscore the need to fix that state of mind *on the basis of the evidence*; it is not evidence-free, it does not boil down to the purely subjective and psychological feeling about the defendant's guilt. It is true that the appeal to "silence and meditation" strikes an Anglo-American scholar as utterly different from the idea of hot disputes between jurors, with arguments and counter-arguments being debated until a decision is made. But—let me repeat it—we should not take *intime conviction* as a legitimation of any kind of subjectively-formed or emotionally-driven decision.[21] Originally it was adopted with the idea of importing into the French system the decision-making form of English juries.[22] Given the Enlightenment trust in the judgment of lay people (and a lesser trust in the opinion of professionals like judges) the French system adopted the jury method for criminal trials and set the *intime conviction* as the check on the defendant's guilt. Eventually the idea of *intime conviction* was extended to mixed courts and to professional fact-finders such as judges.

Now, is *intime conviction* a standard of proof? Is it rather an evidence assessment criterion? Is it both? Anglo-American commentators tend to think it is a standard and a bad one, for it is simply subjective.[23] Ironically, French jurists and civil lawyers tend to think it is not a threshold, being instead a way to assess the evidence and make a decision.[24]

Spain is in a similar situation. According to the civil procedure code, evidence must be assessed according to the "rules of sound criticism" (e.g. art. 348 on expert testimony, art. 376 on witnesses). The situation is similar to that of other civil law countries in that "legal proof" is the exception, not the rule. Here the rule is assessment according to "sound criticism". Some commentators appreciate this criterion insofar as it is more epistemologically inspired than the mere idea of a "free assessment".[25] Many Spanish-speaking countries in Latin America adopt the same criterion. Such an epistemological flavor is missing instead from the Spanish code of criminal procedure, according to which judges decide assessing the evidence with their "conscience" (art. 741.1). The appeal to conscience seems to be in line with the French appeal to *intime conviction*. Similarly for other Spanish-speaking countries, but Chile uses both "sound criticism" and "beyond a reasonable doubt" in criminal cases (see art. 369-bis of the criminal code and art. 340 of the criminal procedure

[20] See e.g. Taruffo (2003b: 81–82), Laudan (2005: 98–99), Ferrer (2007: 144–145) and Bayón (2008: 17–18).

[21] Hans and Germain (2011: 755): "the decision based on *intime conviction* is not best viewed as the expression of a feeling, but rather as a considered opinion based on the charges, evidence, and defenses presented by the parties."

[22] See Padoa-Schioppa (1994).

[23] See e.g. Clermont and Sherwin (2002: 256).

[24] See e.g. Delmas-Marty (1996: 59–60) and Taruffo (2003a: 666).

[25] Lluch (2012: 194–196) and Taruffo (2003b: 82).

code).[26] The situation of Chile is interesting because it provides a confirmation of the conceptual intuition that assessment criteria and standards of proof are different things. Chile adopts "sound criticism" as the evidence assessment criterion to be used in criminal cases, and "beyond a reasonable doubt" as the standard of proof to be used there. This double discipline is not redundant, for criteria and standards are different.[27]

Finally a few words on Germany. According to the civil procedure code, the court is to decide "at its discretion and conviction" and taking into account "the entire content of the hearings and the results obtained by evidence being taken" (sec. 286.1). The emphasis on judicial discretion confirms the trend of civil law systems in civil matters. Coming to the German code of criminal procedure, the court shall decide "on the result of the evidence according to its free conviction gained from the hearing as a whole" (sec. 261). Again, the emphasis is on "free" assessment and conviction. But still on the basis of the evidence and the hearing. So, both in the civil and criminal context German law emphasizes the role of judicial discretion in evidence assessment. No standard of proof is made explicit in positive law (unless one reads "free conviction" as a standard, which is puzzling as much as the idea of having a "free standard"). But the attentive reader would have noted that those provisions concern judicial decision, not only evidence assessment. So, such evidence assessment criteria are supposed to play also the role of decision rules.

In sum, contemporary civil law countries are similar in that, with respect to civil matters, "legal proof" is a residual eventuality and "free assessment" of evidence is the rule. According to some scholars there are degrees of such freedom, for the French way is somewhat extreme in giving discretion to fact-finders, while the Spanish way puts more epistemological constraints on their judgment, and the Italian way of "prudent assessment" appears to be in the middle. Michele Taruffo has claimed that "free assessment" is not to be taken as freedom from the rules of logic, the requirements of rationality, and the constraints that science and epistemology put to the triers of fact.[28] It is simply freedom from fixed rules of evidential weight; freedom from the old system of "legal proof" (with the few exceptions that still remain). It is the idea of having evidence assessed case by case by sensible fact-finders, by working minds that are able to appreciate the particularities of concrete cases and make responsible factual determinations. Contemporary civil law countries are less similar with respect to criminal matters, since some of them (like Chile and Italy) have explicitly adopted the BARD standard of proof, while others just stick to the Continental idea of a "free" assessment of evidence.

The following tables summarize the above findings (the considered countries are in the columns[29]; in the rows "AC" stands for "assessment criterion" and "SP"

[26] See also art. 297 of the criminal procedure code, that establishes the free assessment of evidence in compliance with logic, experience and science. This provision does not mention "sound criticism" but is read as a reformulation of this traditional idea.

[27] See Accatino (2011) and Coloma (2012).

[28] Taruffo (2009: 160ff, 219). See also Gascón (2010: 32).

[29] To avoid any misunderstanding: "US" stands for the United States, "UK" for the United Kingdom, "I" for Italy, "F" for France, "S" for Spain, "C" for Chile, and "G" for Germany.

for "standard of proof"; in the cells, "?" stands for a missing or unclear aspect, "POTE" stands for the "preponderance of the evidence", "BARD" stands as usual for "beyond a reasonable doubt" and "IC" stands for *intime conviction* or a similar mental state) (Tables 1, 2).

Overall it is not an equilibrium situation, because the general discretion of evidence assessment and the lack of explicit standards in some contexts generate not only additional litigation and controversy but also confusion as to what is missing in, and prescribed by, a given legal system (like the civil procedure one in Italy, for instance).

As an additional point, note that the scope of argumentation performed by parties and fact-finders about the evidence is significantly reduced when fixed criteria of evidence assessment are mandated. This was pretty obvious in the old system of "legal proof", where any discussion about the witnesses' credibility or the circumstances of testimony was beside the point once the fact-finders were bound by some measure like the two-witness rule. The more discretion fact-finders have, the more prominent is the role of legal argumentation about evidence and the litigated facts. Parties are supposed to provide persuasive arguments about evidence and fact-finders are supposed to provide arguments that justify their decisions.[30] This is the case in particular when judges have to provide written opinions that (supposedly) justify the decisions they made. *Intime conviction* is frequently ridiculed as epistemologically unsound, but consider how frustrating it is to be bound by fixed criteria that impede any appreciation of the particular circumstances of the case and block any argumentative effort to make sense of the evidentiary items at disposal.

Before moving on, remember that fixed criteria are not to be equated with explicit ones: the fixed criteria of the "legal proof" system were of course explicitly mandated by authoritative legal sources, but these can also mandate relaxed criteria that leave discretion to fact-finders.

4 Some analysis and questions

A conceptual analysis of the foregoing remarks focuses on the relevant systemic differences. Some legal systems have explicit standards of proof (as decision rules) and lack explicit assessment criteria. Some other systems have explicit assessment criteria and lack explicit standards of proof (again, understood as decision rules). In addition it is hard to say whether the non-explicit things are implicit or undetermined. For instance, it is *prima facie* hard to say whether in the Italian civil procedure system there is an implicit standard of proof or none.

If all of that is correct, two further research questions arise after the one we posed above (in Sect. 2) concerning the confusion between criteria and standards in scholarly work and judicial opinions. Here's the second question, simply put: Why do systems differ? That is, why do some of them have assessment criteria and lack

[30] The literature is abundant. See, among others, Anderson et al. (2005), Verheij (2014) and Verheij et al. (2016).

Table 1 Civil cases

	US	UK	I	F	S	C	G
AC	?	?	Prudent	Free	Sound criti-cism	Sound criti-cism	Free
SP	POTE	POTE	?	?	?	?	?

Table 2 Criminal cases

	US	UK	I	F	S	C	G
AC	?	?	Free	IC?	Conscience	Sound criticism	Free
SP	BARD	Certainty	BARD	IC?	?	BARD	IC?

standards of proof, while others have it the other way round? And now the third question: How can systems work without one of such elements? That is, if assessment criteria and standards of proof are necessary components of legal fact-finding, how can it be the case that a system which lacks one of them can work nevertheless?

The second of our three research questions mainly finds its answer in history. The above differences between systems (and the relative confusion in scholarly works and judicial opinions) are generated by different historical, procedural and institutional concerns. The present paper is not a piece in legal history, and the present Journal is not devoted to it, so I won't expand on this point.[31] Suffice it to say that the common law world shows a prominent concern for standards of proof and this finds an explanation in the features of jury trial. The fact that the triers of fact are lay people, who need instructions from judges but don't give written reasons for the decisions they make, militates in favor of a simple system were fact-finders have to consider a standard of proof and can be instructed by judges about the meaning of the standard and the ways to assess whether the standard is met. This facilitates the task of the jury. On the other hand, the civil law world shows a prominent concern for evidence assessment and this finds an explanation in the features of trials with professionals like judges who once were supposed to apply the complex system of "legal proof" and eventually, in contrast with the previous binding system, were asked to freely assess the evidence.[32]

The puzzling nature of *intime conviction* is less puzzling if we consider it in the perspective of this historical process. It was introduced in France to emulate the English way to decide criminal cases with lay jurors and eventually it was extended to professional judges freed from the complexities of the "legal proof" system; so it is Janus-faced in that it reflects different legal cultures and traditions and because it has absorbed, so to say, the different concerns of these cultures and traditions.

[31] I refer to Damaška (1986: 29ff) and to Jackson and Summers (2012: 57ff) among others.

[32] Still, many discussions on things like "moral certainty" show that the concern for a decision threshold was also present on the Continent. See among others Pagano (1828: 59ff) and Ippolito (2008: 218–220). Cf. Whitman (2008).

That also explains some conceptual confusion about evidence and proof. Civil lawyers often take standards as assessment criteria because they are more familiar with the latter (consider again the ECtHR decision quoted above). And vice versa common lawyers often take assessment criteria as standards because they are more familiar with the latter (think again of the widespread critique to the French *intime conviction*). But consider that those views are not entirely unjustified as far as there are some functional connections between criteria and standards: some standards incorporate assessment criteria and vice versa. Consider again BARD: it is arguable that the reasonable doubt standard incorporates the idea of a reasonable assessment of the evidence (which is not very informative, but still a form of guidance). Similarly for *intime conviction*: if it is an assessment criterion, it incorporates the threshold of the fact-finder's conviction as a decision rule.

These last remarks brings us to my tentative answer to the third research question specified above. *There is some connection* between criteria and standards, notwithstanding their conceptual difference. What kind of connection? My guess is that there is a *functional connection* between them. Now a functional connection must be distinguished from a sort of functional identity or equivalence. The idea of a *functional equivalence* would consist in the fact that standards and criteria perform the same procedural and systemic function. This idea has to me some intuitive interest, for it would explain away the puzzle of systems working fairly well without one of the apparently necessary components of factual decision-making. Remember what the question was: How can systems work without one of such elements, if we assume that assessment criteria and standards of proof are necessary components of legal fact-finding? The functional equivalence idea would explain the puzzle away because it would claim that criteria and standard perform the same function. So there would be no mystery in the fact that systems which have one but not the other component work fairly well. It would be unnecessary to have both, it would be redundant. Because they perform the same function, namely the function of guiding and justifying factual decision-making—or, even more simply (if guiding and justifying appear to be different things), the function of justifying fact-finding.

The functional equivalence idea is interesting but has some drawbacks. One is the (likely annoying) generality of the function appealed to: "guiding and justifying factual decision-making" is remarkable for its lack of specificity—even more so for the simpler idea of fact-finding justification. But it's true that functional accounts always abstract from many details, insofar as they consider that such details are not relevant to the performance of the function envisaged (if the function is writing you don't bother about the differences between pens and pencils, which you bother about if you need devices which draw signs that cannot be canceled).[33] A further and more significant drawback is the failure of the functional equivalence idea to make sense of some interesting cases. What about the Italian system of civil procedure? As we said, it has an explicit assessment criterion but no explicit standard. And there is an increasing debate about the standard or the standards that judges should use. If criteria and standards were functionally the same, this debate would be meaningless

[33] On functional accounts of law see Green (1998) and Ehrenberg (2016).

because the problem would not subsist. On the contrary there is a debate because it is unclear whether there is a standard and what this standard might be, whereas no one has doubts about the existence of an assessment criterion (art. 116 of the civil procedure code) however vague it may be. The existence of this assessment criterion doesn't solve by itself the standard of proof problem. So, there is no equivalence. Moreover, what about the Chilean system of criminal procedure? As we said, in this system you find both a standard (BARD) and an assessment criterion ("sound criticism"). In this respect, assuming the functional equivalence approach, the system would be redundant. This conclusion would follow from the idea that the two perform the same function. It would be redundant to have a standard once you have a criterion, or vice versa, because they perform the same function. Now, this unreasonable or at least uncharitable reading of that kind of system militates against the functional equivalence approach.

A more promising idea is that of a functional connection between criteria and standards: *criteria indicate how to assess evidence and standards indicate how to make a decision once the evidence at disposal has been assessed*. A system like the Chilean code of criminal procedure makes everything explicit in requiring an assessment according to the principle of "sound criticism" and a decision according to the BARD standard. The systems which are less explicit have just one component stated in positive law; then I see two possibilities: (1) that the other component is somehow incorporated in the explicit one; (2) that the other component is undetermined.

According to the first scenario, there can be explicit standards of proof and implicit assessment criteria, or vice versa explicit criteria and implicit standards. The incorporation idea would consist in the fact that some system implicitly conveys a certain criterion or standard, through the explicit component and the rest of the normative system (its rules, principles, etc.). I would locate the Italian criminal procedure system in this category, since it is characterized by an explicit standard (the BARD one) and an implicit assessment criterion consisting in a free but rational assessment of the evidence. Similarly, common law systems that adopt the BARD standard appear to convey at the same time the requirement of a reasonable assessment of the evidence; judicial instructions to juries perfectly fit this picture.[34] Similarly, "legal proof" criteria usually incorporated standards of proof in that they set evidence thresholds and decision rules. Consider again the traditional two-witness rule, according to which two independent testimonies of the same amounted to "full proof" of the fact: it was a rule about the value of testimonial evidence and, at the same time, it set a threshold for decision.

According to the second scenario, there can be systems where one of the necessary components of legal fact-finding is made explicit and the other is undetermined. What happens then in such systems? The analytical gap is filled by *legal practice*, with the possible support of legal theory and scholarly opinions. Legal practitioners and participants in legal disputes take sides about the undetermined component; arguments are developed and decisions are made; and in the long run, if no legislation intervenes, some component entrenched in legal practice

[34] Cf. on this point Picinali (2015).

will emerge as dominant through widely accepted arguments, judicial precedents, etc. The gap will be filled by legal practice. And in this kind of context the importance of legal argumentation is greater than elsewhere in this respect because legal practitioners and scholars need to construct and develop persuasive arguments to fill in that gap. As an example, I would locate the Italian civil procedure system in this category, since it is characterized by an explicit assessment criterion and the lack of a positive standard of proof. This system differs from the criminal procedure system of Italy because in the latter, according to most commentators and practitioners, there is an implicit component (the implicit assessment criterion); as to the former, on the contrary, no general agreement exists about the component which is not explicit (the standard of proof); therefore it is arguable that this component is undetermined and it is up to legal practice to fill in the gap. Recent judicial opinions in Italy confirm this reading. In fact, one might ask the following: How could it be the case that Italian civil judges have been deciding disputes over the years if the standard of proof, being a necessary component of fact-finding, were completely undetermined? Decisions would have been impossible. However, this *reductio ad absurdum* of my argument would be unsound because it does not take into account the gap-filling role of practitioners, attorneys and judges in particular: even before they used the Anglo-American vocabulary and conceptual apparatus of standards of proof, for sure Italian legal practitioners had in their mind some idea about the evidence threshold that justifies a certain civil decision, for it is plausible to think that they had some intuition about such sufficiency threshold. When is the evidence sufficient to justify a certain belief about the litigated facts? When is it sufficient for the acceptance of a certain version of the litigated facts? You need not have the vocabulary of the "standards of proof" to grasp this idea and the problem. Then solutions come with legal practice and argumentation.

So, in some way or other, both evidence assessment criteria and standards of proof are conceptually needed. Some might think, though, that current legal theory is "forcing" the practice of civil law systems. It is putting pressure on legal practitioners, so the arguments goes, to compel them to use the standards vocabulary. To address this remark, recall some of the facts. Why do civilians not usually discuss standards? Because they do criteria. Why do common lawyers not discuss assessment criteria? Because they do standards. To have an explicit standard is, most of the time, to have an implicit criterion. And to have an explicit criterion is, most of the time, to have an implicit standard. Therefore theory does not force practice if it makes explicit what is implicit in the system because of some functional connection. And theory does not force practice if it tries to give an accurate account of its gap-filling role when a necessary component of fact-finding is undetermined.

In sum, the functional connection account better explains how systems work without one of such analytically necessary components of factual decision-making. Or, better, without one such component being made explicit in positive law. Systems work nevertheless either because there is some incorporation of the lacking element in the explicit one, or because legal practice fills the gap when one such component is undetermined.

5 Conclusion

The first question we asked (in Sect. 2) was why some jurists and courts do not distinguish between assessment criteria and standards of proof. The second question (in Sect. 4) was why some systems have criteria and others have standards. And the third question (again in Sect. 4) was how systems work without one of such necessary components of legal fact-finding.

The answer to the first question came with the answer to the second. Some historical, institutional and procedural differences explain why some of the systems have assessment criteria and others have standards. This is basically the answer to the second question above. And this answer suggests that jurists belonging to one context are prone to see also elsewhere what they are more familiar with. This means that common lawyers have some tendency to take assessment criteria as standards and vice versa civil lawyers have some tendency to see standards as assessment criteria. The less familiar thing is understood in terms of the more familiar one. More importantly, when different systems come to overlap to a certain extent (as is the case in the jurisdiction of courts like the ECtHR) parties and judges who come from different contexts and backgrounds run the risk of mixing things up. Some terms, concepts and institutions that belong to one context are used in combination with terms, concepts and institutions of a different context. Some merging of the issues (assessment and standards) comes from the work of supranational courts such as the ECtHR: they use some traditional common law terminology and conceptual apparatus in dealing with civil law countries issues. Not surprisingly, the result can be confused and confusing to a certain extent. This is one of the prices of a more integrated or even globalized legal world, where different concepts and institutions play together and interact in ways that are not always foreseeable.[35]

The answer to the third question above is more theoretically intriguing in my view. It consists in the claim that there is a functional connection between assessment criteria and standards of proof. Because of that connection, one element can incorporate the other. And because of that connection, where one of those things is neither made explicit in positive law nor incorporated in the explicit one, legal practice and scholarly literature make an effort to fill in the relevant gap, providing answers and solutions to the questions and problems that actual cases pose to legal practitioners and scholars.

Acknowledgements I wish to thank two anonymous reviewers of this Journal, for their very detailed and helpful comments on an earlier version of this work.

References

Accatino D (2011) Certezas, dudas y propuestas en torno al estándar de la prueba penal. Revista de Derecho de la Pontificia Universidad Católica de Valparaíso 37(2):483–511

[35] For a similar point see Hans and Germain (2011: 738, 761–763).

Allen RJ (2014) Burdens of proof. Law Probab Risk 13:195–219
Allen RJ, Pardo M (2007) The problematic value of mathematical models of evidence. J Leg Stud 36(1):107–140
Allen RJ, Stein A (2013) Evidence, probability, and the burden of proof. Ariz Law Rev 55:557–602
Anderson T, Schum D, Twining W (2005) Analysis of evidence, 2nd edn. Cambridge University Press, Cambridge
Barzun C (2008) Rules of weight. Notre Dame Law Rev 83(5):1957–2017
Bayón JC (2008) Epistemología, moral y prueba de los hechos: hacia un enfoque no benthamiano. Analisi e diritto 2008:15–34
Bell RS (1987) Decision theory and due process: a critique of the Supreme Court's lawmaking for burdens of proof. J Crim Law Criminol 78:557–585
Caprioli F (2009) L'accertamento della responsabilità penale "oltre ogni ragionevole dubbio". Rivista italiana di diritto e procedura penale 52:51–92
Cheng E (2013) Reconceptualizing the burden of proof. Yale Law J 122(5):1254–1279
Clermont K, Sherwin E (2002) A comparative view of standards of proof. Am J Comp Law 50:243–275
Cohen LJ (1977) The probable and the provable. Clarendon Press, Oxford
Coloma R (2012) ¿Realmente importa la sana crítica? Revista Chilena de Derecho 39(3):753–781
Damaška M (1986) The faces of justice and state authority. A comparative approach to the legal process. Yale University Press, New Haven
Damaška M (2003) Epistemology and legal regulation of proof. Law Probab Risk 2:117–130
Della Torre J (2014) Il lungo cammino della giurisprudenza italiana sull'"oltre ogni ragionevole dubbio". Diritto Penale Contemporaneo. https://www.penalecontemporaneo.it. Accessed 20 June 2014
Delmas-Marty M (1996) La preuve pénale. Droits 23:53–65
Ehrenberg KM (2016) The functions of law. Oxford University Press, Oxford
Ferrajoli L (1989) Diritto e ragione. Teoria del garantismo penale. Laterza, Roma-Bari
Ferrer J (2007) La valoración racional de la prueba. Marcial Pons, Madrid
Gascón M (2010) Los hechos en el derecho. Bases argumentales de la prueba, 3rd edn. Marcial Pons, Madrid
González Lagier D (2013) Quaestio facti. Ensayos sobre prueba, causalidad y acción. Fontamara, Mexico
Green L (1998) The functions of law. Cogito 12(2):117–124
Haack S (2014) Evidence matters. Science, proof, and truth in the law. Cambridge University Press, Cambridge
Hans V, Germain C (2011) The French jury at a crossroads. Chic Kent Law Rev 86(2):737–768
Iacoviello FM (2006) Lo standard probatorio dell'al di là di ogni ragionevole dubbio e il suo controllo in Cassazione. Cassazione penale 46:3869–3884
Ippolito D (2008) Mario Pagano. Il pensiero giuspolitico di un illuminista. Giappichelli, Torino
Jackson JD, Summers SJ (2012) The internationalisation of criminal evidence. Beyond the common law and civil law traditions. Cambridge University Press, Cambridge
Kaplan J (1968) Decision theory and the factfinding process. Stanf Law Rev 20:1065–1092
Kaplow L (2012) Burden of proof. Yale Law J 121:738–859
Kaye DH (1986) Do we need a calculus of weight to understand proof beyond a reasonable doubt? Boston Univ Law Rev 66:657–672
Laudan L (2005) Por qué un estándar de prueba subjetivo y ambiguo no es un estándar. Doxa 28:95–113
Laudan L (2006) Truth, error, and criminal law. An essay in legal epistemology. Cambridge University Press, Cambridge
Lillquist E (2002) Recasting reasonable doubt: decision theory and the virtues of variability. U C Davis Law Rev 36:85–197
Lluch XA (2012) La dosis de la prueba: entre el common law y el civil law. Doxa 35:173–200
McNaughton JT (1955) Burden of production of evidence: a function of a burden of persuasion. Harv Law Rev 68:1382–1391
Padoa-Schioppa A (1994) La giuria penale in Francia. LED, Milano
Padoa-Schioppa A (2003) Italia ed Europa nella storia del diritto. il Mulino, Bologna
Pagano FM (1828) Principj del codice penale e Logica de' probabili. Raffaello di Napoli, Napoli
Picinali F (2013) Two meanings of "reasonableness": dispelling the "floating" reasonable doubt. Mod Law Rev 76(5):845–875
Picinali F (2015) The threshold lies in the method: instructing jurors about reasoning beyond reasonable doubt. Int J Evid Proof 19:139–153
Putnam H (1981) Reason, truth and history. Cambridge University Press, Cambridge

Putnam H (1995) Are moral and legal values made or discovered? Leg Theory 1:5–19
Redmayne M (1999) Standards of proof in civil litigation. Mod Law Rev 62:167–195
Redmayne M (2003) Objective probability and the assessment of evidence. Law Probab Risk 2:275–294
Roberts P, Zuckerman A (2010) Criminal evidence, 2nd edn. Oxford University Press, Oxford
Taruffo M (1992) La prova dei fatti giuridici. Giuffrè, Milano
Taruffo M (2003a) Rethinking the standards of proof. Am J Comp Law 51:659–677
Taruffo M (2003b) Algunos comentarios sobre la valoración de la prueba. Discusiones 3:81–97
Taruffo M (2009) La semplice verità. Il giudice e la costruzione dei fatti. Laterza, Roma-Bari
Vázquez C (ed) (2013) Estándares de prueba y prueba científica: ensayos de epistemología jurídica. Marcial Pons, Madrid
Verheij B (2014) To catch a thief with and without numbers. Arguments, scenarios and probabilities in evidential reasoning. Law Probab Risk 13:307–325
Verheij B et al (2016) Arguments, scenarios and probabilities: connections between three normative frameworks for evidential reasoning. Law Probab Risk 15:35–70
Whitman J (2008) The origins of reasonable doubt. Theological roots of the criminal trial. Yale University Press, New Haven