

The civil standard of proof—what is it, actually?

The International Journal of
Evidence & Proof

2016, Vol. 20(3) 217–234

© The Author(s) 2016

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1365712716645227

eji.sagepub.com



Mark Schweizer

Max Planck Institute for Research on Collective Goods, Germany
University of St. Gallen, Switzerland

Abstract

Unlike common law, Continental European civil law does not strictly distinguish between the standards of proof applicable in civil and criminal matters, respectively. In civil law countries such as Germany and Switzerland, judges are supposed to use the same (high) standard of ‘full conviction’ in both criminal and civil cases. This study is the first to look at the standard of proof *actually used* by judges and judicial clerks in a civil law country (Switzerland). It is shown that the standard actually used differs from the one normatively prescribed. No difference between the estimated decision threshold for members of the courts and members of the general population is found. The results suggest that the standard of proof actually employed by judges of a civil law system is not much different from the common law’s ‘preponderance of the evidence’ standard, despite the doctrinal insistence to the contrary.

Keywords

comparative law, decision theory, empirical legal study, preponderance of the evidence, standard of proof

Introduction

Common law knows (at least) two different standards of proof, the ‘preponderance of the evidence’—or ‘balance of probabilities’ in English law—for civil cases and the ‘proof beyond reasonable doubt’ in criminal cases.¹ In US law, a further intermediate standard of proof known as ‘clear and convincing evidence’, which is applicable in certain civil cases (e.g. civil fraud), is well-established,² while it is a matter of controversy whether English law recognises such an intermediate standard of proof (Anderson et al., 2006: 243; McBride, 2009: 325ff).

1. See *Addington v Texas*, 441 U.S. 418 (1979), 422, 423; for English law *In Re H & Others* (minors) [195] UKHL 16, AC [1995] 563, s. 76; Wright (2009: 80).

2. See e.g. *Addington v Texas* 441 U.S. 418 (1979) at 422, 423.

Corresponding author:

Mark Schweizer, Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn, Germany.
Email: mark.schweizer@gmail.com

Unlike common law, Continental European civil law does not generally distinguish between standards of proof for civil and criminal matters (Engel, 2009: 435; Motsch, 2009: 242). Standard of proof is always the (full) conviction of the judge, be it a *conviction intime* or a *conviction raisonnée*, a reasoned or reasonable conviction (meaning that the judge must justify his or her decision by valid arguments).³ The insistence of civil law that the standard of proof in civil cases is ‘full conviction’, meaning ‘near certainty’—notwithstanding many exceptions—has left common law lawyers puzzled. In a strongly worded article, Kevin M Clermont and Emily Sherwin ‘rudely wonder[ed] how civilians can be so wrong’ (Clermont and Sherwin, 2002: 244). The article has met with an equally strongly worded rebuttal (Taruffo, 2003), but also a renewed interest of civilians in standards of proof (Engel, 2009; Kinsch, 2009; Motsch, 2009). Clermont and Sherwin have not, however, been convinced that their original analysis, which came to the conclusion that civil law values the perceived legitimacy of the court system higher than a rational approach to judicial decision making, was wrong (Clermont, 2009).

Scholars have wondered for a long time whether civilian judges actually adhere to the high standard proclaimed by case law and doctrine (Clermont and Sherwin, 2002: 261; Einmahl, 2001: 474ff.; Rechberger, 1990: 490; Zamir and Ritov, 2012: 172). So far, no empirical study tried quantifying the standard of proof *actually employed* in civil matters by judges in a civil law country. This paper reports the results of the first such study. It demonstrates that the standard of proof that Swiss judges and judicial clerks proclaim to adhere to when asked directly is much higher than the standard of proof that would result if the decision threshold was chosen to minimise the expected error costs of the decision *even when the error costs are obtained from the subjects*. It also estimates that there is a 50% probability that a request in a civil action will be granted when the judge is convinced only to a degree of 63% that the factual allegations supporting the claim are true. This decision threshold is no different from that estimated for a sample of the general German population. It is also comparable to the decision threshold of 70 estimated for Israeli trial lawyers by Zamir and Ritov, although Israel adheres to the common law’s ‘preponderance of the evidence’ standard in civil matters.⁴ The results suggest that the standard of proof actually employed by courts in Switzerland, a traditional civil law country, is much lower than the standard proclaimed by the doctrine and case law.

The rest of this paper is structured as follows: First, the different standards of proof in civil matters by common law and by Continental European civil law are exposed. Subsequently, three different methods of measuring the standard of proof are described. In the next section, the hypotheses to be tested are stated. This is followed by a description of the method of the study and the sample of Swiss judges and judicial clerks participating in the study. The results of the study are then reported, followed by a discussion of them.

Different standards of proof in civil matters in common law and Continental European Law

The ‘preponderance of the evidence’ standard of US law is explained in the Federal Jury Practice and Instructions as follows (O’Malley et al., 2001: § 166.51):

To ‘establish by a preponderance of the evidence’ means to prove that something is more likely so than not so. In other words, a preponderance of the evidence in the case means such evidence as, when considered and compared with that opposed to it, has more convincing force, and produces in your minds belief that what is sought to be proved is more likely true than not true. This rule does not, of course, require proof to an absolute certainty, since proof to an absolute certainty is seldom possible in any case.

3. The standard of proof in Germany is better described as a ‘conviction raisonnée’ rather than the French ‘conviction intime’; see Deppenkemper (2004: 208ff., 421 and the references cited therein).

4. Zamir and Ritov (2012: 177). Zamir and Ritov did not use a percentage scale, but rather a scale from 0 to 100 (without indication of percentages).

For English law, a definition by Lord Denning in *Miller v Minister of Pensions* is the most cited formulation of the civil ‘balance of probabilities’ standard. ‘If the evidence is such that the tribunal can say “we think it more probable than not” then the burden is discharged, but if the probabilities are equal it is not.’⁵ If the fact finder is inclined to believe the plaintiff more than the defendant, even to the slightest degree, then he or she must find for the plaintiff.⁶ In other words, it is sufficient if the plaintiff’s allegations are more probably true than not.⁷ On the other hand, to reach a guilty verdict in a criminal case, the jury must be convinced beyond any reasonable doubt that the facts alleged by the prosecution are true. A ‘reasonable’ doubt is one that is based upon reason and not purely on speculation, a merely possible doubt does not prevent a finding against the defendant.⁸

Traditionally, the doctrinal discussion has emphasised that the standard of proof in civil matters is expressed by reference to the evidence or probabilities (‘objectively’), while in criminal matters it is expressed by reference to the state of mind of the fact finder (‘convinced beyond reasonable doubt’) (Anderson et al., 2006: 242). However, as the jury instructions for the civil standard of the preponderance of the evidence—cited above—show, in civil cases, too, the belief or state of mind of the fact finder is what counts (‘produces in your minds *belief* that what is sought to be proved is more likely true than not true’). The difference between the criminal and civil standard of proof lies in the *degree of belief* (or conviction) required for finding for the party bearing the burden of proof. This degree of belief is much higher in criminal cases than it is in civil cases.

It has long been suggested that normative decision theory provides an elegant explanation for the different standards (the seminal papers are Kaplan (1968) and Cullison (1969)): since wrongly convicting an innocent person is widely considered to be a graver mistake than erroneously acquitting a guilty person, the expected error costs are minimised if the standard of proof in criminal cases is well above 50% (whether it can be quantified at all is highly controversial; see the references cited in Tillers and Gottfried (2007)), but nobody would dispute that a civil jury may find for the plaintiff under circumstances that would not permit a criminal jury to convict the accused). On the other hand, it is a commonly held assumption that in civil cases, the disutility of erroneously finding for or against the plaintiff is similar, which means the error-cost minimising decision threshold is $\geq 50\%$.⁹ As one commentator put it, ‘civil cases are the paradigm for symmetrical error costs’ (Lee, 1997: 25). Therefore, ‘Bayesian decision theory seems to provide a pleasing and harmonious interpretation of civil litigation’s usual requirement of proof by a preponderance of the evidence’ (Kaye, 1987: 55).

Unlike common law, Continental European civil law does not generally distinguish between standards of proof for civil and criminal matters (Engel, 2009: 435; Motsch, 2009: 242). Standard of proof is always the (full) conviction of the judge, be it a *conviction intime* or a *conviction raisonnée*, a reasoned or reasonable conviction (meaning that the judge must justify his or her decision by valid arguments).¹⁰

-
5. *Miller v Minister of Pensions*, 3 All ER 372 (1947) at 373ff.
 6. Redmayne (1999: 172); for US law *Livanovitch v. Livanovitch*, 131 A. 799, 800 (Vt. 1926) (‘If [. . .] you are more inclined to believe from the evidence that he did so deliver the bonds to the defendant, even though your belief is only the slightest degree greater than that he did not, your verdict should be for the plaintiff’ (quoting the jury instructions); Pennsylvania Suggested Standard Civil Jury Instructions, 3rd ed. 2005, § 1.42.
 7. Illinois Supreme Court Committee on Pattern Jury Instructions in Civil Cases (eds.), Illinois Pattern Jury Instructions: Civil, § 21.01 (‘more probably true than not true’), available from www.state.il.us/court/CircuitCourt/CivilJuryInstructions/21.00.pdf (accessed 1 April 2016); Sand et al. (2007), Bd. 4, § 73.01, Instruction 73–2 (‘by a preponderance of the evidence’ means ‘more likely true than not true’).
 8. Jury instructions according to the Ninth Circuit Model Criminal Jury Instructions, 2003 edition, § 3.5—Reasonable Doubt—Defined. Sheppard (2003) shows the development from ‘moral certainty’ to ‘reasonable doubt’ to ‘articulate doubt’, which, according to him, explains the current practice in criminal law better.
 9. *In re Winship* 397 U.S. 358, 371 (1970) (Harlan J. Concurring); Ball (1960: 817); Kaye (1987: 72); Lee (1997: 25); Posner (1999: 1504); Redmayne (1999: 171); Clermont and Sherwin (2002: 268); Stein (2005: 148); Zamir and Ritov (2012: 189); but see Tyree (1982: 93ff.)
 10. The standard of proof in Germany is better described as a ‘conviction raisonnée’ rather than the French ‘conviction intime’; see Deppenkemper (2004: 208ff., 421 and the references cited therein).

This standard is described in the leading case of the German Federal Supreme Court (Bundesgerichtshof) as a ‘personal conviction [...] in doubtful cases, the judge may and must be content with a degree of certainty useful for practical life that silences doubt without completely excluding it.’¹¹ The Swiss Federal Supreme Court (Bundesgericht) uses a similar definition according to which ‘a court must be convinced of the truth of a factual allegation based on objective grounds. Absolute certainty is not required. It is sufficient if the court has no serious doubt or any remaining doubt appears insubstantial.’¹² Neither German nor Swiss courts have ever expressed the decision threshold as a (quantified) subjective probability. The traditional doctrine is also reluctant to do so, but when it does quantify the standard of proof, the decision threshold is said to be above 90%;¹³ sometimes figures of 95%¹⁴ or even 99.8%¹⁵ are given.

There are, certainly in German and Swiss law, many exceptions to the standard of full conviction in civil cases, namely for allegations that are notoriously difficult to prove, such as causality in medical malpractice or the theft of an insured item to be proven by the policy holder.¹⁶ In summary proceedings, such as preliminary injunction proceedings, an even lower standard of ‘prima facie evidence’ (Glaubhaftmachung) is used, which is met if the factual allegations supporting the claim are more probably true than not.¹⁷ But the exceptions—and the considerable doctrinal effort required for their justification—prove the rule (Walter, 1979: 184). Few would dispute that the degree of conviction required for finding for the plaintiff in civil law is much higher than that required by the common law’s ‘preponderance of the evidence’ or ‘balance of probabilities’ standard.¹⁸ In Germany in the 1970s, a number of scholars influenced both by (possibly misunderstood) Scandinavian doctrine and the decision theoretic framework of Kaplan have suggested introducing a general standard of a ‘balance of probabilities’ in civil matters (Kegel, 1967: 335; Maassen, 1975: 5ff.; Motsch, 1978: 335ff.; Nell, 1983: 211; more recently Wagner, 2009: 172ff.). These propositions have been met with almost visceral rejection and are today widely considered heterodox (Prütting, 2010: 142). This scholarly debate is further evidence that, at least in theory, the civil standard of proof in Germany is indeed (much) higher than in US or English law. I am inclined to believe that this also holds for other civil law countries, but Taruffo does make the valid point that it is dangerous to generalise based on a few examples (Taruffo, 2003: 660).

Measuring standards of proof

Several methods of measuring standards of proof as subjective probabilities are known (Dane, 1985: 143ff.). The direct rating or self-report method consists of simply asking the subjects to quantify the minimal threshold they require for a guilty verdict (or a grant of the request) on a meaningful numerical scale (Dane, 1985: 143; Hastie, 1993: 101). Among the indirect methods, two approaches are distinguished: the decision-theoretic approach and the parallel-ranking or rank-order method (Dane, 1985: 143; Hastie, 1993: 101). It is known that the three methods do not lead to the same decision thresholds (Dhami, 2008; Hastie, 1993: 102ff.).

11. BGHZ 53, 245 = NJW [Neue Juristische Wochenschrift] 1970, 946 (translation from German by the author).

12. DFT [Decisions of the Federal Tribunal] 130 III 321 sect. 3.2 (translation from German by the author).

13. For German law, Kadner Graziano (2011: 189); for Swiss law, Berger-Steiner (2008: s. 6.81); Walter (2009: 53); Bühler (2010: s. 9).

14. For German law Greger (1978: 110); for Swiss law Summermatter and Jacober (2012: 142).

15. For German law Bender (1981: 258); Fuchs (2005: 80).

16. See e.g. for German law BGH NJW 1995, 2169; NJW 2004, 777; for Swiss law DFT 130 III 321 s. 3.3; 132 III 715 s. 3.2.

17. For German law Scherer (1996: 39ff.); for Swiss law Berger-Steiner (2008: s. 6.155).

18. But see Gottwald (2000: 175) and Brinkmann (2005: 3), who argue against any difference in principle between the German and the common law’s standard of proof in civil cases.

Direct rating or self-report method

The direct rating or self-report method is useful for assessing whether different verbal definitions of the standard of proof, e.g. in jury instructions, are actually perceived as requiring different subjective probabilities of guilt for a conviction. The problem with using this approach with sophisticated subjects such as judges is that the judges know the theoretically required standard of proof, and they will likely give the answer that is expected and not the threshold actually used. Obtaining a decision threshold by the direct method is still useful for comparison with the decision thresholds obtained using other measurement methods, but it is not suitable to answer the question, relevant here, of what *actual* standard of proof the judges employ.

Decision theory-based method

The first of two indirect methods is called the decision theory-based method.¹⁹ It obtains from the subjects the disutilities (or costs) required to parametrise the following inequation

$$\Pr(p) \geq \frac{(D_{fp} - D_{cn})}{(D_{fn} - D_{cp}) + (D_{fp} - D_{cn})} = \frac{1}{1 + \left(\frac{D_{fn} - D_{cp}}{D_{fp} - D_{cn}}\right)} \quad (1)$$

$\Pr(p)$ is the probability that the plaintiff's (or prosecution's) allegations are true. The expected costs are minimised when $\Pr(p)$ meets or exceeds the value calculated according to the above equation. D_{fp} is the disutility of a false positive decision, i.e. convicting an innocent person or granting an unfounded claim; D_{fn} is the disutility of a false negative, i.e. acquitting a guilty person or denying a well-founded claim; D_{cp} is the disutility of a correct positive, i.e. convicting the guilty or granting a well-founded claim; and D_{cn} is the disutility of a correct negative, i.e. acquitting the innocent or denying an unfounded claim. The equation can be rewritten for utilities instead of disutilities, but most legal scholars work with disutilities or costs.

Estimating the decision threshold using the (dis)utilities associated with each possible outcome from the subjects reliably leads to lower decision thresholds than those elicited using the direct method. Even in criminal cases, the decision thresholds calculated using the decision theory-based method often barely exceed 50%.²⁰ For my purposes, the problem with the decision theory-based method is that it only allows the estimation of a *normative* decision threshold, i.e. where the threshold would have to be in order to maximise utility (minimise disutility) given the subject's expressed (dis)utilities. But it again does not measure where the threshold *actually* is. Still, obtaining the parameters necessary to estimate the normative decision threshold from sitting judges is interesting in and of itself, because the judges may or may not share the common belief that in civil cases, the error costs of a false positive and a false negative decision are equal.

Binary logistic regression

This study estimates the decision threshold using a binary logistic regression with the 'grant' decision as dependent variable and the expressed subjective probability as independent variable. Both dichotomous decision (for or against plaintiff) and subjective probability of the plaintiff's allegations being true were obtained from the same subjects (within-subject design). To control for order effects,²¹ the order of questions was randomised.

19. Terminology according to Hastie (1993: 103).

20. See values reported in table 4.3 in Hastie (1993: 105); Dhimi (2008: 360).

21. Some studies found that subjects who first made a dichotomous choice gave higher subjective probabilities than subjects who first gave their probability rating and then made the dichotomous choice (Simon and Mahan, 1971: 322); but see Dane (1985: 149ff.), who did not find any order effect.

Table 1. Illustration of the parallel-ranking method.

Guilty?	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Subj. probability	77	76	76	75	75	71	70	69	66
Rank	1	2	3	4	5	6	7	8	9

The binary logistic regression allows estimating the degree of belief at which there is a 50% probability that a randomly chosen subject will grant the request, i.e. decide for the plaintiff. The advantage of this method is that it estimates the degree of belief *actually required* for a probability of x% that the subject will grant the claim, rather than the threshold explicitly stated by the participants, which may be subject to demand effects.

Another advantage is that the results of the binary logistic regression may be compared to the results of other studies obtained using the parallel-ranking method. The parallel-ranking (also rank-order) method uses an in-between subject design (Simon, 1970). Half the subjects express their belief in the guilt of the accused as a subjective probability, the other half only makes a dichotomous guilty-innocent verdict judgment. The subjective probabilities are ranked from highest to lowest, and the minimum threshold for a guilty verdict is determined by counting down the probability ratings to the rank number corresponding to the proportion of guilty verdicts obtained in the dichotomous decision condition (in the illustrative example given in Table 1, the minimum value obtained would be 70).

Arguably, the 50% probability of grant computed using a binary logistic regression conforms to the mid-point between the upper and lower threshold obtained using the parallel-ranking method, because at this point the best one can say (based on the rank order method) is that there is a 50% probability that the request will be granted. This allows a comparison of the data obtained in the present study with the data obtained in earlier studies that used the parallel-ranking method (namely Zamir and Ritov, 2012).

Hypotheses to be tested

Based on the previous research on different elicitation techniques for the standard of proof, I expect to observe a higher decision threshold when using the direct rating method than when using the decision theory-based method. I also expect that the actual decision threshold is lower than the one stated overtly, which may be subject to demand effects. This hypothesis is tested using a logistic regression with the decision threshold as dependent variable, which allows exploring where the actual decision threshold lies, rather than the one stated explicitly or the one calculated based on normative considerations.

Method and participants

Online questionnaire for direct rating and estimation using binary logistic regression

For the direct rating method and to obtain the probability judgments and verdicts to estimate the decision threshold using a binary logistic regression, an online questionnaire was used. The participants first answered some demographic questions and then read a scenario that was adapted from the 'loan' scenario used by Zamir and Ritov.²² According to the 'action' condition of the scenario, the plaintiff has allegedly given a long-time friend a loan of CHF 20,000 (equivalent to about USD 20,000 at the time) and requests that the court orders the defendant to pay him back the CHF 20,000. There is no written contract and no receipt; there is, however, some circumstantial evidence such as a deposit of CHF 20,000 into the defendant's bank account at the time he allegedly received the loan, and witnesses

22. Zamir and Ritov (2012: 200ff.) ('Experiment 1: Loan'). An English translation of the German version used in this study is reported in Appendix A.

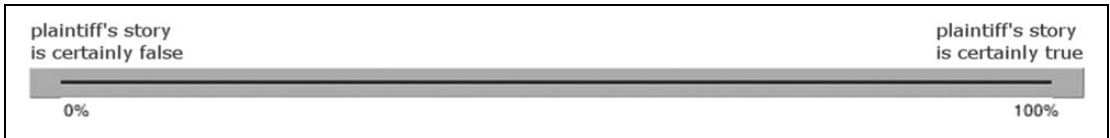


Figure 1. Slider used for rating degree of belief.

that report that there was talk about a loan. In the ‘negative declaratory action’ condition of the same scenario, the plaintiff is the alleged debtor and requests a declaration that he owes nothing to the alleged creditor.

After having read the scenario, the participants answer three control questions to check whether they have understood the scenario and, importantly, know who the plaintiff is (manipulation check). Participants who do not correctly answer the three control questions are excluded from further participation. Those who pass the control question then state their degree of belief in the truth of the plaintiff’s allegation using a sliding scale, where 0% indicates ‘the plaintiff’s story is certainly false’ and ‘100%’ means ‘the plaintiff’s story is certainly true’ (see Figure 1; a movable button appears after a participant clicks on the slider). Then the participants are asked whether they would grant or deny the claim. The order of these two questions was randomly varied to allow controlling for order effects.

Finally, the participants are given a choice of three verbal definitions of standards of proof used in civil proceedings in Switzerland, i.e. the default standard of ‘full conviction’, the intermediate standard of ‘high probability’ (used, for example, to establish causality) and the lower standard of ‘prima facie evidence’ (Glaubhaftmachung, used primarily in summary proceedings), and asked to identify the standard of ‘full conviction’.²³ Afterwards, they are asked to express the degree of belief required by the default standard of full conviction on an identical slider to the one used to rate the belief in the plaintiff’s story.

The link to the questionnaire was sent out by email in the early summer of 2012 by the High Courts of Berne and Zurich to all judges and judicial clerks (Gerichtsschreiber) in the respective Canton. Participation was voluntary; the recipients were encouraged to participate but were not compensated for their participation. Answering the questionnaire takes 10 to 15 minutes. Judicial clerks are lawyers who help draft the opinions of the court, do research and generally discuss the case with the judge. The tasks of judicial clerks in Switzerland are similar to those of judicial clerks in the US, but, unlike in the US, employment as a judicial clerk in Switzerland is not necessarily temporary. While most move on to other jobs after a couple of years, some stay for much longer (a fraction become judges; almost all judges are former judicial clerks).

A total of 186 members of courts (88 judges and 98 judicial clerks) participated in the survey. Twenty-six (14%) were excluded after incorrectly answering at least one of the control questions. Among those excluded are as many men as women, and their mean age does not differ significantly from that of the non-excluded participants. Of the 160 participants who completed the questionnaire, 76 are judges and 84 judicial clerks. Further descriptive statistics are reported in Table 2. A majority of judges (and a slight majority of judicial clerks) work primarily on civil cases, the others mostly on criminal cases and a small minority on enforcement and bankruptcy cases.

There are a total of approximately 343 judges and 476 judicial clerks in the Cantons Berne and Zurich.²⁴ Hence roughly 22% of the judges and 18% of the judicial clerks participated in the study.

23. The verbal definitions corresponded to those used by the Federal Supreme Court in its case law; for the standard of full conviction, see DFT 130 III 321 s. 3.2; for the intermediate standard see DFT 132 III 715 s. 3.1; for the ‘prima facie’ standard, see DFT 138 III 232 s. 4.1.1.

24. Data for Zurich from the annual report (Rechenschaftsbericht) of the High Court of Zurich for 2012, available from www.gerichte-zh.ch/organisation/obergericht/rechenschaftsbericht.html (accessed 1 April 2016). Data for Berne from the human resource department of the High Court, Mrs Sonja Hartmann (on file with author).

Table 2. Descriptive statistics for the court member sample.

	Men	Mean age (SD)	Mean work experience in years (SD)	Primarily working in civil law
Judges (N = 76)	47 (62%)	48.7 (8.3)	18.7 (8.9)	45 (59%)
Judicial clerks (N = 84)	40 (48%)	32.7 (7.3)	4.8 (5.9)	44 (52%)
Total	87 (54%)	40.2 (11.2)	11.6 (10.8)	89 (56%)

Sixty-six per cent of all judges are male, while only 40% of all judicial clerks are male.²⁵ As can be seen from the data reported in Table 3, these proportions are roughly reflected in the sample, which indicates that the sample is quite representative. Given that one usually is elected as a judge between the age of 30 and 35 and retires at 65,²⁶ the mean age of the judges in the sample of 48.7 years seems quite representative of the population, too.

To allow a comparison of the court members with members of the general population, the same questionnaire was administered to a sample of the general German population between 18 and 60 years of age in February 2012. The participants were recruited by a commercial panel provider and remunerated for their participation with credit points redeemable for goods. The questionnaire deviated in two points from the one used for the court members. First, the question on how to decide the case was re-phrased by adding the text in italics: ‘How would you decide this case *if you were acting as a judge?*’ Secondly, the final questions regarding the definitions of the standard of proof and the expression of the normatively required decision threshold as a degree of belief were omitted.

A total of 247 participants (mean age 41 years, SD = 12.6; 49.7% men) completed the questionnaire. According to self-reports, 141 (53%) of the participants were employed, 11 (4%) employed in a managerial position, 19 (8%) self-employed, 26 (11%) students and 60 (24%) ‘other’.

Paper and pencil questionnaire for decision theory-based method

For the estimation of the error costs associated with each outcome of a civil action, a thought experiment inspired by Laudan and Saunders was implemented using a simple one-page paper and pencil questionnaire (Laudan and Saunders, 2009: 23). The participants were attending a seminar on case management in civil proceedings organised by the High Court of Berne on 23 August 2012. All 49 attendees completed the simple questionnaire, which consisted of the scenario described in the next paragraph and some demographic questions on the back of a single sheet of paper. Completion of the questionnaire took less than 10 minutes; during this time, the participants were not observed to be talking to each other.

The participants were told that an anonymous donor had given them CHF 100,000 under the condition that they must spend the entire amount, but not more than the entire amount, on preventing the possible outcomes of the following simple case: An individual brought a civil action against another individual before a competent court, requesting the payment of CHF 100,000. Both parties have similar wealth and income. *The only issue is an issue of fact, namely, whether the plaintiff actually gave the defendant a loan in the amount of CHF 100,000.* It is undisputed that the loan, if it had been disbursed, had not been paid back despite being overdue.

The four possible outcomes of the case are (italics in the original. The order of the outcomes was randomised):

25. These figures are based on categorisation by first name, which is not absolutely accurate, as some first names do not allow determination of the gender. However, in German, such names are very rare.

26. In Switzerland, judges at lower courts are elected by popular vote, at higher courts they are usually appointed by parliament.

- a) the plaintiff *has* given the defendant a loan of CHF 100,000, and the defendant *is* ordered to pay CHF 100,000 to the plaintiff;
- b) the plaintiff *has not* given the defendant a loan of CHF 100,000, but the defendant *is* ordered to pay CHF 100,000 to the plaintiff;
- c) the plaintiff *has not* given the defendant a loan of CHF 100,000, and the defendant *is not* ordered to pay CHF 100,000 to the plaintiff;
- d) the plaintiff *has* given the defendant a loan of CHF 100,000, but the defendant *is not* ordered to pay CHF 100,000 to the plaintiff.

Because the total amount to be invested in preventing the outcomes is fixed, the scenario avoids the difficult question of estimating the absolute error costs of each outcome and instead allows identifying the (only relevant) *ratio* of the error costs.

Two participants are excluded from the following analysis because they invested more than the donated amount. Of the 47 participants who correctly completed the questionnaire, 20 (43%) were judges, 10 were persons in charge of conducting the mandatory conciliation hearing and 17 (36%) were judicial clerks. The persons in charge of conducting the mandatory conciliation are fully qualified lawyers. Under Swiss civil procedural law, in most cases, a party can only bring a suit after an attempt at settlement has failed, which must be conducted before the competent authority (see art. 202 ff. of the Swiss Civil Procedure Act). At 57%, women were in the majority overall, but they were, unsurprisingly, a minority of 35% among the judges in the sample. Participants were between 29 and 61 years old, with a mean age of 40, the mean age of judges (47) being higher than that of the judicial clerks (32) and the persons in charge of reconciliation (38).²⁷ The participants had an average of 7.3 years of work experience; judges a mean of 12 years and judicial clerks a mean of 4 years. Thirteen of the participants had also participated in the online questionnaire (about two months earlier). For the reasons outlined above, the demographics of the sample are quite representative for the population.

Results

Results from the direct rating / self-report method

One hundred and eighteen (73.8%) court members identified the correct verbal definition of the default standard of proof of 'full conviction'. Forty (25%) chose the verbal definition of the intermediate standard of 'high probability', and only 2 the definition of 'prima facie evidence'. No significant differences in the proportion of correct answers by judges versus judicial clerks, men versus women and those who self-identified as working primarily in civil law versus those who work in other areas of law were found.

Nineteen participants indicated that a 100% certainty was required under the standard of full conviction; the lowest threshold indicated was 51%. The median decision threshold under the standard of full conviction was 91% (M = 88.8, SD = 11.9). No difference in the mean decision threshold between those who correctly identified the verbal definition of the standard of full conviction (N = 118, M = 87.5, SD = 12.9) and those who incorrectly identified the verbal definition (N = 42, M = 90.4, SD = 8.3) was observed. No difference in the required decision threshold between those who granted the request (N = 69, M = 87.4, SD = 11.3) and those who denied the request (N = 91, M = 89.9, SD = 12.4) was found either.

27. Nine participants did not state their age.

Table 3. Mean apportionment of CHF 100,000 to prevent outcomes (in brackets median).

	loan was given	loan was not given
request is granted	CHF 2,979 (CHF 0)	CHF 47,766 (CHF 50'000)
request is denied	CHF 46,276 (CHF 50'000)	CHF 2,979 (CHF 0)

Results from the decision theory-based method

Table 3 reports the descriptive statistics for the apportionment of the CHF 100,000 among the four possible outcomes of the civil action.

The modal answer, given by 34 (72%) of the participants, was to invest CHF 50,000 each in the prevention of a false positive (upper right cell in Table 3) and a false negative (lower left cell in Table 3) and nothing in the prevention of the correct decisions. The second most common answer, given by three participants, was to invest CHF 25,000 in the prevention of each outcome.

Taking the means from Table 3 and calculating the decision threshold using equation (1) results in a decision threshold of 51%. Calculating the decision threshold in this way depends on the strong assumption that each additional Swiss Franc invested in the prevention of an outcome has the same effect as the last. However, given that almost the same amounts were invested in the prevention of each of the correct and each of the incorrect decisions, the assumption hardly matters.

Ten (21%) participants have invested different amounts in the prevention of the two types of errors. Seven of those invested more in the prevention of a false positive (grant of request despite no loan given) than in the prevention of a false negative outcome. The decision thresholds resulting from equation (1) using the implied error costs indicated by these seven participants result with a single exception in thresholds below 63%. The implied error costs of just one participant result in a decision threshold of 87.5%, close to the mean threshold obtained with the direct rating method.

Results from the binary logistic regression

Excluded from the following analyses were the 24 participants (four court members and 20 members of the general population) that granted the request, although they indicated that their degree of belief in the truth of the allegations supporting the claim was less than 50% or denied the request although they indicated that they were 100% certain that the allegations supporting the claim were true.²⁸

Table 4 reports the descriptive statistics from the online questionnaire. The first four columns report the results for the 'action' condition, the last four columns those for the 'negative declaratory action' condition (since the order of questions did not have an effect, as will be shown below, the results are pooled across the 'order' condition).

Unsurprisingly, the mean conviction of those who granted the request is higher than the mean conviction of those who denied the request for both conditions and both samples. A higher percentage of court members grants the claim than members of the general population (71% versus 58%; χ^2 [3 df, N = 383] = 21.323, $p < 0.001$). Only three members of court granted the negative declaratory action.

Looking at the mean conviction irrespective of whether the request was granted or denied, the allegations of the plaintiff in the 'action' condition (the alleged creditor) were believed with a degree of belief of 76.5% (members of court) and 66.9% (general population). The allegations of the plaintiff in the negative declaratory action condition (the alleged debtor) were believed to a degree of 27.8% (members of court) and 41.1% (general population). Since the two propositions are incompatible and

28. The results including those 24 subjects are reported in the Appendix B. All effects remain significant with the exception of type of request in Model 3. Model fit as measured by pseudo R^2 decreases with the additional 24 subjects.

Table 4. Descriptive statistics for the degree of belief and decision.

	Action							
	grant		denial		negative declaratory action			
	obs. (share)	Ø conviction (SD)	obs. (share)	Ø conviction (SD)	grant	denial		
				obs. (share)	Ø conviction (SD)	obs. (share)	Ø conviction (SD)	
members of court (N = 156)	62 (71%)	84.1 (10.2)	25 (29%)	57.7 (19.0)	3 (4%)	81.3 (11.1)	66 (96%)	25.4 (23.0)
general population (N = 227)	93 (58%)	79.9 (13.9)	68 (42%)	49.2 (23.5)	13 (20%)	63.5 (16.3)	53 (80%)	36.8 (26.2)

Table 5. Binary logistic regressions.

	Model 1	Model 2
[B]elief	0.081*** (0.009)	0.085*** (0.009)
[C]ourt	-0.291 (0.320)	-0.358 (0.330)
[O]rder	0.044 (0.300)	0.045 (0.308)
[T]ype of request	1.290*** (0.386)	1.295*** (0.386)
[M]ale		0.368 (0.315)
[A]ge		-.040** (0.013)
Intercept	-6.295*** (0.699)	-5.039*** (0.798)
Observations	383	383
Pseudo R ²	0.470	0.490

****p* < 0.001; ***p* < 0.01; **p* < 0.05.

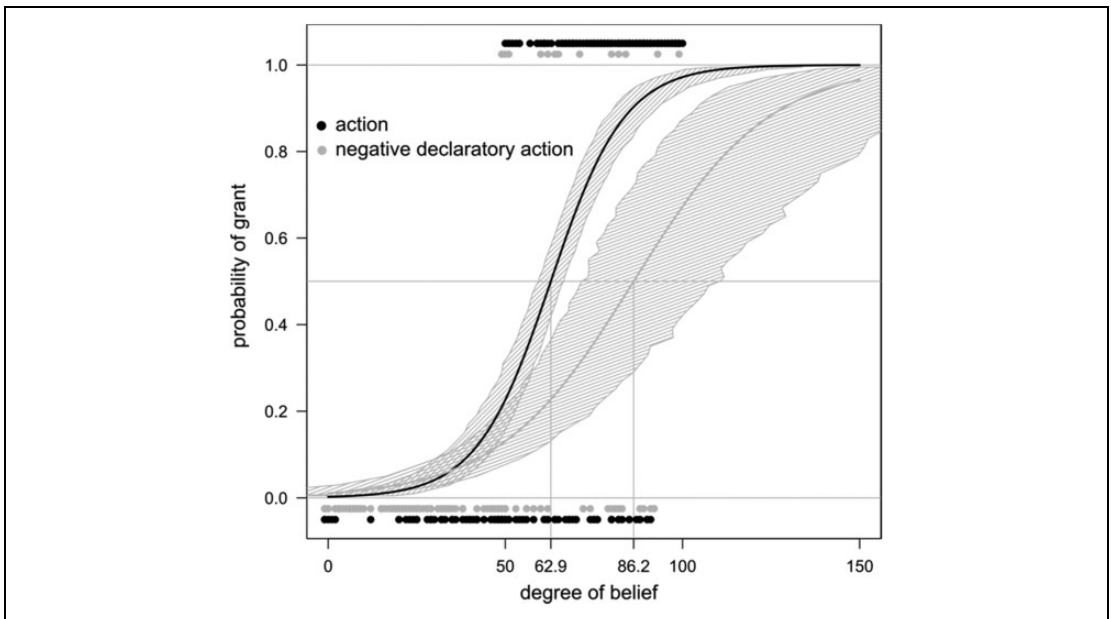


Figure 2. Degree of belief required for granting the claim.

one must be true, the total in each condition should add up to 100%. Both for the general population (108%) and the members of court (104.3%), the total is slightly above 100%.

Next, the coefficient weights for a binary logistic regression with grant of request as the dependent variable and conviction (degree of belief on a scale from 0 to 100), membership of court (dummy variable, 1 = judge or judicial clerk), order of question (dummy variable, 1 = dichotomous decision before degree of belief) and type of action (dummy variable, 1 = ‘action’ condition) were estimated (Model 1). The results are reported in Table 5. Unsurprisingly, the probability of grant (more precisely, the natural logarithm of the odds ratio $\Pr(\text{grant}_i)/\Pr(1-\text{grant}_i)$) increases with an increase in the conviction that the allegations supporting the claim are true. The order of questions and membership of court do not exert a significant influence, but the type of request does, with the negative declaratory action being granted at a much lower rate.

Continuing with the discussion of Model 1, the relationship between degree of belief and probability of grant is displayed graphically. Figure 2 plots the function of degree of belief on probability of grant for

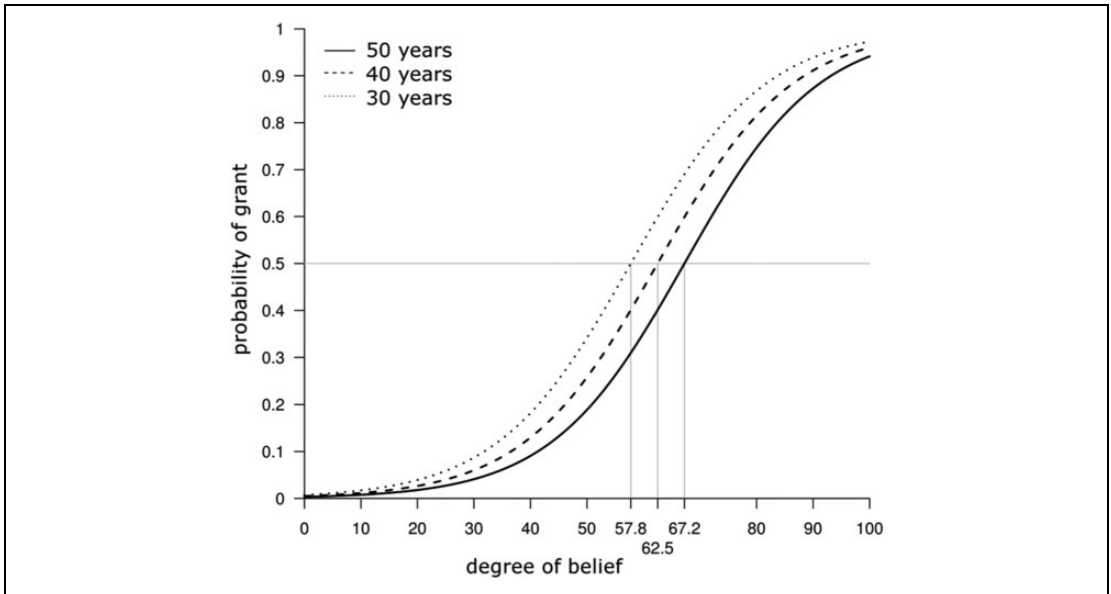


Figure 3. Degree of belief required for granting the claim at three different age levels.

both types of claims. The grey shaded areas indicate the 95%-confidence intervals (bias corrected and accelerated), estimated by bootstrapping.²⁹ Since few participants granted the negative declaratory action, the confidence intervals for the negative declaratory action are much wider than for the action.

At a conviction of 62.9% (95%-confidence interval 59.4–66.5) the probability that the request in the action condition (order for payment) is granted is 50%. For the request in the negative declaratory action condition (declaration that nothing is due) to be granted with a probability of 50%, the degree of belief must be 86.2% (95%-confidence interval 71.6–100).

In a next step, Model 1 is expanded by including the age and gender of the participant as independent variables (Model 2). Gender has no influence on the probability of grant. Age has a negative influence, i.e. the older the participant, the less likely he or she is to grant the request. Figure 3 shows this relationship graphically for the three age levels of 30, 40 and 50 years. As can be seen, a 50-year-old (bold line) requires a degree of conviction of 67.2 to grant the claim with a probability of 50%, while a 30-year-old only requires 57.8.

Discussion

The results reveal stark differences in the estimated decision threshold in civil cases, depending on the measurement method. The direct rating method reveals a decision threshold of judges and judicial clerks of about 90%, which aligns with the quantification suggested in the German and Swiss legal literature and with results from an empirical study by Berger-Steiner with a sample of Swiss judges (2008: 221). The results from the direct rating method also show that there is a small, but non-negligible, minority of judges that believe any decision granting a request is only permissible if there is a 100% certainty that the factual allegations supporting the claim are true.

29. 2,000 bootstrap replicates were generated using the R (R Core Team, 2013) command 'boot ()' (Canty and Ripley (2013) implementing Davison and Hinkley (1997)). The 95% confidence interval was calculated as adjusted bootstrap percentile (Bca) as suggested by Wright et al. (2011: 96ff.)

A decision threshold of 90% implies that a false positive error (erroneously granting the request) has a disutility that is about 10 times that of a false negative decision (erroneously denying the request). However, when the error costs associated with each decision are elicited directly, judges and judicial clerks do not express error costs that would support the high default standard of proof of ‘full conviction’ applicable in Continental European civil law jurisdictions such as Germany and Switzerland. They rather share the assumption of symmetrical error costs underlying the ‘preponderance of the evidence’ or ‘balance of probabilities’ standard of proof of the common law. This makes it doubtful whether the civil law’s higher standard of proof is defensible in a decision theoretic framework.

Finally, the degree of belief in the truth of the allegations supporting the request in the action condition at which the probability of grant is 50% is only 63%, while it is 86% for the request in the declaratory action condition. Zamir and Ritov have estimated the decision threshold of Israeli trial lawyers using the rank-order method at 70, using two different scenarios, and at 60 for Israeli law students using the same ‘loan’ scenario that was used here (Zamir and Ritov, 2012: 176ff.). While caution must be applied when comparing these results because of the different answer formats—Zamir and Ritov asked for a number between 0 and 100, rather than a position of a slider on a scale from 0% to 100%—they do seem to indicate that the decision threshold of the Israeli lawyers is arguably too high, since Israel applies the common law’s ‘preponderance of the evidence’ standard in civil disputes, while that of the Swiss judges and judicial clerks should be higher if it were to conform to the normatively required ‘full conviction’. The results suggest that there may be a ‘natural’ or ‘intuitive’ decision threshold that is largely unaffected by the normative standard of the respective legal system. They also lend empirical support to the claim that *in fact*, rather than in theory, the standard of proof used by courts in Continental Europe in civil cases is not much different from the common law’s standard. The heated debate between Clermont/Sherwin and Taruffo may be much ado about nothing.

No fully convincing explanation exists for the higher decision threshold observed in the negative declaratory action condition. It is certainly true that a negative declaratory action is a rare beast. As a general rule, it is up to the alleged creditor to decide when and whether to bring suit. Swiss law therefore requires that the plaintiff in a negative declaratory action shows a particular legal interest (‘Rechtsschutzinteresse’) in an immediate decision.³⁰ The scenario was silent whether the conditions for a negative declaratory action were met. This may explain why many judges were hesitant to grant the request. On the other hand, the sample of the general population that shows the same hesitation was most likely not aware of the legal requirements for a negative declaratory action.

The influence of age on the degree of belief required for a grant of the request was not expected. It is known that the probability of convicting an accused in a criminal trial increases with the age of the fact finder (Anwar et al., 2012). The effect of age in the civil case goes in the opposite direction, i.e. the higher the age, the lower the probability of grant. Neither for the effect of age on conviction in criminal cases nor for the effect on the grant in civil cases are there convincing causal explanations.

A limitation of the present study is that it only used one scenario. It may be that different scenarios lead to different decision thresholds.³¹ Based on the current data, this cannot be ruled out.

Conclusion

Judges from a civil law country (Switzerland) seem to use a much lower decision threshold in civil cases than required by legal doctrine and case law and indicated by themselves when asked directly. The actual difference in standards of proof in civil cases between common law and civil law may therefore be much smaller than thought. The error costs that the Swiss judges and judicial clerks associate with each

30. DFT 131 III 319 s. 3.5; 120 II 20 s. 3a.

31. The results of Zamir and Ritov (2012: 177), who used three different scenarios, indicate some influence of type of scenario on the decision threshold; however, none that would be big enough to question the overall narrative.

possible outcome of a civil action imply an error-cost-minimising decision threshold of just 51%, the same as the threshold usually stated for the common law's 'preponderance of the evidence' or 'balance of probabilities' standard of proof.

Appendix A

Imagine that you are a judge that has to decide the following case

Mr Arnold, the plaintiff, has sued Mr Graf, the defendant, for payment of CHF [EUR] 20,000. [Note: in the 'negative declaratory action' condition, the preceding sentence was replaced by: 'Mr Graf, the plaintiff, wants a declaration that he does not owe Mr Arnold, the defendant, any money.']

Mr Arnold claims that he lent Mr Graf CHF [EUR] 20,000 and Mr Graf had not paid him back. Mr Graf denies ever having received a loan from Mr Arnold.

It is undisputed that Mr Arnold and Mr Graf have known each other for years, since they attended college together, and have a long-running business relationship. Mr Arnold is a business lawyer, Mr Graf a businessman running his own business.

According to Mr Arnold, Mr Graf asked him for a loan of CHF [EUR] 30,000 for a couple of months to overcome financial difficulties. Mr Arnold agreed to lend Mr Graf CHF [EUR] 20,000. Because the two knew each other for so long, he did not insist on a written contract. According to Mr Arnold's testimony, he gave the money in cash to Mr Graf, without getting a receipt, on 1 December 2010 (roughly one year before filing suit).

A bank statement of Mr Graf's account shows a deposit of CHF [EUR] 20,000 on 1 December 2010.

Mr Graf testifies that he had indeed asked Mr Arnold for a loan, and Mr Arnold initially agreed to lend him CHF [EUR] 20,000, but then changed his mind and did not give him any money. He, i.e. Mr Graf, had gotten a loan from another party to overcome his financial difficulties. He would not identify the third party.

A witness called by Mr Arnold testifies that Mr Graf told her on 2 December 2010 that he had received a loan from Mr Arnold in the amount of CHF [EUR] 20,000 on 2 December 2010. Mr Graf remarks that he cannot remember having had a conversation with the witness.

The wife of Mr Arnold testifies that her husband had told her at the end of November 2010 that he intended to loan CHF [EUR] 20,000 to Mr Graf, for tax reasons in cash, without a receipt.

Mr Graf says that it is inconceivable that Mr Arnold, a business lawyer, would give him such a large amount of money without a written contract and without demanding a receipt.

Note: The scenario was in German. In Germany, the amount was given in Euro, in Switzerland, in Swiss Franc. One Swiss Franc at the time of the survey was approximately EUR 0.9. Respondents could access the full text of the scenario while answering the questions using a help button.

Appendix B

Table B1. Descriptive statistics for the degree of belief and decision without exclusion of 24 subjects.

	Action				negative declaratory action			
	grant		denial		grant		denial	
	obs. (share)	Ø convic- tion (SD)	obs. (share)	Ø convic- tion (SD)	obs. (share)	Ø convic- tion (SD)	obs. (share)	Ø convic- tion (SD)
members of court (N = 160)	63 (72%)	83.0 (13.5)	25 (29%)	57.7 (19.0)	6 (8%)	65.2 (19.0)	66 (92%)	22.4 (23.0)
general population (N = 247)	99 (59%)	77.8 (15.7)	70 (41%)	50.6 (24.7)	24 (30%)	51.6 (20.8)	54 (70%)	37.9 (27.3)

Table B2. Binary logistic regressions without exclusion of 24 subjects.

	Model 1	Model 2
[B]elief	0.055*** (0.006)	0.057*** (0.006)
[C]ourt	-0.253 (0.269)	-0.299 (0.257)
[O]rder	-0.008 (0.250)	0.003 (0.259)
[T]ype of request	0.652* (0.302)	0.651* (0.306)
[M]ale		0.312 (0.263)
[A]ge		-0.035** (0.011)
Intercept	-3.772*** (0.425)	-2.612*** (0.561)
Observations	407	407
Pseudo R ²	0.323	0.344

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by a personal grant to the author by the Swiss National Science Foundation (SNF), grant no. PA00P1_131519.

References

- Anderson T, Schum DA and Twining WL (2006) *Analysis of Evidence*. Cambridge: Cambridge University Press.
- Anwar S, Bayer P and Hjalmarsson R (2012) A fair and impartial jury? The role of age in jury selection and trial outcomes. *NBER Working Paper* (17887). Available at: <http://www.nber.org/papers/w17887.pdf> (accessed 31 January 2015).
- Ball V (1960) The moment of truth: Probability theory and standards of proof. *Vanderbilt Law Review* 14: 807–830.
- Bender R (1981) Das Beweismaß. In: Grunsky W, Stürmer R, Walter G, et al. (eds) *Festschrift für Fritz Baur*. Tübingen: Mohr Siebeck, 247–271.
- Berger-Steiner I (2008) *Das Beweismass im Privatrecht: Eine dogmatische Untersuchung mit Erkenntniswert für die Praxis und die Rechtsfigur der Wahrscheinlichkeitshaftung*. Bern: Stämpfli.
- Brinkmann M (2005) *Das Beweismaß im Zivilprozess aus rechtsvergleichender Sicht*. Köln, Berlin, München: Heymann.
- Bühler A (2010) Beweismass und Beweiswürdigung bei Gerichtsgutachten—unter Berücksichtigung der jüngsten Lehre und Rechtsprechung. *Jusletter*.
- Canty A and Ripley B (2013) *boot: Bootstrap R (S-Plus) Functions: R package version 1.3-9*. Available at: <http://cran.r-project.org/web/packages/boot/> (accessed 11 April 2016).
- Clermont KM (2009) Standards of proof revisited. *Vanderbilt Law Review* 33: 469–488.
- Clermont KM and Sherwin E (2002) A comparative view of standards of proof. *American Journal of Comparative Law* 50(2): 243–276.
- Cullison AD (1969) Probability analysis of judicial fact-finding: A preliminary outline of the subjective approach. *University of Toledo Law Review* 1(3): 538–598.

- Dane F (1985) In search of reasonable doubt. *Law and Human Behavior* 9(2): 141–158.
- Davison AC and Hinkley DV (1997) *Bootstrap Methods and Their Applications*. New York: Cambridge University Press.
- Deppenkemper G (2004) *Beweiswürdigung als Mittel prozessualer Wahrheitserkenntnis: Eine dogmengeschichtliche Studie zu Freiheit, Grenzen und revisionsgerichtlicher Kontrolle tatrichterlicher Überzeugungsbildung* (§ 261 StPO, § 286 ZPO). Göttingen: V&R Unipress.
- Dhami MK (2008) On measuring quantitative interpretations of reasonable doubt. *Journal of Experimental Psychology: Applied* 14(4): 353–363.
- Einmahl M (2001) Zeugenirrtum und Beweismaß im Zivilprozess: Eine Fallstudie am Beispiel des Verkehrsunfallprozesses. *Neue Juristische Wochenschrift*: 469–475.
- Engel C (2009) Preponderance of the evidence versus intine conviction: A behavioral perspective on a conflict between American and Continental European law. *Vermont Law Review* 33(3): 435–467.
- Fuchs B (2005) *Das Beweismaß im Arzthaftungsprozess*. Frankfurt am Main: Peter Lang.
- Gottwald P (2000) Das flexible Beweismaß im englischen und deutschen Zivilprozess. In: Gottwald P, Jayme E and Schwab D (eds) *Festschrift für Dieter Henrich zum 70. Geburtstag, 1. Dezember 2000*. Bielefeld: Giesecking, 165–176.
- Greger R (1978) *Beweis und Wahrscheinlichkeit: Das Beweiskriterium im Allgemeinen und bei den sogenannten Beweiserleichterungen*. Köln: Heymann.
- Hastie R (ed.) (1993) Algebraic models of juror decision processes. In: *Inside the Juror: The Psychology of Juror Decision Making*. Cambridge: Cambridge University Press, 84–115.
- Kadner Graziano T (2011) ‘Alles oder nichts’ oder anteilige Haftung bei Verursachungszweifeln?—Zur Haftung für perte d’*un*ce chance/loss of chance und eine Alternative. *Zeitschrift für Europäisches Recht* 19(1): 171–200.
- Kaplan J (1968) Decision theory and the factfinding process. *Stanford Law Review* 20(6): 1065–1092.
- Kaye DH (1987) Apples and oranges: Confidence coefficients and the burden of persuasion. *Cornell Law Review* 73(1): 4–77.
- Kegel G (1967) Der Individualanscheinsbeweis und die Verteilung der Beweislast nach überwiegender Wahrscheinlichkeit. In: Biedenkopf KH, Coing H and Mestmäcker E (eds) *Das Unternehmen in der Rechtsordnung: Festgabe für Heinrich Kronstein aus Anlass seines 70. Geburtstages am 12. September 1967*. Karlsruhe: C.F. Müller, 321–344.
- Kinsch P (2009) Probabilité et certitude dans la preuve en justice. In: *Actes de la Section des Sciences Morales et Politiques*. XII edn. Luxemburg: Institut Grand-Ducal, 65–103.
- Laudan L and Saunders HD (2009) Re-thinking the criminal standard of proof: Seeking consensus about the utilities of trial outcomes. *International Commentary on Evidence* 7(2): 1–36.
- Lee TR (1997) Pleading and proof: The economics of legal burdens. *1997 Brigham Young University Law Review* 1(1): 1–34.
- Maassen BM (1975) *Beweismaßprobleme im Schadenersatzprozess*. Köln: Heymann.
- McBride E (2009) Is the civil ‘higher standard of proof’ a coherent concept? *Law, Probability & Risk* 8(4): 323–351.
- Motsch R (1978) Vom Prozess als Beweis zum Überwiegensprinzip. In: Klug U, Ramm T, Rittner F, et al. (eds) *Gesetzgebungstheorie, juristische Logik, Zivil- und Prozessrecht: Gedächtnisschrift für Jürgen Rödig*. Berlin, New York: Springer, 334–348.
- Motsch R (2009) Comparative and analytical remarks on judicial fact-finding. In: Rübmann H (ed.) *Festschrift für Gerhard Käfer*. Saarbrücken: Juris GmbH, 241–271.
- Nell EL (1983) *Wahrscheinlichkeitsurteile in juristischen Entscheidungen*. Berlin: Duncker & Humblot.

- O'Malley KF, Grenig JE and Lee WC (2001) *Federal Jury Practice and Instructions*. Eagan, MN: Westlaw.
- Posner RA (1999) An economic approach to the law of evidence. *Stanford Law Review* 51(6): 1477–1546.
- Prütting H (2010) Beweislast und Beweismaß: Der Einfluss Leo Rosenbergs und Karl Heinz Schwabs auf die Entwicklung eines modernen Beweisrechts. *Zeitschrift für Zivilprozess* 123(2): 135–145.
- R Core Team (2013) R: A language and environment for statistical computing. Available at: <http://www.R-project.org> (accessed 11 April 2016).
- Rechberger WH (1990) Maß für Maß im Zivilprozess?: Ein Beitrag zur Beweismaßdiskussion. In: Prütting H (ed.) *Festschrift für Gottfried Baumgärtel zum 70. Geburtstag*. Köln: Heymann, 471–490.
- Redmayne M (1999) Standards of proof in civil litigation. *Modern Law Review* 62(2): 167–195.
- Sand LB, Rakoff JS, Reiss SA, et al. (2007) *Modern Federal Jury Instructions—Civil Volumes*. Los Angeles: LexisNexis.
- Scherer I (1996) *Das Beweismaß bei der Glaubhaftmachung*: Cologne, Berlin: Heymann.
- Sheppard S (2003) The metamorphoses of reasonable doubt: How changes in the burden of proof have weakened the presumption of innocence. *Notre Dame Law Review* 78(4): 1165–1250.
- Simon RJ (1970) 'Beyond a reasonable doubt': An experimental attempt at quantification. *Journal of Applied Behavioral Science* 6(2): 203–209.
- Simon RJ and Mahan L (1971) Quantifying burdens of proof: A view from the bench, the jury, and the classroom. *Law & Society Review* 5(3): 319–330.
- Stein A (2005) *Foundations of Evidence Law*. Oxford, New York: Oxford University Press.
- Summermatter D and Jacober C (2012) Zum Beweismass beim Kausal- und Motivationszusammenhang: Versuch einer begrifflichen Klärung anhand der jüngeren Rechtsprechung im Zivil-, Straf- und Sozialversicherungsrecht. *HAVE* (2): 136–149.
- Taruffo M (2003) Rethinking the Standards of Proof. *American Journal of Comparative Law* 51(3): 659–677.
- Tillers P and Gottfried J (2007) Case comment—United States v. Copeland, 369 F. Supp. 2d 275 (E.D.N.Y. 2005): A collateral attack on the legal maxim that proof beyond a reasonable doubt is unquantifiable? *Law, Probability and Risk* 5(2): 135–157.
- Tyree AL (1982) Proof and probability in the Anglo-American legal system. *Jurimetrics* 23(1): 89–100.
- Wagner G (2009) Ökonomische Analyse der Rechtsmittel. In: Bork R, Eger T and Schäfer H (eds) *Ökonomische Analyse des Verfahrensrechts: Beiträge zum XI. Travemünder Symposium zur Ökonomischen Analyse des Rechts* (26–29 March 2008). Tübingen: Mohr Siebeck, 156–193.
- Walter G (1979) *Freie Beweiswürdigung*. Tübingen: Mohr Siebeck.
- Walter HP (2009) *Beweis und Beweislast im Haftpflichtprozess*. Vienna: Haftpflichtprozess, 47–68.
- Wright DB, Strubler KA and Vallano JP (2011) Statistical techniques for juror and jury research. *Legal and Criminological Psychology* 16(1): 90–125.
- Wright RW (2009) Proving facts: Belief versus probability. In: Koziol H and Steininger BC (eds) *Tort and Insurance Law*. Vienna: Springer, 79–105.
- Zamir E and Ritov I (2012) Loss aversion, omission bias, and the burden of proof in civil litigation. *Journal of Legal Studies* 41(1): 165–207.