View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

Preprints of the
Max Planck Institute for
Research on Collective Goods
Bonn 2008/33

Preponderance of the Evidence versus *Intime Conviction*

A Behavioural Perspective on a Conflict between American and Continental European Law

Christoph Engel

MAX PLANCK SOCIETY

Preprints of the
Max Planck Institute
for Research on Collective Goods

Bonn 2008/33

Preponderance of the Evidence versus *Intime Conviction*

A Behavioural Perspective on a Conflict between American and Continental European Law

Christoph Engel

August 2008

Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn
http://www.coll.mpg.de

Preponderance of the Evidence versus *Intime Conviction*

A Behavioural Perspective on a Conflict between American and Continental European Law*

Christoph Engel

I. The Issue 2
II. Doctrine 3
III. Norms 6
  1. Acceptable Probability of Error 6
  2. Utilitarian Norm 8
IV. A Behaviourally Informed Perspective 9
  1. Violations of the Rational Choice Benchmark 9
  2. A Proper Definition of the Task 10
  3. Sense Making 10
  4. Consistency Maximisation 12
  5. Rescued by Emotions 14
V. Conclusion 19

* Helpful comments by Stephan Dickert and Martin Hellwig are gratefully acknowledged.

I. The Issue

European law is irrational. American law is irresponsible. These beliefs are the essence of one of the few true conflicts between (US) common law and (European) continental law. At the surface, the conflict is confined to an apparently technical issue in the law of evidence. On the European continent, for the court to hold against the defendant, the judge must be convinced that the facts brought forward by the plaintiff in support of the claim are indeed true. In principle, continental law does not make a difference between civil law and criminal law. The standard of proof is *intime conviction* throughout (*French* Code de Procédure Pénale, Art. 353[1]; *German* Zivilprozessordnung, § 286 I 1; *German* Strafprozessordnung, § 261). By contrast, US law has three different standards of proof (*Addington v. Texas* 441 U.S. 418, 422 (1979)). In criminal law, the charge must be established "beyond a reasonable doubt". In civil law, normally the plaintiff wins if only "the preponderance of the evidence" is in her favour. Only in a limited number of civil law matters, of particular gravity for the defendant, the intermediate standard of "clear and convincing evidence" must be met.

At closer sight the conflict roots deeper. US law not only differentiates; it conceptualises proof differently. On the continent, proof is understood as the strictly subjective impression in the judge's mind. By contrast, in US law, proof is an objective concept. The conflict in the law of evidence is closely related to a fundamental divide in epistemology. In line with the prevalent position in the sciences, the American law of evidence aims at objectivity. The continental law of evidence is closer to the opposite view that, in epistemology and in the history of sciences, is characterised as a quest for truth (Daston and Galison 2007). Partisans of the latter view stress that even science, if understood as practice, has an art component. Ultimately, strict intersubjectivity is prohibitively costly, the position believes.

As usual, at closer sight, the conflict is more nuanced. There are partisans of probabilistic approaches on the continent (a prominent voice in German law is Kegel 1967). And, in particular in criminal law, the strict objectivist position is hotly debated in the US (e.g. Franklin 2006). Yet the underlying difference is net. It is both normative and positive. Which is the goal of judicial procedure? And by which means is this goal best attained? All legal orders agree that courts have to decide under conditions of uncertainty (Brook 1982). It is less clear, however, whether the appropriate way of dealing with this uncertainty is some form of (utilitarian) calculus (as advocated by, e.g. Tillers and Gottfried 2006). Alternatively, and more modestly, judicial procedure might only strive at making error sufficiently unlikely (as proposed by, e.g. Newman 1993).

The deepest divide, however, is psychological. Those believing in objectivity point to the fact that individual judgement is error prone. It is the purpose of procedural law to impose as much rationality as possible as a check (Koehler 2006). Those believing in truth counter that rational calculus is not what empirical judges and jurors do. A norm for decision quality is pointless if actual judges and jury members have no chance to live up to it, they believe. Empirically, from

1 The French Code de Procédure Civile 1976 does not contain an explicit provision on intime conviction, Taruffo in American Journal of Comparative Law (2003) 665.

the evidence presented to them, judges and jury members construct stories. They decide for the plaintiff, or for the prosecution for that matter, if it has been able to induce a coherent story, with no striking gaps, that resonates with the world knowledge of the juror (Hastie 1993). While having deliberate components, the bulk of this mental activity is performed by the automatic system. Only the result is propelled back to consciousness as an intuition (Simon 2004). Specifically, consciousness is not only informed that the automatic system has come to a resolution. This information is combined with a level of confidence (Simon 2004).

Actually, at this point emotions come into play. It is quite plausible that the level of confidence is itself tagged by a somatic marker. While this hypothesis still awaits the empirical test, a related effect is well established. If an individual is aware of a serious risk, this prospect is emotionally highlighted by a somatic marker. It induces the individual to steer clear from the risk, even if the ensuing benefit seems appealing (Damasio 1994). Judges and jurors decide on other people's lives. In so doing, they exercise sovereign powers. In this perspective, asking them for their *in-time conviction* is a way of making personal accountability salient (Tetlock 1983). The jury instruction is a technology to trigger the somatic marker, and thereby to influence how the automatic system assesses the evidence.

The automatic system is very powerful. It handles vast amounts of information in almost no time. It elegantly combines online information with the traces of previous experiences stored in memory. Most importantly: it enables human decision-makers to make meaningful decisions on an incomplete factual basis (Engel and Singer 2008). The automatic system is, however, not foolproof. It can be betrayed by spurious analogies. It is influenced by stereotypes and predilections. And it can be manipulated by outsiders (Simon 2004). At the heart of the conflict between US and continental European law is therefore the different assessment of cost and benefit. While continental law is attracted by the benefit, US law is scared by the cost. At the end of the day, this is a normative issue. Once one becomes aware of what the conflict is really about, one however is in a better position to decide. And hopefully clever institutional arrangements might be able to overcome the apparent yes or no decision.

The remainder of this paper is organised as follows: section II takes a closer look at doctrine on both sides of the Atlantic. Section III discusses the norm. Section IV establishes the psychological underpinnings. Section V draws normative conclusions.

II. Doctrine

Comparative law tends to be a disinterested affair. Not so in this case. Two American observers wondered “how civilians can be so wrong” (Clermont and Sherwin 2002:243). To them, the civil law of evidence “is downright bizarre” (Clermont and Sherwin 2002:254). They put it down to the “disruption in legal procedure” resulting from the French Revolution (Clermont and Sherwin 2002:258), and to a lack of interest in later years (Clermont and Sherwin 2002:257), which led to “the persistence of seriously suboptimal provisions” (Clermont and Sherwin 2002:258). They

call on civilian lawyers to improve their system in light of the superior US solution (Clermont and Sherwin 2002:274).

An Italian discussant did not mince his words either. Clermont and Sherwin have fallen prey to the "reductivist fallacy" (Taruffo 2003:659). "They adduce some scattered statements made by some European scholars (mainly French), taking them out of their theoretical context and without checking to which extent they correctly represent the practice of courts in civil law countries" (Taruffo 2003:664 f.). Moreover they have failed to understand that *intime conviction* is not the same as "beyond a reasonable doubt" (Taruffo 2003:666). "They rely upon a rather naive idea of probability, roughly corresponding to the popular concept of statistical or quantitative probability, and upon a naive and unqualified idea of truth" (Taruffo 2003:669).

Acrimony is not the recipe for improving scientific understanding. Let us take a disinterested view at doctrine on both sides of the Atlantic. There is a clear difference in the handling of civil law disputes. In the US, the plaintiff prevails if only "the preponderance of the evidence" speaks in her favour. "The litigants thus share the risk of error in roughly equal fashion" (*Addington v. Texas* 441 U.S. 418, 422 (1979)). The Pennsylvania jury instructions put the standard thus:

> "Preponderance of the evidence means the claim is more likely true than not [...]. Think about an ordinary balance scale with a pan on each side to hold objects. Imagine using the scale as you deliberate in the jury room. Place all the evidence favourable to the plaintiff in one pan. Place all the evidence favourable to the defendant in the other. If the scales tip, even slightly, to the plaintiff's side, then, you must find for the plaintiff" (Pa. SSJI (Civ) 1.42).

There is some criticism of the standard in American legal literature. But this criticism is chiefly concerned with extending the standard into matters that are (close to) criminal law, with a focus on sentencing under the Federal Sentencing Guidelines (Salky and Brown 1992; Lyons 1993; Sargent 1993; Greenwald 1994; Slatkin 1997; Wysocki 2006). Another line of criticism is concerned with the robustness of judicial outcome to small perturbations, and recommends an alternative standard that weights probability with the normative desirability of the respective outcome (Orloff and Stedinger 1983:1159, 1169; Franklin 2006:161). American legal scholars thus seem to feel comfortable with the objectivists approach, and with the probabilistic interpretation of the standard of proof.

By contrast, the standard of proof in civil law countries is predominantly interpreted as not being probabilistic (Clermont and Sherwin 2002:265). The classic formulation is in art. 353 of the French Code of Criminal Procedure:

> "The law does not ask judges for an explanation of the means by which they are convinced, it does not set any particular rules by which they must assess the fullness and adequacy of the evidence; it stipulates that they must search their conscience in good faith and silently and thoughtfully ask themselves what impression the evidence given against the accused and the defence's arguments have made upon them. The law asks them only one question which sums up all of their duties 'Are you personally convinced'?"

This standard is also applied in civil law disputes (Bredin 1996:23), although there is no explicit provision to the effect in French law (Taruffo 2003:667).

In German law, the situation is even clearer. According to § 286 I 1 Code of Civil Procedure

> "Paying due regard to the entirety of the proceedings, including the evidence presented, if any, it is for the court to decide, based on its personal conviction, whether a factual claim is indeed true or not".

In the leading case, the German Supreme Court has made it clear that the judge may not content herself with a mere assessment of probabilities. Even a very high probability would not be enough. Initial doubt is acceptable. But the judge must have overcome this doubt (*Bundesgerichtshof* BGHZ 53, 245, 255 f. – *Anastasia*). This is not meant to defer to judicial discretion, but to judicial intuition (MUSIELAK-FOERSTE § 286 ZPO, R 17). The standard is an empirical one (Schulz 1992:42). The crucial feature is "the psychic state of taking a fact for true" (Schulz 1992:43). The test is built on "ethos, experience and intuition" (Schulz 1992:168).

Interestingly, in criminal procedure, the conflict is less pronounced. In the US, many instructions for criminal law juries have a subjectivist flavour. For instance, the Pennsylvania instructions read

> "To find the defendant guilty beyond a reasonable doubt, you must be convinced of [his] [her] guilt to the same degree you would be convinced about a matter of importance in your own life in which you would act with confidence and without restraint or hesitation" (Pa. SSJI (Crim) 2.01).

> "A reasonable doubt is a doubt that would cause a reasonably careful and sensible person to hesitate before acting upon a matter of importance in his or her own affairs" (Pa. SSJI (Crim) 7.01).

In the explanatory part, the document says

> "In essence, to prove facts beyond a reasonable doubt is simply to bring a jury to a particular state of certainty. It is akin to raising the temperature in the room to 75 degrees; no one may hold in their hand or find the 75th degree in the room, but one may sense the presence of that state nonetheless. Were the courts ever to develop a meter that worked with thermometer-like precision and ease of determination, judges could advise jurors on when the evidence in a case rose to a level that guilt would be shown beyond a reasonable doubt, much as a scientist could tell his or her class at what temperature they might expect the water to boil" (Pa. SSJI (Crim) 7.01).

The Massachusetts Supreme Court once put it

> "Then, what is reasonable doubt? [...] It is that state of the case, which, after the entire comparison and consideration of all the evidence, leaves the minds of jurors in that condition that they cannot say they feel an abiding conviction, to a moral certainty, of the truth of the charge [...] For it is not sufficient to establish a probability, though a strong one arising from the doctrine of chances, that the fact charged is more likely to be true than the contrary, but the evidence must establish the truth of the fact to a reasonable and moral certainty" (*Commonwealth v. Webster*, 59 Mass. 295, 320 (1850)).

This remains the law (Weinstein and Dewsbury 2006:170 f.), despite voices in the literature calling for a quantification of the standard (Saunders 2005; Tillers and Gottfried 2006; Weinstein and Dewsbury 2006), (but see Franklin 2006).

III. Norms

Whether the US or the continental approach is superior depends on the goals the legal order strives to achieve by defining the standard of proof. In the legal literature, a more modest and a more ambitious approach compete (for an overview see Stoffelmayr and Seidman Diamond 2000). At the most basic level, a consequentialist approach looks at the implications for the decision taken in this one case. At a more sophisticated level, one might also be concerned with the implications for the functioning of jurisdiction and for the place of jurisdiction in the social fabric, in particular, for perceptions of its legitimacy. In a consequentialist approach, one is concerned with both, the probabilities of different types of errors and their consequences. A utilitarian version of this approach would postulate a social welfare function and would assess the different norms in terms of their welfare implications. A cruder approach would postulate rule of thumb values for the cost of false convictions versus false acquittals and, on this basis, develop norms for assessing the tradeoff between the probabilities of the different types of errors.

All agree that the problem originates in the social cost of error. The modest approach confines itself to making the probability of certain errors sufficiently unlikely (1). The more ambitious approach is utilitarian (2). From the latter perspective, it is easy to incorporate additional concerns that have been voiced in the complementary discussion in economics. The standard of proof will be anticipated by private parties in their pre-trial decision-making. They for instance might decide to increase their level of care. Moreover, society might also care about the transaction cost incurred by the parties and the courts. These concerns can be incorporated as additional arguments into society's welfare function.

1. Acceptable Probability of Error

Frequently, the courts are not in a position to establish the facts with near certainty. Even if they subjectively believe a stated fact to be true, there is next to always at least some small objective probability that the statement is actually false (Brook 1982; Stoffelmayr and Seidman Diamond 2000:770). The most basic normative question to be addressed by the standard of proof is defining the maximum tolerable probability of this happening (Newman 1993:981). The US Supreme Court has put the question thus:

> "In a criminal case [...] the interests of the defendant are of such magnitude [...] that they have been protected by standards of proof designed to exclude as nearly as possible the likelihood of an erroneous judgement. In the administration of criminal justice, our society imposes almost the entire risk of error upon itself"
> (*Addington v. Texas*, 441 U.S. 418, 422 (1979)).

The legal order is in the same position as a statistician (Kaye 1982:501; Miceli 1990:189, 191). Both are in the business of drawing inferences from evidence they are aware might be biased. Statisticians are happy to make the inference if the probability of an unwarranted statement is sufficiently small. To do so, they compare the tested hypothesis to some nul hypothesis. In criminal law, the analogue to the tested hypothesis is the charge. The nul hypothesis is the presumption of innocence (Davis 1994:346).

A statistician will only accept the tested hypothesis if she is able to refute the competing nul hypothesis with very high probability. In the social sciences, the convention is this: the probability that the nul hypothesis is true must be below 5%. That way, statisticians make sure that alpha errors are very unlikely. An alpha error occurs if the tested hypothesis is taken to be true although, in reality, it had been false. If this standard has not been met by the available data, this is not to say that the opposite is proven to be true. In statistical jargon, the finding is simply not significant. This does not exclude that, with better data, the statement could be proven to the requisite standard (Hays 1994). Likewise, in criminal law, guilt must be proven "beyond a reasonable doubt". The law focuses on alpha errors, and requires them to be very rare. By implication, "preponderance of the evidence" requires much less certainty. The legal order tolerates a substantially higher error rate (Brook 1982:85).

One may wonder why society dislikes false convictions. Obviously, the constitutional presumption of innocence matters (Davis 1994:346). Moreover, society might want to make sure that effort, by the defendant, to convince the court of her innocence is more productive if the defendant is actually innocent (Rubinfeld and Sappington 1987:310). Last, but not least, the general acceptance of the legal system might suffer if the courts make patently false decisions (Clermont and Sherwin 2002:269, 271; Lillquist 2002:176, 185; Demougin and Fluet 2005; Franklin 2006:159).

Every empirical researcher has made the experience: she is subjectively convinced that she has indeed found the effect. Yet she is unable to refute the nul hypothesis. The finding is not significant, and therefore not publishable. A promising line of research has been a waste of time and resources. The legal order is in the same situation. The stricter one is with alpha errors, the more likely it becomes that one commits the opposite error. In statistical jargon it is called the beta error. One wrongly refutes the tested hypothesis. Empirical researchers are trained to respond by calculating test power before they engage in the generation of data. The smaller the expected effect, the larger the sample must be to meet the significance level (Hays 1994). The analogue in judicial proceedings is the amount of energy that goes into producing evidence. The stricter the standard of proof, the more effort is required by the plaintiff, or by prosecution for that matter.

Ultimately, this effort may become prohibitive. Moreover, the legal order might consider it unfair to impose such a heavy burden on the plaintiff. This concern drives the harsh criticism of the continental standard of proof cited at the beginning of this article (Clermont and Sherwin 2002). Lowering the standard of proof is, however, not the only technology for helping plaintiffs where the legal order deems this appropriate. The strongest intervention is shifting the burden of proof

to the defendant (Taruffo 2003:672); (cf. Weinstein and Dewsbury 2006:167). Making prima facie evidence acceptable is a more cautious intervention (for an instance from US law see Gertken 2003). The plaintiff may content herself with proving the preconditions of an established factual regularity. It then is for the defendant to cast sufficient doubt on the claim that the case at hand is in line with this regularity. Finally, instead of lowering the standard of proof across the board, a more lenient standard may exclusively be applied to facts that are particularly hard to prove. An example is the amount of damage, once liability has been established (as stipulated in § 287 *German Zivilprozessordnung*).

2. Utilitarian Norm

In the US literature, there is mounting unease with the exclusive focus on error rates (Kaye 1982; Orloff and Stedinger 1983:1160, 1167; Lillquist 2002). In essence, this criticism is utilitarian (Brook 1982:86; Davis 1994:348). Authors advocate defining a social welfare function. The main advantage of this approach is conceptual. Not only becomes it possible to address alpha and beta errors within one and the same framework (Tillers and Gottfried 2006:152). Quite naturally, two more arguments feature in the welfare function: the conviction of guilty defendants, and the acquittal of innocent defendants (Miceli 1990:189). Each of these four events can be weighted. The social optimum requires choosing the standard of proof such that this function is maximised.

Once one starts from a social welfare function, it is easy to add further arguments. Economists have claimed that the standard of proof should also be chosen such that innocent defendants stand a chance to credibly signal their type (Rubinfeld and Sappington 1987). Others have suggested that the standard of proof might be used to calibrate would-be defendants' incentives to exert care (Demougin and Fluet 2005; Demougin and Fluet 2006). Moreover, in maximising welfare, society might also take the transaction cost of judicial proceedings into account that has to be borne by defendants (Rubinfeld and Sappington 1987), plaintiffs (Miceli 1990), and the courts (Sanchirico 1997).

While appealing from a theoretical perspective, an utilitarian norm has a number of undesirable properties. The positive utility from convicting guilty defendants, and from not committing beta errors, is bound to differ between classes of cases. Likewise, the negative utility from committing an alpha error will not in all cases be the same (Lillquist 2002:149, 184). A variable standard of proof may be difficult to administer, and it may meet with public resistance. Actually, if the utilitarian approach is taken seriously, it is not enough to differentiate between types of offences. One would also need knowledge about the composition of the respective population (Davis 1994:347). It will be next to impossible to make this knowledge available in court. And it might normatively not be acceptable to have a different standard of proof for each and every case (as suggested by Davis 1994:350). Quite a few concerns that would matter from a strict utilitarian perspective might also not be acceptable in court, like the crime record of the defendant (Franklin 2006:164). Both with respect to practicality and to the legitimacy of the judicial sys-

tem, the more robust definition of the maximum acceptable alpha error might therefore be preferable. At any rate, court practice on both sides of the Atlantic seems to be in line with this.

IV. A Behaviourally Informed Perspective

The section on norms informs those who design the law of evidence. In the perspective advocated here, the standard of proof has to make sure that the probability of alpha errors is kept below society's level of tolerance. Achieving this goal is an empirical matter. Ultimately, whether courts make materially wrong decisions depends on what judges and jury members do. A behaviourally informed perspective is paramount. There are two competing views of this.

One line of research shows how often courts have indeed erred. If measured against the standards of rational choice decision theory, judges and jurors perform poorly (1). The competing line of research starts from a different definition of the task. Judicial decision-making is not applied science. In limited time and, more importantly, on a patently incomplete factual basis, courts have to decide on overly complex issues (2). Empirically, jurors and judges decide by constructing stories. They decide in favour of the plaintiff, or the prosecution for that matter, if it has been able to induce a coherent story that resonates with the jury member's world knowledge (3). The underlying psychological mechanism strives at making sense of the evidence. It aims at generating coherence. This is an unconscious activity (4). From this perspective, a second function of the jury instructions, and of professional legal training for the judges, becomes visible. By stressing the undesirability of alpha errors, by holding jury members accountable, the unconscious generation of coherence is altered. The intervention has clout since it musters emotions. In so doing it creates a somatic marker (5).

2. Violations of the Rational Choice Benchmark

“Statistics, not Experts” (Meadow and Sunstein 2001)! “Train Our Jurors” (Koehler 2006)! Among behavioural researchers there is not much trust in either (legal) experts, i.e. judges, nor in laymen on jury duty. If one compares (judicial) performance to the rational choice benchmark, there is indeed reason for scepticism. Statistically untrained subjects have been demonstrated to get about everything wrong that should matter according to statistical theory (Conlisk 1996). Since statistics is the most sophisticated approach to decision-making under uncertainty, this is no minor concern. Moreover, while professional judges only marginally outperform laypeople on such tasks as assessing the credibility of an eyewitness, they are massively more confident in their judgement (Sporer 2007). These concerns resonate with colourful remarks by Benjamin Cardozo (Cardozo 1921):

> “The great tides and currents which engulfed the rest of men do not turn aside in their course and pass the judges by” (168).

> "The spirit of the age, as it is revealed to each of us, is too often only the spirit of the group in which the accidents of both our education or occupation or fellowship have given us a place. No effort or revolution of the mind will overthrow utterly and at all times the empire of these subconscious loyalties" (174).
>
> "Deep below consciousness are the forces, the likes and dislikes, the predilections and the prejudices, the complex of instincts and emotions and habits and convictions, which make the man, whether he be litigant or judge" (167).

Specifically, jury instructions have been shown to do a rather poor job. To many jurors "jury instructions are like foreign movies without subtitles" (Wald 1993:111, quoting Stephen Adler). Verbal instructions have misled subjects in mock juries (Kagehiro and Stanton 1985). Estimates of what "beyond a reasonable doubt" meant ranged from 51 to 91% of certainty (Hastie 1993:101-108). A recent study even produced a range between 30 and 100% (Saunders 2005). Further studies show high variance in these estimates (Horowitz and Kirkpatrick 1996; Horowitz 1997; Stoffelmayr and Seidman Diamond 2000:777; Tillers and Gottfried 2006:156). In line with this, it could be demonstrated that subjects also were confused on other elements of the instructions (Wiener, Pritchard et al. 1995; Saxton 1998).

2. A Proper Definition of the Task

This is certainly troubling news. It makes understandable why the literary voices for quantifying the standard of proof are so strong in the US (Saunders 2005; Tillers and Gottfried 2006; Weinstein and Dewsbury 2006). The critics of the continental European approach claim: The European quest for "truth" is futile and obnoxious. It opens the door to mistake, randomness, prejudice and outright abuse (Clermont and Sherwin 2002:263, 267, 271). They counsel turning judicial procedure into an exercise in objectivity (cf. Schafer and Wiegand 2004:95). Legal decisions should be made as science like as possible.

Critics object: judicial decision-making is an extremely complex task (Pennington and Hastie 1992:189). Judges and jurors are supposed to handle vast amounts of information in limited time (Pennington and Hastie 1986:242; Tillers and Gottfried 2006:155). They are only allowed to pass judgement once they have heard the entire evidence. Pieces of evidence are not additively separable (Pennington and Hastie 1992:190). In the language of probability theory: conditional probabilities matter (but see Miceli 1990:190). The evidence is presented in a scrambled order. It normally remains incomplete (Pennington and Hastie 1986:243). Courts are aware of strategic attempts by the parties to mislead them.

3. Sense Making

The most powerful criticism of the objectivist definition of the task is, however, empirical. Judges and jury members do simply not act like miniature scientists. This is not a defect, but best they can do, given the patently incomplete evidence that is characteristic for almost all court

cases. Empirically, judges and jury members are engaged in sense making (Pennington and Hastie 1991:519); (also see Weick 1995). Decision-making is explanation based (Pennington and Hastie 1986:242; Pennington and Hastie 1988:521; Pennington and Hastie 1992:189). It is interpretative (Pennington and Hastie 1988:524). It relies on reasoning about the evidence, rather than an algebra like process (Pennington and Hastie 1988:531). Jurors attempt at creating a narrative story from the pieces of evidence they have heard (Pennington and Hastie 1986:243; Pennington and Hastie 1988:521; Pennington and Hastie 1993b:136; Pennington and Hastie 1993a:194). Decision-making is based on the construction of mental models (Pennington and Hastie 1988:521); (also see Johnson-Laird 1983).

This is an inherently constructive activity (Pennington and Hastie 1993a:194). The constructive element is particularly important, since the evidence is not presented in story-like form (Pennington and Hastie 1986:243; Pennington and Hastie 1993a:195). In this mental activity, jurors not only build on the evidence. They combine it with their world knowledge (Pennington and Hastie 1993a:195) and with their expectations of what constitutes an adequate explanation in the respective area of life (Pennington and Hastie 1988:521). In so doing, they look out for causal and intentional relations (Pennington and Hastie 1986:243). Due to this memory component, by necessity, different jurors will construct different narratives (Pennington and Hastie 1993a:196).

If the case is not utterly simple, the ultimate narrative will be composed of episodes that themselves have story form (Pennington and Hastie 1993a:197). Each episode consists of initiating events, goals, actions, consequences, and accompanying states, in a particular causal configuration (Pennington and Hastie 1988:522). Typically, the result is a hierarchy of embedded episodes (Pennington and Hastie 1986:243). It is possible that jurors try out competing stories (Pennington and Hastie 1988:522).

Story construction is instantaneous. It starts early on, with hearing the first pieces of evidence (Hastie, Penrod et al. 1983:18; Pennington and Hastie 1988:521, 524). Actually the fact that jurors have to hold on explicit judgement until they have heard the entire evidence fosters story construction (Pennington and Hastie 1992:189). The order matters in which the evidence is presented. Stories that fit better to the pieces of evidence heard initially are more likely to persist (Pennington and Hastie 1988:521).

Jurors decide by matching stories to the representation of the verdict categories given to them in the judge's instructions on the law (Pennington and Hastie 1986:243). Whether a story is accepted, or whether it is selected, depends on its goodness of fit (Pennington and Hastie 1993a:201). In this assessment, jurors look out for coverage, coherence and uniqueness (Pennington and Hastie 1992:190). They check the extent to which the story accounts for the evidence presented at trial (Pennington and Hastie 1992:190). They require the story to be coherent, meaning that it must be consistent, plausible and complete (Pennington and Hastie 1993a:198 f.). Consistency is an internal criterion, applying the laws of logic to the elements of the story. Plausibility is an attempt at matching the story with the juror's world knowledge. A

story is taken to be complete if it has all the elements a story of this kind ought to have (Pennington and Hastie 1993a:198 f.). It lacks completeness if it has conspicuous gaps (Pennington and Hastie 1988:522); (also see Kaye 1986). Finally jurors check for uniqueness. They are more likely to accept a story if they are unable to come up with a plausible alternative story, based on the same evidence (Pennington and Hastie 1992:190). If there are competing stories, jurors prefer the one that is more coherent, in the sense just explained (Pennington and Hastie 1988:530).

Coverage, coherence and uniqueness also determine the level of confidence (Pennington and Hastie 1988:521, 528; Pennington and Hastie 1992:190). Subjects are the more confident the more they feel they have a clear view of the case. In line with this, confidence is higher if both the prosecution and the defence have presented their evidence in story order (Pennington and Hastie 1988:530).

4. Consistency Maximisation

In the legal literature, many have noted that a trier of fact has to rely on her intuition (Schulz 1992:43, 168)(MUSIELAK-FOERSTE § 286 ZPO, R 17). Some have even linked this to Gestalt psychology (Schafer and Wiegand 2004). It has been said that, to convict the defendant, the fact-finder must be in the "subjective state of near certitude" (*Jackson v. Virginia*, 443 U.S. 307, 315 (1979)).

The underlying mental process is relatively well understood. In line with the basic claim from Gestalt psychology (Markus and Zajonc 1985), the assessment of the evidence is holistic (Simon 2004:560, 562). Decision-makers aim at forging coherence (Simon, Pham et al. 2001). They strive at parallel constraint satisfaction (Simon and Holyoak 2002). They treat the evidence somehow like the pieces of a jigsaw puzzle, and try to fit as many pieces as possible, with the twist that they are happy to go with a less than perfect fit of the elements, as long as the resulting picture looks good. Based on a connectionist cognitive architecture, decision-making progresses bidirectionally (Holyoak and Simon 1999). Not only do facts determine conclusions. Potential conclusions also affect the perception of the evidence. The mental model reconfigures itself until maximal coherence is achieved (Simon 2004:522).

The process has been modelled mathematically. Based on the model, and on assumptions about the parameters, simulations have been run. The results of these simulations have been compared to experimental findings, with a fairly good fit.

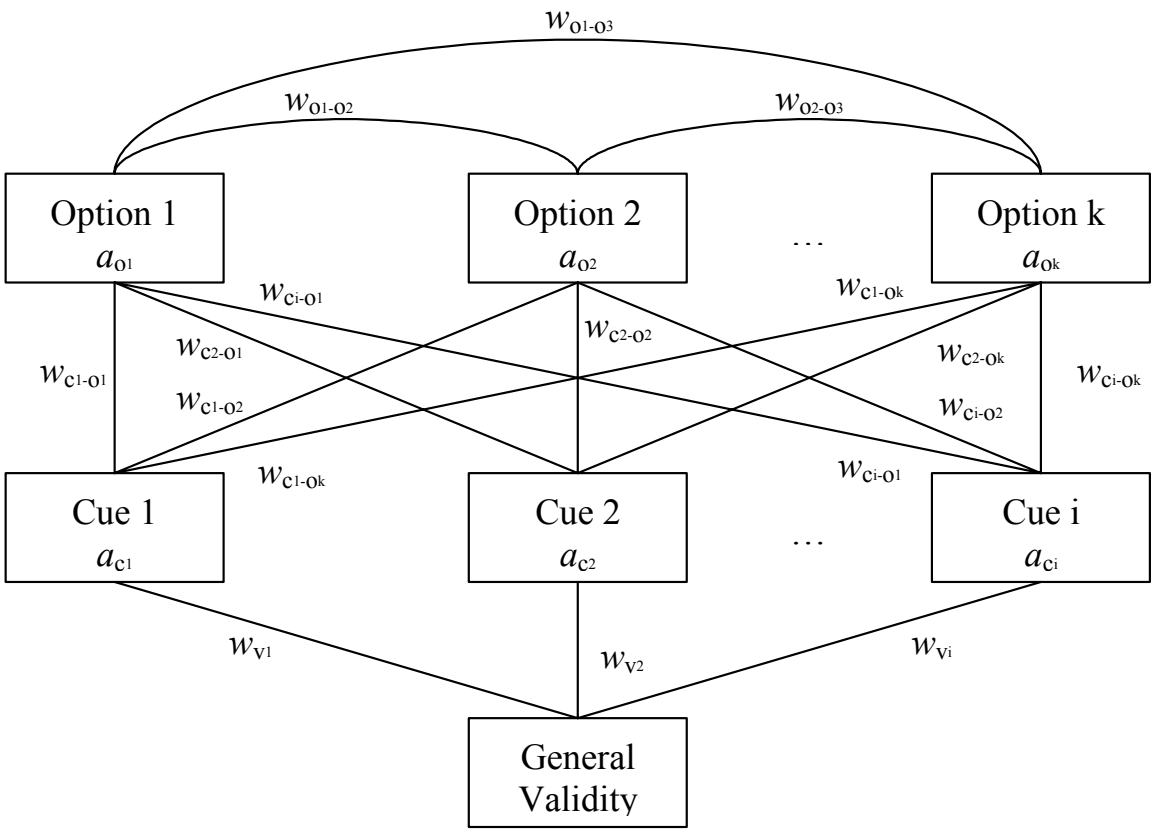


Figure 1
Model of Consistency Maximisation
Glöckner Betsch 2008: 218

The basic components of the process are these: the evidence activates cues. These cues are positively or negatively related to decision options. If options exclude each other, they are negatively related. As a first step, the evidence provides support for at least some of the options. But the process does not stop here. Depending on how strongly an option is activated initially, it propels positive activation back to the supporting evidence, and negative activation to the conflicting evidence, and to the competing options. Based on the resulting re-assessment of the evidence, the process enters the next iteration. It is repeated as long as the marginal changes in the activation of the options are substantial. The most simple decision criterion is the difference in the final activation of options. The option with the highest positive activation is chosen, provided all competing options are sufficiently less activated. A more elaborate decision criterion also takes the structure of the network into account that supports the decision. It also checks for the overall consistency of this network (Glöckner 2008; Glöckner and Betsch 2008).

For the decision maker, this mechanism has a number of advantages. First and foremost, the mechanism is able to handle vast amounts of information (Simon 2004:544; Engel 2008); (also see Glimcher 2003). The severe limitations for the deliberate handling of information are not present. As is well known, the memory span for deliberate processing is limited to seven items plus minus one in most people (Baddeley 1986). More importantly even, parallel constraint satis-

faction enables the individual to decide although the problem is overly complicated (Simon 2004:517), or known to be ill-defined (Engel 2008); (on these two categories of tasks see Gigerenzer, Todd et al. 1999). The process makes the best out of partial evidence. In gradually accentuating the differences between options, the individual is much more often in a position to decide, although full certainty is not to be had. The mechanism is therefore best interpreted as an enabling device (cf. Simon 2004:517, 546). It empowers the individual to make relatively good decisions although there is no chance to fully understand the problem.

These advantages come at a price (Engel 2008). Intuition is not a calculator. The mechanism is programmed to transform the information input. Over the consecutive iterations, information supporting the final decision is overvalued. Conflicting evidence is undervalued. Information is thus polarised (Simon 2004:523). This process has been dubbed a coherence shift (Simon, Pham et al. 2001). It has been shown that these shifts are pronounced (Simon 2004:532). The higher the confidence of the decision maker, the stronger the re-evaluation of evidence (Simon, Snow et al. 2004:819). Since the intuition is formed in the automatic system, decision-makers are not aware of the coherence shift (Simon 2004:545). Ultimately, coherence shifts fade away (Simon 2004:542). There is, however, the possibility that a shift resulting from one decision spills over to another decision in the near future (Simon 2004:546).

These general findings have been shown to hold in the assessment of evidence by the members of mock juries (Pennington and Hastie 1988:521; Simon 2004:523). Individual predilections are able to influence their judgement via the automatic system (Simon 2004:550). The effects are even present if jury members are admonished to withhold judgement until they have heard the entire evidence (Simon 2004:533).

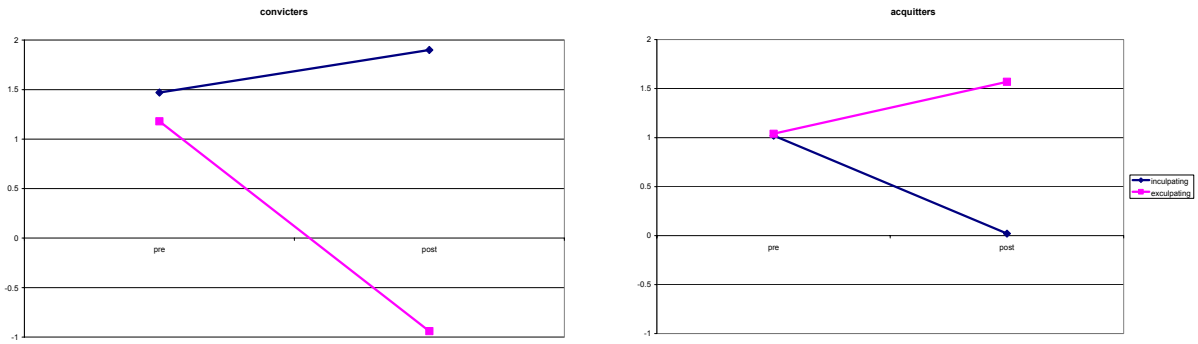
5. Rescued by Emotions?

Continental observers have said: the subjectivist model of proof is built on “ethos, experience, and intuition” (Schulz 1992:168) (also see Taruffo 2003:667). Apparently, the civilian legal orders do not blindly trust intuition. They are (at least intuitively) sensitive to the risk that judicial intuition is misled: by clever parties, by prejudice, or simply by heedless decision-making. Yet the response to this risk is not an attempt at turning judicial procedure into a quasi scientific affair. Rather these legal orders have recourse to yet another determinant of human behaviour that is not under direct conscious control. They appeal to emotions. Specifically they call on judges' professional ethos, and on jurors' sense of republican duty. If this intervention is to have the desired effect, two conditions must be met: it must be possible to calibrate the subconscious mechanism of constraint maximisation. And the institutional interventions that make judicial responsibility salient must be sufficiently powerful to bring about the normatively expected level of scrutiny.

For the purposes of this paper, the most relevant sub-question is this: are standards of proof mute if triers of fact rely on their automatic system? One might be afraid that this would happen. After

all it is the function of parallel constraint satisfaction to give individuals more scope for decision-making under uncertainty. In the extreme, this could mean that the mind always comes down on one side. This would be at variance with the two cornerstones of the law of evidence: the standard of proof, and the burden of proof. The law wants two things: the court shall only take a fact for true if the requisite standard of proof has been met; if the evidence remains inconclusive, the court shall decide as defined by the burden of proof. Behaviourally, this doctrinal model presupposes two things: constraint satisfaction must be able to fail. In the definition of success vs. failure, individuals must be able to apply different, exogenously imposed standards of strictness.

The psychological correlate of the standard of proof is confidence. It has been shown that reported confidence levels do not mirror the degree to which the evidence has remained un-altered in the process of forming an intuition. On the contrary: subjects that exhibit particularly strong coherence shifts are also most confident of their final judgement (Simon 2004:532). The most troubling piece of evidence is an experiment with two stages. In the first stage, subjects were asked to rate the credibility of unrelated vignettes. In the second stage, material from these vignettes was presented as evidence in a mock jury trial. The trial was framed as a criminal case. Subjects were reminded that the respective standard of proof was beyond a reasonable doubt. In line with the general parallel constraint satisfaction model, subjects inflated the relevance of evidence that supported their final decision, and they deflated conflicting evidence. However those who ultimately convicted the defendant deflated exculpating evidence much stronger than acquitters deflated inculpating evidence (Simon 2004:531)[2]. One interpretation of this result is this: the only effect of the stricter standard of proof is to deflate conflicting evidence more strongly.



	convicters		acquitters	
	pre	post	pre	post
inculpating	1.47	1.90	1.02	0.02
exculpating	1.18	-0.94	1.04	1.57

Figure 2
Dan Simon's Experiment
evidence was rated on a scale from -5 to +5

2 The difference was, however, not statistically significant, personal communication from Dan Simon.

Happily, this interpretation seems to be wrong. In our lab, we have conducted three studies with a total of 238 subjects. In two of the studies, we have given two groups of subjects the same evidence. One group has been instructed to convict the defendant only if her guilt was beyond a reasonable doubt. The other group has been instructed to decide to the detriment of the defendant if the preponderance of the evidence speaks against her. For these experiments, we have used (translated) jury instructions from the US. Results were as follows:

Figure 3
Erfurt Experiment

Subjects only convict if their estimate of guilt is above 50 %[3]. While some convict under beyond reasonable doubt at a disturbingly low subjective probability of guilt, the majority of red diamonds is in the normatively expected places: subjects require a higher subjective probability under the stricter standard. If they acquit under preponderance of the evidence, their subjective probability of guilt is not above 50%. If they acquit although the subjective probability is above 50 %, this is always under the beyond a reasonable doubt instruction. Subjects are most confident in their decision, if they believe the case to be clear, i.e. if they acquit and subjective probability is low, or if they convict and subjective probability of guilt is high.

In our experiments, total coherence shifts are indeed more pronounced when subjects convict. However, on closer inspection this effect does not have a normatively troublesome cause. As Figure 4 shows, it is caused by the fact that, under the beyond a reasonable doubt instruction,

3 The one outlier notwithstanding.

subjects do not devalue inculpating evidence as strongly as under the preponderance of the evidence instruction[4]. Or more legally: many subjects acquit "for want of evidence", and are not induced to devalue conflicting evidence to come up with a decision "for cause". This is exactly what the legal order wants to achieve with the stricter standard of proof (Glöckner and Engel 2008).

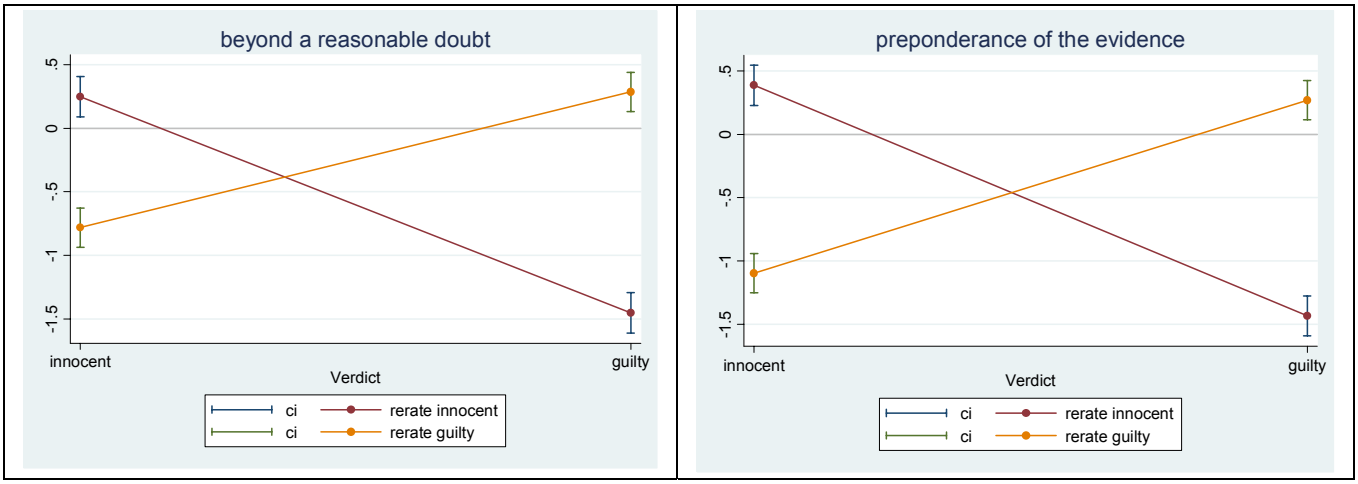


Figure 4
Rerating of Evidence by Standard of Proof

It is relatively easy to explain this finding in light of the storytelling model. When instructed to convict only if guilt is beyond a reasonable doubt, subjects apply a stricter standard of coverage, coherence and uniqueness. They acquit already if minor pieces of evidence do not find a place in the story told by prosecution. They do not tolerate a shade of doubt regarding the logical consistency of the story. In light of their world knowledge, they require high plausibility for this story. They apply a strict standard of scrutiny when testing the completeness of the prosecution story. Finally they impose a large minimum gap in plausibility if there is more than one story that can be told, based on the evidence.

The easiest way to translate these techniques of calibrating judgement into mental mechanism is manipulating the threshold. Irrespective of the standard of proof, the automatic system always processes the information as long as this results in more than minimal changes in the activation of options. However, depending on the standard of proof the mind asks for a larger minimum difference in the activation of the preferred alternative, compared to all competing alternatives. This explanation fits best for the minimum gap in plausibility between competing stories. It also works for the level of plausibility, given world knowledge. For the remaining elements of the story model, one needs a richer decision criterion. It must also take the structure of the network into account that supports the decision, as proposed by the parallel constraint satisfaction model.

4 In a regression with the rerating of conflicting evidence as the dependent variable, frame and probability manipulations as controls, and robust standard errors, the regressor for standard of proof is (weakly) significant if subjects convict, beta = .18, p = .09)

Thus far, it has been shown that different standards of proof are psychologically feasible. But how is the legal order able to induce these standards? What does “ethos” mean in psychological terms? By which mental mechanism are the thresholds for discriminating options, and for assessing the overall consistency of the supporting network, calibrated? Which is the psychological effect of the many institutional interventions that make judicial responsibility salient? To name only a few: only a very small number of lawyers are chosen to be judges. By the profession, being a judge is regarded as an honour. Judges are appointed in a ceremonial way. Throughout judicial procedure, all are reminded repeatedly that the judge holds an office of the people, and speaks in the people's name. Likewise: which is the psychological effect of being selected as a jury member, and of the jury instructions in particular?

Psychologically speaking, these interventions make accountability salient (Engel 2007). Accountability induces subjects to invest more cognitive effort (McAllister, Mitchell et al. 1979; Weldon and Gargano 1988; Tetlock, Skitka et al. 1989:633; Ashton 1992) and to become open to greater complexity of thought (Tetlock and Kim 1987). Specifically, the subjectivist standard of proof in continental law directly appeals to personal responsibility. “Do only convict if you are sure you can take on responsibility for this decision”! The wording of jury instructions for the reasonable doubt standard in the US sounds pretty similar to this. It seems that, in judicial practice, the objectivist norm that wants to minimise the alpha error is translated into a requirement with a pronounced subjectivist flavour. By contrast, the preponderance of the evidence instructions can be interpreted as a tool for exonerating jury members from personal responsibility. Society is happy with quite a number of materially wrong judgements. Accountability is reduced to avoiding gross errors.

Ultimately, jury instructions are just words. Society has no chance to check, in the individual instance, whether these words have had their desired effects. There is no ex post control. There are no sanctions. Why is one nonetheless justified to believe that the instructions do their job? Arguably, at this point emotions come into play. The instructions, together with the described institutional arrangement, set a somatic marker (Damasio 1994; Bechara and Damasio 2005). Somatic markers are particularly instrumental if decisions have to be taken under conditions of high complexity and uncertainty (Damasio 1994:167 f.). They reduce options and focus attention (Damasio 1994:173). They work as a biasing device (Damasio 1994:174, 198). They induce the individual to take a risk very seriously (Damasio 1994:212). They are particularly instrumental if low probability events are paired with high affective reactions (Loewenstein, Weber et al. 2001; Slovic, Finucane et al. 2002). Much the same way, the jury instructions set a somatic marker that biases judgement away from committing alpha errors. For the argument to go through, the somatic marker must be attached to one of the potential outcomes, to conviction.

The mental mechanism has been studied in the Iowa gambling task. Experimental subjects have been given a choice between two classes of lotteries. In the first class, gains were higher than in the second. However, in the first class losses were dramatic, whereas they were mild in the second class. Subjects did not get this information in advance. They had to learn from experience. Ordinary subjects did so quite rapidly, and quite reliably. However, subjects with brain damage

in the ventromedial prefrontal cortex, the region of the brain responsible for integrating emotional information with decisionmaking, failed miserably on this task. They were overly attracted by the prospect of immediate gain, and went quickly bankrupt (Bechara, Damasio et al. 1994).

V. Conclusion

The difference between US and continental standards of proof is not just doctrinal. It also is not just a matter of degree. These legal orders conceptualise proof differently. Some US authors even advocate a different norm. They would want the standard of proof to result from maximising a social welfare function. The same approach is prevalent in the discussion of standards of proof in the economics literature. If one accepts the utilitarian norm, one is however forced to have different standards of proof for different classes of offences. Ultimately one even needs information about the respective composition of the population of defendants. This is not what legal orders do, on either side of the Atlantic. Judicial practice is captured by a more modest norm. It defines the maximum tolerable number of materially wrong decisions. In statistical jargon, it is exclusively concerned with the alpha error.

Judicial proceedings are characterised by a high degree of complexity. Normally, decisions have to be taken on an incomplete factual basis. Human decision-makers are unable to take such decisions in a quasi scientific manner. By contrast, the consciously not accessible automatic system is made for such tasks. The parallel constraint satisfaction mechanism calculates the relatively best way to make sense of the available evidence. In so doing it polarises judgement. Evidence supporting the final decision is inflated. Conflicting evidence is discounted.

While the automatic system polarises, it nonetheless is sensitive to exogenous intervention. A high standard of proof increases the minimum distance in activation between the preferred and all alternative decisions, and it increases the minimum level of overall consistency necessary for taking action. The stricter standard is imposed on the automatic system by a somatic marker. The jury instructions appeal to jurors' ethos. They make accountability towards the defendant, and towards society at large, salient.

The subjectivist standard of proof in continental legal orders therefore mirrors the mental activity of real judges and jurors. It is descriptively correct, and it in principle is able to reach the stated normative goal. Some US observers do not deny the descriptive statement. But they are opposed to the prescriptive part (Clermont and Sherwin 2002:271; Simon 2004:579). Judicial intuition is indeed not foolproof. Since it partly relies on idiosyncratic memory (Simon, Snow et al. 2004:536 f.), the outcome is not fully predictable (cf. Engel 2005; Engel 2008). For instance, the members of mock juries differ in the strictness of standards (Saunders 2005; Franklin 2006). Inadmissible evidence (Simon 2004:538) and stereotypes influence judgement (Simon 2004:542). If jurors believe in an outcome, this biases how they assess the evidence (Simon, Snow et al. 2004:827). Coherence shifts continuously, even if jurors are exhorted to withhold judgement until they have heard the entire evidence (Pennington and Hastie 1992:190; Simon 2004:534,

539, 551). This can be exploited by the parties. They can manipulate the order in which they present the evidence, in the interests of tilting juror judgement.

Continental law therefore has no reason to be haughty. The subjectivist standard of proof is far from perfect. Judicial performance might well be improved by adding objectivist elements. Actually, this is already the case. For instance, continental courts are happy to rely on DNA evidence, and to assess its quality with formal statistical tools. The practically most important, and the furthest reaching objectivist element is the obligation to write down reasons. This is not meant to be a description of mental process. But it also is not just lip service. The need to later give an explicit justification is anticipated when the decision is taken. It influences the mental representation of the case (Engel 2007).

Conversely, US law, in light of the knowledge about juror psychology, might become more sceptical about academic attempts to erase subjectivist traces from the law of evidence. And it might want to explore the psychological feasibility of the preponderance of the evidence standard. How good are jurors at finding that a claim is "probably true"? Do they perform more reliably if they have a chance to compare the plausibility of two competing stories? Does the procedure of civil law disputes pay due regard to the conditions for the formation of juror intuition?

References

ASHTON, ROBERT H. (1992). "Effects of Justification and a Mechanical Aid on Judgment Performance." Organizational Behavior and Human Decision Processes **52**: 292-306.

BADDELEY, ALAN D. (1986). Working Memory. Oxford; New York, Clarendon Press; Oxford University Press.

BECHARA, ANTOINE and ANTONIO DAMASIO (2005). "The Somatic Marker Hypothesis. A Neural Theory of Economic Decisions." Games and Economic Behavior **52**: 336-372.

BECHARA, ANTOINE, ANTONIO DAMASIO, HANNA DAMASIO and STEVEN W. ANDERSON (1994). "Insensitivity to Future Consequences Following Damage to Human Prefrontal Cortex." Cognition **50**: 7-15.

BREDIN, JEAN-DENIS (1996). "Le doute et l'intime conviction." Droits. Revue Francaise de Théorie Juridique **23**: 21-29.

BROOK, JAMES (1982). "Inevitable Errors. The Preponderance of the Evidence Standard in Civil Litigation." Tulsa Law Review **18**: 79-109.

CARDOZO, BENJAMIN N. (1921). The Nature of the Judicial Process. New Haven, Yale University Press.

CLERMONT, KEVIN M. and EMILY SHERWIN (2002). "A Comparative View of Standards of Proof." American Journal of Comparative Law **50**: 243-276.

CONLISK, JOHN (1996). "Why Bounded Rationality?" Journal of Economic Literature **34**: 669-700.

DAMASIO, ANTONIO R. (1994). Descartes' Error. Emotion, Reason, and the Human Brain. New York, Putnam.

DASTON, LORRAINE and PETER GALISON (2007). Objectivity. New York; Cambridge, Mass., Zone Books.

DAVIS, MICHAEL L. (1994). "The Value of Truth and the Optimal Standard of Proof in Legal Disputes." Journal of Law, Economics and Organization **10**: 343-359.

DEMOUGIN, DOMINIQUE and CLAUDE FLUET (2005). "Deterrence vs. Judicial Error. A Comparative View of Standards of Proof." Journal of Institutional and Theoretical Economics **161**: 193-206.

DEMOUGIN, DOMINIQUE and CLAUDE FLUET (2006). "Preponderance of Evidence." European Economic Review **50**: 963-976.

ENGEL, CHRISTOPH (2005). Generating Predictability. Institutional Analysis and Institutional Design. Cambrige, Cambridge University Press.

ENGEL, CHRISTOPH (2007). The Psychological Case for Obliging Judges to Write Reasons. The Impact of Court Procedure on the Psychology of Judicial Decision Making. Christoph Engel und Fritz Strack. Baden-Baden, Nomos**:** 71-109.

ENGEL, CHRISTOPH (2008). Institutions for Intuitive Man. Better Than Conscious? Christoph Engel und Wolf Singer. Cambridge, MIT Press**:** 391-410.

ENGEL, CHRISTOPH and WOLF SINGER, Eds. (2008). Better Than Conscious? Cambridge, MIT Press.

FRANKLIN, JAMES (2006). "Case Comment – United States v. Copeland, 369 F.Supp.2d 275 (E.D.N.Y. 2005): Quantification of the 'Proof Beyond Reasonable Doubt' Standard." Law, Probability and Risk **5**: 159-165.

GERTKEN, RHEA-T. (2003). "Causation in Retaliation Claims: Conflict Between the Prima Facie Case and the Plaintiff's Ultimate Burden of Pretext." Washington University Law Quarterly **81**: 151-174.

GIGERENZER, GERD, PETER M. TODD and ABC RESEARCH GROUP (1999). Simple Heuristics that Make us Smart. New York, Oxford University Press.

GLIMCHER, PAUL W. (2003). Decisions, Uncertainty, and the Brain. The Science of Neuroeconomics. Cambridge, Mass., MIT Press.

GLÖCKNER, ANDREAS (2008). How Evolution Outwits Bounded Rationality. The Efficient Interaction of Automatic and Deliberate Processes in Decision Making and Implications for Institutions. Better Than Conscious? Christoph Engel und Wolf Singer. Boston, MIT Press**:** 259-284.

GLÖCKNER, ANDREAS and TILMANN BETSCH (2008). "Modeling Option and Strategy Choices with Connectionist Networks. Towards an Integrative Model of Automatic and Deliberate Decision Making." Judgement and Decision Making **3**: 215-228.

GLÖCKNER, ANDREAS and CHRISTOPH ENGEL (2008). Can We Trust Intuitive Jurors? An Experimental Analysis

GREENWALD, LAUREN (1994). "Relevant Conduct and the Impact of the Preponderance Standard of Proof under the Federal Sentencing Guidelines. A Denial of Due Process." Vermont Law Review **18**: 529-564.

HASTIE, REID (1993). Inside the Juror. The Psychology of Juror Decision Making. Cambridge England; New York, NY, Cambridge University Press.

HASTIE, REID, STEVEN PENROD and NANCY PENNINGTON (1983). Inside the Jury. Cambridge, Mass., Harvard University Press.

HAYS, WILLIAM LEE (1994). Statistics. Fort Worth, Harcourt College Publishers.

HOLYOAK, KEITH J. and DAN SIMON (1999). "Bidirectional Reasoning in Decision Making by Constraint Satisfaction." Journal of Experimental Psychology: General **128**: 1-29.

HOROWITZ, IRWIN A. (1997). "Reasonable Doubt Instructions. Commonsense Justice and Standard of Proof." Psychology, Public Policy, and Law **3**: 285-302.

HOROWITZ, IRWIN A. and LAIRD C. KIRKPATRICK (1996). "A Concept in Search of a Definition. The Effects of Reasonable Doubt Instructions on Certainty of Guilt Standards and Jury Verdicts." Law and Human Behavior **20**: 655-670.

JOHNSON-LAIRD, PHILIP N. (1983). Mental Models. Towards a Cognitive Science of Language, Inference and Consciousness. Cambridge, Cambridge University Press.

KAGEHIRO, DOROTHY K. and W. CLARK STANTON (1985). "Legal vs. Quantified Definitions of Standard of Proof." Law and Human Behavior **9**: 159-178.

KAYE, DAVID (1982). "The Limits of the Preponderance of the Evidence Standard. Justifiable Naked Statistical Evidence and Multiple Causation." Law and Social Inquiry **7**: 487-516.

KAYE, DAVID H. (1986). "Do We Need a Calculus of Weight to Understand Proof Beyond a Reasonable Doubt?" Boston University Law Review **66**: 657-672.

KEGEL, GERHARD (1967). Der Individualanscheinsbeweis und die Verteilung der Beweislast nach überwiegender Wahrscheinlichkeit. Das Unternehmen in der Rechtsordnung. Festgabe für Heinrich Kronstein. Kurt Biedenkopf. Karlsruhe, C.F.Müller**:** 321-344.

KOEHLER, JONATHAN (2006). Train our Jurors. Heuristics and the Law. Gerd Gigerenzer und Christoph Engel. Cambridge, MIT Press**:** 303-326.

LILLQUIST, ERIC (2002). "Recasting Reasonable Doubt. Decision Theory and the Virtues of Variability." UC Davis Law Review **36**: 85-198.

LOEWENSTEIN, GEORGE, ELKE U. WEBER, CHRISTOPHER K. HSEE and NED WELCH (2001). "Risk as Feelings." Psychological Bulletin **127**: 267-286.

LYONS, DANIEL J. (1993). "Federal Sentencing Guidelines. Retaining the Preponderance Standard of Proof." St. John's Law Review **67**: 639-653.

MARKUS, HAZEL and ROBERT B. ZAJONC (1985). The Cognitive Perspective in Social Psychology. Handbook of Social Psychology. Gardner Lindzey. New York, Random House**:** 137-230.

MCALLISTER, DANIEL W., TERENCE R. MITCHELL and LEE R. BEACH (1979). "The Contingency Model for the Selection of Decision Strategies. An Empirical Test of the Effects of Significance, Accountability, and Reversibility." Organizational Behavior and Human Decision Processes **24**: 228-244.

MEADOW, WILLIAM and CASS R. SUNSTEIN (2001). "Statistics, not Experts." Duke Law Journal **51**: 629-646.

MICELI, THOMAS (1990). "Optimal Prosecution of Defendants Whose Guilt is Uncertain." Journal of Law, Economics and Organization **6**: 189-201.

NEWMAN, JON (1993). "Beyond 'Reasonable Doubt'." New York University Law Review **68**: 979-1002.

ORLOFF, NEIL and JERRY STEDINGER (1983). "A Framework for Evaluating the Preponderance of the Evidence Standard." University of Pennsylvania Law Review **131**: 1159-1174.

PENNINGTON, NANCY and REED HASTIE (1986). "Evidence Evaluation in Complex Decision Making." Journal of Personality and Social Psychology **51**: 242-258.

PENNINGTON, NANCY and REED HASTIE (1988). "Explanation-Based Decision Making. Effect of Memory Structure on Judgement." Journal of Experimental Psychology: Learning, Memory and Cognition **14**: 521-533.

PENNINGTON, NANCY and REED HASTIE (1991). "A Cognitive Theory of Juror Decision Making. The Story Model." Cardozo Law Review **13**: 519-557.

PENNINGTON, NANCY and REED HASTIE (1992). "Explaining the Evidence. Tests of the Story Model for Juror Decision Making." Journal of Personality and Social Psychology **62**: 189-206.

PENNINGTON, NANCY and REED HASTIE (1993a). The Story Model for Juror Decision Making. Inside the Juror. The Psychology of Juror Decision Making. Reed Hastie. Cambridge, Cambridge University Press**:** 192-221.

PENNINGTON, NANCY and REID HASTIE (1993b). "Reasoning in Explanation-Based Decision-Making." Cognition **49**: 123-163.

RUBINFELD, DANIEL L. and DAVID E.M. SAPPINGTON (1987). "Efficient Awards and Standards of Proof in Judicial Proceedings." Rand Journal of Economics **18**: 308-315.

SALKY, STEVEN M. and BLAIR G. BROWN (1992). "The Preponderance of Evidence Standard in Sentencing." American Criminal Law Review **29**: 907-918.

SANCHIRICO, CHRIS WILLIAM (1997). "The Burden of Proof in Civil Litigation. A Simple Model of Mechanism Design." International Review of Law and Economics **17**: 431-447.

SARGENT, JOSEPH P. (1993). "The Standard of Proof under the Federal Sentencing Guidelines. Raising the Standard to Beyond a Reasonable Doubt." Wake Forest Law Review **28**: 463-485.

SAUNDERS, HARRY D. (2005). Quantifying Reasonable Doubt: A Proposed Solution to an Equal Protection Problem http://law.bepress.com/expresso/eps/881.

SAXTON, BRADLEY (1998). "How Well do Jurors Understand Jury Instructions? A Field Test Using Real Juries and Real Trials in Wyoming." Land and Water Law Review **33**: 59-190.

SCHAFER, BURKHARD and OLAV K. WIEGAND (2004). "Incompetent, Prejudiced and Lawless? A Gestalt-Psychological Perspective on Fact Finding in Law as Learning." Law, Probability and Risk **3**: 93-108.

SCHULZ, JOACHIM (1992). Sachverhaltsfeststellung und Beweistheorie. Elemente einer Theorie strafprozessualer Sachverhaltsfeststellung. Köln, Heymanns.

SIMON, DAN (2004). "A Third View of the Black Box. Cognitive Coherence in Legal Decision Making." University of Chicago Law Review **71**: 511-586.

SIMON, DAN and KEITH J. HOLYOAK (2002). "Structural Dynamics of Cognition. From Consistency Theories to Constraint Satisfaction." Journal of Personality and Social Psychology **6**: 283-294.

SIMON, DAN, LIEN B. PHAM, QUANG A. LE and KEITH J. HOLYOAK (2001). "The Emergence of Coherence Over the Course of Decision Making." Journal of Experimental Psychology: Learning, Memory and Cognition **27**: 1250-1260.

SIMON, DAN, CHADWICK J. SNOW and STEPHEN J. READ (2004). "The Redux of Cognitive Consistency Theories: Evidence Judgments by Constraint Satisfaction." Journal of Personality and Social Psychology **86**: 814-837.

SLATKIN, STEPHANIE C. (1997). "The Standard of Proof at Sentencing Hearings under the Federal Sentencing Guidelines. Why the Preponderance of the Evidence Standard is Constitutionally Inadequate." University of Illinois Law Review: 583-609.

SLOVIC, PAUL, MELISSA FINUCANE, ELLEN PETERS and DONALD G. MACGREGOR (2002). "Risk as Analysis and Risk as Feelings: Some Thoughts about Affect, Reason, Risk, and Rational." Risk Analysis **24**: 311-322.

SPORER, SIEGFRIED LUDWIG (2007). Evaluating Eyewitness Testimony. The Fallacies of Intuition. The Impact of Court Procedure on the Psychology of Judicial Decision Making. Christoph Engel und Fritz Strack. Baden-Baden, Nomos**:** 111-150.

STOFFELMAYR, ELISABETH and SHARI SEIDMAN DIAMOND (2000). "The Conflict Between Precision and Flexibility in Explaining 'Beyond a Reasonable Doubt'." Psychology, Public Policy, and Law **6**: 769-787.

TARUFFO, MICHELE (2003). "Rethinking the Standards of Proof." American Journal of Comparative Law **51**: 659-677.

TETLOCK, PHILIP E. (1983). "Accountability and Complexity of Thought." Journal of Personality and Social Psychology **45**: 74-83.

TETLOCK, PHILIP E. and JAE II KIM (1987). "Accountability and Judgment Processes in a Personality Prediction Task." Journal of Personality and Social Psychology **52**: 700-709.

TETLOCK, PHILIP E., LINDA SKITKA and RICHARD BOETTGER (1989). "Social and Cognitive Strategies for Coping with Accountability. Conformity, Complexity, and Bolstering." Journal of Personality and Social Psychology **57**: 632-640.

TILLERS, PETER and JONATHAN GOTTFRIED (2006). "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**: 135-157.

WALD, PATRICIA (1993). "Guilt Beyond a Reasonable Doubt. A Norm Gives Way to Numbers." University of Chicago Legal Forum **3**: 101-126.

WEICK, KARL E. (1995). Sensemaking in Organizations. Thousand Oaks, Sage Publications.

WEINSTEIN, JACK B. and IAN DEWSBURY (2006). "Comment on the Meaning of 'Proof Beyond a Reasonable Doubt'." Law, Probability and Risk **5**: 167-173.

WELDON, ELIZABETH and GINA M. GARGANO (1988). "Cognitive Loafing. The Effects of Accountability and Shared Responsibility on Cognitive Effort." Personality and Social Psychology Bulletin **14**: 159-171.

WIENER, RICHARD L., CHRISTINE C. PRITCHARD and MINDA WESTON (1995). "Comprehensibility of Approved Jury Instructions in Capital Murder Cases." Journal of Applied Psychology **80**: 455-467.

WYSOCKI, MICHAEL D. (2006). "Beyond a Reasonable Doubt. The Effects of Blakely v. Washington, United States v. Booker, and the Future of the Federal Sentencing Guidelines." Texas Tech Law Review **38**: 495-530.