Social Studies of Science

http://sss.sagepub.com/

The Bayesian approach to forensic evidence: Evaluating, communicating, and distributing responsibility

Corinna Kruse

Social Studies of Science 2013 43: 657 originally published online 15 March 2013

DOI: 10.1177/0306312712472572

Article

S|S|S

Social Studies of Science
43(5) 657–680

DOI: 10.1177/0306312712472572
sss.sagepub.com
SAGE

The Bayesian approach to forensic evidence: Evaluating, communicating, and distributing responsibility

Corinna Kruse
Department of Thematic Studies – Technology and Social Change, Linköping University, Linköping, Sweden

Abstract
This article draws attention to communication across professions as an important aspect of forensic evidence. Based on ethnographic fieldwork in the Swedish legal system, it shows how forensic scientists use a particular quantitative approach to evaluating forensic laboratory results, the Bayesian approach, as a means of quantifying uncertainty and communicating it accurately to judges, prosecutors, and defense lawyers, as well as a means of distributing responsibility between the laboratory and the court. This article argues that using the Bayesian approach also brings about a particular type of intersubjectivity; in order to make different types of forensic evidence commensurable and combinable, quantifications must be consistent across forensic specializations, which brings about a transparency based on shared understandings and practices. Forensic scientists strive to keep the black box of forensic evidence – at least partly – open in order to achieve this transparency.

Keywords
black box, forensic evidence, mechanical objectivity, uncertainty

Forensic evidence is receiving more and more attention within Science and Technology Studies (STS). Discussion ranges from histories of how particular types of evidence have been assigned important roles in criminal justice (Cole, 2001; Lynch et al., 2008), to forensic data exchange (Prainsack and Toom, 2010), the consequences of forensic databases (e.g. Dahl and Sætnan, 2009; Hindmarsh and Prainsack, 2010; McCartney, 2006), to the relationship between science and law (Jasanoff, 2001, 2006), to forensic technologies

Corresponding author:
Corinna Kruse, Department of Thematic Studies – Technology and Social Change, Linköping University, 581 83 Linköping, Sweden.
Email: Corinna.Kruse@liu.se

that have been described as 'articulate collectives' (M'charek, 2008) of large numbers of actors, both human and nonhuman.

A topic that has received less attention is how forensic evidence is transported between different professional constituencies. Analyses of expert witnessing (Daemmrich, 1998; Halfon, 1998) and of probabilistic evidence in court (Lynch et al., 2008), as well as jury reception of forensic evidence (Kaye and Koehler, 1991) have only touched upon the issue that forensic evidence is produced in order to be communicated across professions.

In this article, I want to draw attention to an important aspect of forensic evidence– that analyses are commissioned and performed in order to provide answers to questions that are relevant to police investigations and court proceedings, and that these answers must be expressed and communicated across the judicial system's professions. I will discuss the Bayesian approach,[1] a quantitative approach to producing and conveying these answers, introduced into forensic science by Cook et al. (1998b) and adopted predominantly by European forensic science laboratories.

A recent judgment in the United Kingdom (*R v T* (2010) EWCA Crim 2439) has stirred up what might develop into a controversy about this approach. The justices in this case criticized an expert opinion on footwear marks for lack of precision and transparency, and members of the forensic science community suggested that the justices, unfortunately, must have misunderstood central aspects of the Bayesian approach (Aitkin et al., 2011; Berger et al., 2011; Robertson et al., 2011). Rather than aiming to resolve their differences, this article will illuminate some of the issues that the Bayesian approach raises about transporting knowledge from the field of expertise where it is produced to another field where it is used.

I will take my point of departure mainly from the forensic scientists' perspective on their work in order to show that they used the Bayesian approach to manage the tension between inevitable uncertainty and the necessity to deliver meaningful results, as well as to communicate with other members of the judicial system. Using the Bayesian approach also made forensic practices visible and distributed responsibility between forensic scientists and the court. Accordingly, I will argue, this particular quantitative method was about more than mechanical objectivity (Daston, 1992, 1995; Daston and Galison, 2007; Porter, 1992a, 1995). Instead of black boxing (Latour, 1987) forensic evidence completely, making it unnecessary for its users to concern themselves with its inner workings, the Bayesian approach aimed to make the uncertainty inherent in forensic evidence visible to others.

My argument is based on ethnographic fieldwork conducted between 2008 and 2010, in which I studied the production of forensic evidence in the Swedish judicial system. This article draws on observations of forensic scientists' work in the state-run Swedish National Laboratory of Forensic Science (SKL after its Swedish name *Statens Kriminaltekniska Laboratorium*), informal and formal interviews with forensic scientists,[2] observations of criminal trials, as well as formal interviews with defense lawyers and district court judges. As SKL has adopted the Bayesian approach as recently as 2008, it was very accessible to an ethnographic study.

Comparing and evaluating

Typically, forensic analyses compare traces recovered during investigation of a crime scene with samples from a known source: bodily fluids with known DNA profiles, a mark on a forced lock with a screwdriver, a shoe print with a suspect's shoe, fibers recovered from a crime scene with a piece of clothing, and a glass fragment found in a suspect's clothes with a smashed window pane at the site of a burglary. The comparisons may help answer the following questions: did the suspect spit on the floor at the crime scene? was the screwdriver found in a suspect's purse used to pry open the lock? did the suspect's shoe leave the mark on the floor? has the suspect's shirt been in contact with the victim's sweater? and did the suspect's garment carry off the glass fragment from the crime scene?

In crime fiction, a match between the suspect or their possessions and the crime scene is often treated as indicating indubitable identification and, by extension, equally indubitable guilt (Kruse, 2010). Nonfictional criminal justice, however, must examine what a match means. Consider the glass fragment possibly carried off from a burglary: the forensic scientist might examine thickness, refraction, and chemical composition of both fragment and window pane. If they differ in only one of these characteristics, the forensic scientist can – barring errors and mix-ups – rule out the possibility that the fragment was carried off from the crime scene. Conversely, however, a match does not inevitably mean that the fragment once was a part of the window pane. Window glass is mass-produced, and thus, there are numerous window panes and perhaps other objects that might match the fragment equally well. Thus, even if the trace matches a suspected source, one cannot be certain that the trace originates from that source:

> A match isn't automatically something terrific, … but what's good and what we want to point to is that we know what it usually looks like and that the result we got is unusually similar – which we hadn't expected, it's so unusual that we hadn't expected it, not even once in a million analyses. (Forensic scientist)

What a match means depends thus on its circumstances: a forensic scientist expects one glass fragment to look much like any other, thus an unusual fragment that still matches a source might be more 'terrific' than a usual one.

In order to do justice to these complexities, SKL's forensic scientists not only produced laboratory results but also established what these results meant: they evaluated the results of their comparisons. With the exception of fingerprints – which due to membership in international organizations were reported as match, nonmatch, or inconclusive – this evaluation was done with a Bayesian approach (Cook et al., 1998b; Robertson and Vignaux, 1995; see also Lawless and Williams, 2010).[3]

According to the forensic scientists at SKL, the Bayesian approach is a way of reasoning as much as a way of calculating. Its underlying philosophy is that while it is impossible to reach certainty that a trace comes from a specific source, it is possible to make inferences from the analytical results:

> There are only two possibilities. Either it comes from that window or it doesn't … and I don't know which is – the truth. And I can never find out, either, through my analyses, the only thing I can do is to analyze these in all possible conceivable ways, and then I can give an opinion on how strongly my results support one or the other. (Forensic scientist)

In Bayesian terms, the evaluation produces a likelihood ratio: the likelihood of obtaining the analytical result (e.g. a match), given that one proposition is true (e.g. the fragment of glass comes from the window at the crime scene), compared to the likelihood of obtaining the same result, given that the assumption that the alternative proposition is true (the fragment of glass comes from some other glass object). Formally,

$$\text{Likelihood ratio} = \frac{\text{Probability of the result if proposition 1 is true}}{\text{Probability of the result if proposition 2 is true}}$$

Typically, the police or the prosecution formulate the propositions; proposition 1 typically is the prosecution's version and proposition 2 the defense's. Especially in complex cases, however, the forensic scientists sometimes give advice on which propositions would be salient to the case.

The likelihood ratio expresses how strongly the results support one proposition over the other. Applied to the example of the glass, the probability of obtaining the match if the fragment came from the window in question is divided by the probability of obtaining the match if it does not. Plainly, it is very probable that the glass matches if the fragment indeed was carried off from the crime scene, so the numerator will be close to 1. The denominator depends on how common the type of glass is. With a commonplace type of glass, a random match would be much more probable than with a rare one. Thus, the denominator will be smaller for a rare type, resulting in a large quotient – the likelihood ratio. A high likelihood ratio would indicate stronger support for the first proposition than would a low one. That is, the likelihood ratio expresses the weight of the evidence, not its accuracy: in this example, the match between trace and suspected source is assumed to have been determined without error.

The propositions can be stated at different levels, such as the so-called source, activity, and offense levels (see also Cook et al., 1998a). Propositions on the source level revolve around the question of a trace's origin – for example, whether a glass fragment did or did not come from a specific window or a bloodstain from a specific person. On the activity level, the propositions concern the activity that left the trace in question; for example, whether or not a specific suspect broke the window at the crime scene or injured the plaintiff. The offense level concerns whether the trace was left during a *criminal* activity, that is, whether the suspect is guilty of, for example, breaking and entering or of attempting homicide.

Often, the propositions SKL's forensic scientists considered involved the source level. Usually only a few types of analysis yielded results at the activity level – for example, fiber analysis – and others might reach it through the circumstances of the case or by combining several pieces of evidence. SKL's forensic scientists emphasized, however, that the offense level – whether or not a suspect had committed a crime – always was solely the court's province.

It is important to note that in the Bayesian approach as used by SKL, the forensic scientists did not give an expert opinion on the propositions. For example, they would not declare whether or not the fragment of glass came from the smashed window. They restricted their opinions to their calculations of how strongly their analytical results supported one proposition compared to the other. The difference may seem trivial, perhaps even unnecessarily complicated, but it was very important to the forensic scientists. Its consequence was that the decision on the propositions – whether or not the glass came from the window, the spit at the crime scene from the suspect, or the mark from the suspect's shoe – was left to the court. By stating the likelihood ratio, the forensic scientists delivered information which, together with the other evidence in the case, the court could use for making a decision about the facts of the case.

Uncertainty associated with the production of forensic evidence was not included in the probability figures. The forensic scientists described issues such as possible contamination or mix-ups in the laboratory in terms of eliminating risks and not of managing uncertainty. The laboratory also did not attend to the uncertainty (or perhaps ambiguity) of what a piece of forensic evidence meant for a defendant's culpability – to use Lynch et al.'s (2008: 345) terms, the field of possibility for 'how to make sense of the evidence'. This type of uncertainty pertains to what forensic scientists called the offense level – whether or not a trace was left during criminal activity – and thus was not in the laboratory's sphere. In the same vein, the forensic scientists made an emphatic distinction between results and evidence, the former being their domain and the latter the court's domain.

SKL expressed the likelihood ratio through a scale of nine grades, from −4 to +4 with a proportionate verbal scale. Each grade corresponded to an interval of likelihood ratios, with +4 expressing the strongest support (see Table 1; see also Nordgaard et al., 2012). The same grade was supposed to express the same 'value', or strength of support, regardless of the evidence for which it was used – a shoe print, a DNA profile, a glass shard, and so forth.

Table 1. SKL's scale.

Numerical scale	Verbal scale	Likelihood ratio interval
+4	The results of the examination extremely strongly support that …	$lr \geq 1{,}000{,}000$
+3	The results of the examination strongly support that …	$6000 \leq lr < 1{,}000{,}000$
+2	The results of the examination support that …	$100 \leq lr < 6000$
+1	The results of the examination support to some extent that …	$6 \leq lr < 100$
0	The results of the examination support neither … nor …	$1/6 < lr < 6$
−1	The results of the examination support to some extent that … *was not* …	$1/6 \geq lr > 1/100$
−2	The results of the examination support that … *was not* …	$1/100 \geq lr > 1/6000$
−3	The results of the examination strongly support that … *was not* …	$1/6000 \geq lr > 1/1{,}000{,}000$
−4	The results of the examination extremely strongly support that … *was not* …	$lr \leq 1/1{,}000{,}000$

SKL: Swedish National Laboratory of Forensic Science; lr: likelihood ratio.
Source: SKL, Sweden.

SKL's forensic scientists underscored that one of the benefits of using a common scale for all types of results was that it made different pieces of evidence commensurable, which in turn could make it possible to state propositions at the activity level rather than the source level. For example, examining a knife for fingerprints as well as for traces of blood could provide results that allowed the court to draw conclusions about whether a particular person used the knife to stab another person.

The Bayesian approach is not the only possible way to evaluate a match. Although the Bayesian approach is widely used by forensic science laboratories in England, Wales, Ireland, and the Netherlands, the standard practice in the United States, at least for DNA evidence, is to use random match probabilities, and there is (sometimes vehement) disagreement on how best to present probabilistic evidence (see, for example, Koehler, 1996; Lindsey et al., 2003; Schklar and Diamond, 1999).

A random match probability might very well be used in a Bayesian evaluation, but one of the differences between the two approaches is that the Bayesian approach produces a grade of support for one specific proposition compared to another, rather than an isolated, often large number – even though SKL's scale is not the only possible way to express the results of a Bayesian evaluation. Other forensic science institutions sometimes use different intervals as well as different verbal expressions and not necessarily a numerical scale (see, for example, Jackson, 2009). In addition, the Bayesian approach enables – as the next section will address – an evaluation of laboratory results even for types of evidence where there is no data or not enough data to calculate a random match probability. It also makes the results of different analysis types commensurable.

Managing epistemic limitations

It may seem that making a Bayesian evaluation is a simple matter of plugging the right numbers into the right slots. But what are the right numbers and how do forensic scientists arrive at them?

For one, there is the question of the reference 'population'. In the case of the glass fragment, for example, when calculating the probability of the result, given that the fragment did not come from the smashed window, should one consider all glass objects in the world? Only window panes? If only windows, should one consider only the ones in the neighborhood? The town? The entire country? These questions are not trivial – a type of glass, shoe sole pattern, or DNA marker may be common in one context and uncommon in another, and thus, different reference populations can lead to different likelihood ratios and thus different grades on the scale. Which reference population the forensic scientist chooses thus has significance for the evidence and the conclusions that the court can draw from it. Amâde M'charek (2000) discusses such 'practices of population' for DNA evidence in the Netherlands.

Second, the forensic scientist needs information about this reference population. For DNA evidence, SKL's forensic scientists could refer to a database of a few hundred DNA profiles, but for other types of evidence, such information is not available – it would be very hard to know exactly how many pairs of shoes with a particular sole pattern and of a particular size are in circulation at a given time. Moreover, this unknown number is subject to constant change as new models are made and sold and old pairs of shoes are thrown away. In the words of a forensic scientist:

> And that's different for different types of cases,[4] how – quickly so to speak the market changes. Flammable liquids, for example, there are new products all the time, and … we need to know at all times, what kinds of products are there on the market, what kinds of flammables are they, what do they look like. If we're talking about shoe prints, we need to have an idea about … new models, and now they've made up new sole patterns and [laughs] … it changes all the time.

Such reference material, such as old cases or samples of fibers stored in file binders or databases, was not understood and used in the same fashion as the DNA database, but as an extension of the forensic scientists' personal experience. The forensic scientists also built up their experience by making field trips to tool factories, by testing different tools and techniques for prying open doors and windows, or by conducting small-scale studies. A study of sole patterns, for example, conducted by students in a forensics course held jointly by SKL and the local university, was received with keen interest:

> And well, that's a study that, I think, that gave many an aha moment about the variation of patterns on shoe soles, and that's terrific. It – well … it gives us – an idea [laughs] about how much variation there actually is. Then again, it doesn't give us the whole truth, because it's a limited study, it was done in a few shopping malls in Linköping,[5] and it was a particular year, so you need to keep that in mind, … but it still is a tremendous help. And that's how it is with almost all databases we've got. However we build it, it has its limitations. (Forensic scientist)

These limitations have to do with both the ever-changing 'markets' for the products in question and the resources it would take to build up and maintain, for example, extensive databases.

Thus, limited knowledge makes a precise calculation of a likelihood ratio impossible in most types of forensic evidence. In addition, strictly speaking, the foundation for the likelihood ratio for DNA evidence is also a reference database, as it does not contain all of the relevant population. In addition, full profile matches were routinely assumed to pass the likelihood ratio threshold required for +4 (at the time of the fieldwork, the DNA profile system was based on 10 loci and a sex marker; since then, SKL has switched to an European standard that uses 12 loci).

The forensic scientists I studied, however, used their experience and the reference material that was available to them to *estimate* a likelihood ratio. The match between the glass fragment and the window pane at the crime scene, for example, might be graded, say, +2 on thickness, composition, and refraction alone, but if it could be pieced together with other fragments from the window pane, it might be graded higher, as that would be a very unexpected result for an unrelated fragment.

In fiber analysis, the intervals of likelihood ratios had been translated into rough guidelines. For example, a fiber specialist describes examining whether a particular shirt had been in contact[6] with a particular chair:

> Let's say we find a few fibers, four or five of them, that might tip it towards a +1. We find a lot of fibers of the same type, hundreds of them, but they're just one type from one shirt. A +2?

> Roughly speaking. For a +3 we need crosswise, I need fibers from the chair on my shirt [and] shirt fibers on the chair … then we think it's strong enough for a grade +3.[7] And +4 – well, in theory we'd be able to reach that, but … by the time we'd done all that, the trial would be over already, so we stop at grade +3 … But these different grades, … we know how much a fiber attaches itself, for how long it stays on, how common it is and how easy it is to find it by coincidence.

The negative side of the scale seemed to be regarded as more difficult than the positive one. The fiber specialist – by no means the only forensic scientist who held such a view – put it like this:

> The negative side is difficult. … To say that I'm -2 certain that they haven't been in contact, why am I -2, not -3 or -4? That's really hard. … because it's difficult to say … that it's a hundred times more likely to get these results if it isn't this way than if it is some other way … to be negatively certain, so to speak [laughs]. Well, if I find more and more, that's one thing, then I get more and more certain, but if I – I've already found nothing, how certain am I that it's nothing? I don't have fewer and fewer results, I still haven't got any fibers … If I'm to be on the negative [side], I almost need to know the details how everything happened, you know, time-wise, what were they wearing, was he wearing anything else, are there other clothes around that might look the same … Because on the positive side, I'd find more and more, [here,] I don't find less and less – it's really difficult.

Fiber analysis takes into account the premise that 'on all clothes, there is a whole carpet of fibers that you have to dismiss', the fiber specialist needs to know what that fiber carpet can be expected to look like – this 'flora' of fibers differs from household to household – and which fibers therefore may be significant for comparison. Having picked out the relevant fibers, she compares them with different microscopes under different kinds of light in order to determine which fibers, if any, are 'the same'.[8] For this, she needs to know how a fiber's appearance is affected by magnification, light, moisture, or dirt in order to determine whether similarities and differences are artifacts or inherent features.

The evaluation requires familiarity with types of fibers, how they look and how easily they spread, attach, and let go. It is thus based on what the forensic scientist can see in this particular case, what she has seen in previous cases, and what she has learned from colleagues and studies. In addition, she needs to know quite a lot about the circumstances of the case[9] – how long since the alleged contact, what happened to the shirt and the chair afterward – as well as about the fabric types involved. Thus, very few chair fibers on the shirt mean something different when the chair sheds a lot of adhesive fibers and both were collected and sealed within minutes of the alleged contact than if the chair fabric should shed few fibers or the shirt had been washed.

In tool mark analyses, comparing and grading are similarly inseparable. When comparing scratch marks, for example, it does not seem to be difficult to find similar gouges in the two samples; the difficult part seems instead to be whether these similarities are coincidental and thus to be expected or whether their numbers and placement make such a coincidence unlikely. As a forensic scientist explained, marks that match 'a little bit everywhere' do not mean very much. As a rule, she said, a +4 does not match at all except for the one (sufficiently large) area that matches perfectly – a match that becomes clearer and clearer with increased magnification.

Such analysis requires an idea of how scratch marks typically look, but it also requires the ability to see and compare individual scratches. According to my interlocutors, this ability to see took them several years to acquire under the tutelage of their colleagues – which was why they saw no point in including photographs of their findings in expert statements or in bringing them to court; 'they wouldn't understand what they're seeing anyway'. In other words, the forensic scientists had acquired what Charles Goodwin (1994) calls 'professional vision', a profession's standardized collective seeing.

Thus, the Bayesian evaluation, while producing comparable and commensurable results, could rest on a conglomerate of different practices and skills. Because limitations of knowledge in many cases made it impossible to calculate a precise likelihood ratio, the forensic scientists estimated an interval of likelihood ratios and expressed them through the graded scale.

This does not mean to say that employing a Bayesian approach and thus delivering quantified forensic evidence is somehow cheating. Rather, SKL's forensic scientists described and used the Bayesian approach as a way to combine professional experience with the statistics that were (and could reasonably be) available, balancing the impossibility to know for certain against the desire to aid criminal justice. According to my informants, some things (such as the exact number of shoes with a particular sole pattern currently in use) cannot be known, much less at reasonable cost, but they could estimate an interval of likelihood ratios and thus assign a grade on their scale.

Quantification: mechanical objectivity and more

From an STS point of view, the Bayesian approach can be understood as an instance of trust in, or rather through, numbers (cf. Porter, 1992a, 1992b, 1995). First, numbers suggest accuracy. Lynch et al. (2008), for example, note that

> the apparently precise measures of uncertainty provided by probability figures [for DNA evidence] became a source of credibility. Then, with the multiplication of markers in currently used STR systems, random match probabilities approached a vanishing point, and match declarations effectively implied certainty. (p. 345)

Second, numbers produced by quantification can be associated with mechanical objectivity, one of the different concepts of objectivity that Lorraine Daston and Peter Galison (Daston, 1992, 1995; Daston and Galison, 1992, 2007) have traced in the history of science:

> This is the form of objectivity that strives to eliminate all forms of human intervention in the observation of nature, either by using machines, such as self-inscription devices or the camera, or by mechanizing scientific procedures, as in deploying statistical techniques to choose the best of a set of observations. (Daston, 1995: 19)

Daston and Galison (1992) argue that ideals of objectivity are an expression of dangers perceived in subjectivity, of 'how, why, and when various forms of subjectivity came to be seen as *dangerously* subjective' (p. 82; emphasis in original). The dangerous subjectivity in forensic science consists of the forensic scientist's personal, individual idiosyncrasies as well as her choices and judgments (Daston and Galison, 1992: 82ff), as they might be biased or even partial.

In addition, the development of mechanical objectivity through quantification associates forensic science and its results with impartiality – the counterpart of forensic science's most dangerous subjectivity – by removing a potentially subjective and thus potentially partial decision, as pointed out by Theodore Porter:

> A decision made by the numbers (or by explicit rules of some other sort) has at least the appearance of being fair and impersonal. Scientific objectivity thus provides an answer to a moral demand for impartiality and fairness. Quantification is a way of making decisions without seeming to decide. (Porter, 1995: 9)

In other words, quantification is perceived as objective (in the sense of fair and impartial), because it produces knowledge by following rules instead of personal (subjective and thus potentially biased) judgment. Thus, quantification also establishes authority:

> 'The impersonality of numbers, I argue, is at least as crucial for their authority as is the plausibility of their claims to truth'. (Porter, 1992a: 20)

A personal judgment, potentially biased and flawed, is replaced with the impersonal result of calculating, thus devoid of subjective flaws and idiosyncrasies. Moreover, an impersonal result associates forensic science with impartiality – there is no subject involved who can take sides.[10] Porter (1995) also draws parallels to making judgments in court, emphasizing an association with morality:

> In most contexts, objectivity means fairness and impartiality. Someone who 'isn't objective' has allowed prejudice or self-interest to distort a judgment. The credibility of courts depends on an ability to elude such charges … Both these senses of objectivity imply that rules should rule, that professional as well as personal judgment should be held in check. They point to the alliance of objectivity as an ideal of knowing and objectivity as a moral value. (pp. 4–5)

In other words, a quantitative approach in forensic science may disassociate the evidence it produces from partiality, bias, and subjectivity. This can be seen in connection with criticism of forensic science as unscientific. That is, an evaluation based on an impersonal and thus impartial calculation makes it possible to address at least some of the issues identified as problematic in forensic science – for example, bias (Ghoshray, 2007) or the evidence's foundation in personal judgment and professional agreement (discussed, for example, by Cole, 2009). It does not automatically do more than conceptually disassociate forensic science from these issues, though. A British forensic science consultant interviewed and quoted by Christopher Lawless and Robin Williams (2010), for example, describes the Bayesian approach as window dressing: 'It's giving a scientific coating to what basically is a human judgment about the belief in something' (p. 748).

In Sweden, state-run SKL seems to enjoy considerable trust, perhaps due to trust in both (forensic) science and the state, but the turn toward the Bayesian approach can be seen as a move to conserve this trust. Using that approach in this context can be seen as a form of credibility work, demonstrating affiliation with scientific values and, even more importantly, loyalty to results and their evaluation and not to one of the parties in a case. As the forensic scientists emphatically do not give opinions on the propositions, and the propositions are developed by the prosecution and, on occasion, the defense,[11] they cannot be accused of compromising their impartiality.

However, I want to argue that there is more to be seen in SKL's use of the Bayesian approach than a contribution to one aspect of the credibility of its forensic science, as the use of the Bayesian approach also affected laboratory practices. For the scale to be truly uniform for all analyses, grading must be consistent throughout the laboratory.

Within the same specialization, consistency was brought about with the standardized seeing of professional vision: by sharing frames of references through discussions, forensic scientists within the same specialization developed shared practices of seeing. These could be explicit, as in the case of the rough guidelines for grading described by the fiber specialist. They could also be tacit, as encountered by a forensic scientist who had only been examining tool marks for a few years at the time about which she is speaking:

> We had a discussion about how small an area there actually can be for a +4 … I was a little skeptical, but the others who know this stuff said, it's a +4, don't you see that? Yes, I see that there's lots of detail there, but it's such a terribly tiny area if you look at it without a microscope. [And they said,] yeah, but that's enough, we've got this and this and this …

As in this account, professional vision was developed through discussions similar to 'talking science' observed in research laboratories, in which interpretations are established from more or less ambiguous data (Lynch, 1985: 155ff). When talking forensic science, forensic scientists referred to the objects under the microscope, changing magnification and angles of light, showing features to each other. Some of these features were (or perhaps could be) described in words, others with gestures.

Such discussions occurred in teaching situations, but they were also a routine part of work. Cases were always assigned to two forensic scientists who performed the same work independently and were expected to arrive at the same result or at least to be able to find and resolve the cause for their disagreement.[12] This practice was meant to catch possible errors and misjudgments as well as to prevent forensic scientists from feeling individually responsible for the result. SKL also explicitly strove for as much standardization as possible, so that it would, as forensic scientists put it, 'not matter who is assigned which case, the result should be the same', and shared analyses also advanced such standardization and developed professional vision.

As these everyday practices were not shared across specialties, uniformity between different specialties required additional effort. This wider uniformity was aimed for through so-called calibration talks during which different specialists discussed and compared their work, particularly more 'knotty' cases with low grades:

> We have these calibration talks … I for instance just had one with a girl who develops fingerprints. That's developing, and they don't grade … so I thought, what are we going to talk about, but it was really interesting … They need to know lots of things. And there we sit, and she's given me a case and I've given her a case and we sit and discuss each other's cases. It's fantastic. It's really interesting to see how other units think about the scale … and their explanations, what's behind it, it's really interesting, super good in fact. (Forensic scientist)

Fingerprints are developed – made visible by chemical means – by one set of specialists and then identified – compared with other fingerprints – by another set of specialists. The person the interviewee is referring to belongs to the first group, and the scale is not used for fingerprints. If it had been used, it would be used by the second set of specialists, so the interviewee's initial hesitation was twofold. However, according to this interviewee, calibration talks enabled him to see 'how other units think' – that is, to understand how other units translated the intervals of likelihood ratios into guidelines and criteria that were relevant for their analyses – referred to in passing as seeing 'what's behind it'. Similarly, other forensic scientists mentioned that calibration talks were very helpful for achieving both transparency and consistency when grading across specializations. Some of them warned, however, against conflating agreement and accuracy. 'We calibrate against each other', one of them said, 'so it's not so odd that we arrive at the same numbers – which does not mean that they are correct'.

These calibration talks could certainly be understood as a version of mechanical objectivity. Given that much forensic science is performed 'by hand' rather than by machines, following explicit and transparent rules would make it appear less personal and idiosyncratic. Accordingly, quantification can move toward mechanical objectivity by disciplining and standardizing judgment (Porter, 1992b: 639). Also SKL's use of the term 'calibration talks' evokes mechanical objectivity: if tasks that are imperiled by subjectivity cannot be delegated to machines, the term 'calibration' suggests at least that people can be disciplined to behave in a machinelike fashion and turn out uniform assessments. To push the metaphor, calibrating people like instruments then suggests that the results produced by these calibrated forensic scientists are as reliable and impartial as results produced by machines. Analogies between the scientist's body and mind and an instrument have been noted before (Knorr-Cetina et al., 1988: 97), and disciplining these personal and individual instruments to fit into particular scientific practices and their values may go hand in hand with establishing or staking out disciplines (Schaffer, 1988).

There was more to the calibration talks than evoking mechanical objectivity, however. The forensic scientists were expected to calibrate themselves. Moreover, they were expected to do so not only by learning and applying rules but also by widening their perspectives through discussing cases outside of the immediate scope of their expertise, such as when the fiber specialist learned about the fingerprint developer's cases and vice versa.

Second, longtime experience, and particularly experience in several specializations, was described as very valuable and as a crucial part of producing solid forensic evidence. People with such experience and skill were often mentioned in discussions and typically described as important resources, as, for example, by the forensic scientist quoted above who deferred to 'the others who know this stuff'. Similarly, leading figures in international forensic science were often referred to as 'gurus'.

These points indicate that SKL did not strive to avoid intervention, which would be impossible, considering that forensic analyses require judgments and decisions. Instead, they aimed for restrained and disciplined intervention, a nonsubjective, standardized intervention. Human subjectivity in the form of personal judgment was counteracted by combining the desired human experience and skill with machinelike self-restraint. This self-restraint was supported and coerced by practices that cultivated professional vision, such as 'calibration talks' and assigning two forensic scientists to crucial parts of each analysis. The objective was similar to that of relying on machines and mechanical objectivity, but both the means and their implications were different.

Thus, the objectivity performed in SKL's forensic practices of quantification was an objectivity produced through divided and shared decisions as well as collective practices aiming to assure uniformity and restraint. Rather than eliminating subjectivity through mechanical use of rules, as Daston and Galison (1992: 83ff) suggest, the forensic scientists strove to avoid subjectivity by achieving intersubjectivity. This was not an intersubjectivity in the Popperian sense of a different person testing or reproducing someone else's results (Popper, 1972[1959]: 44), but an intersubjectivity of shared decisions and perspectives. Inspired by Emile Durkheim's distinction of mechanical and organic solidarity in society, it might be called *organic* objectivity. Division of labor and diversified specializations, Durkheim (1984[1893]) argues, engender organic solidarity based on complementary differences rather than 'mechanical' duplication of similarities.

I certainly do not wish to invoke the cultural evolutionism inherent in Durkheim's argument; however, the division of labor along different specializations seems to be precisely what was used in order to counter the subjectivity of bias, limited perspective, and individual mistakes in forensic practices. Different experts worked together and calibrated their judgments in order to turn out expert opinions that strove both toward uniformity and standardization and toward personal experience and skill. Thus, the objectivity they performed was an objectivity of divided and shared decisions as well as of collective practices of vision and restraint.

This standardization and transparency did not extend to the forensic science community beyond SKL. Apart from talking forensic science and calibration, the forensic scientists also participated in regular accreditation activities as well as in international expert groups, which calibrated their results against the national and international standards of their respective forensic specialties. These activities did not aim at specific uniformities in evaluations, though.

Of course, the use of a Bayesian approach does not guarantee organic objectivity. It is also conceivable that transparency and organic objectivity could be arrived at with other tools than a Bayesian approach. For SKL, transparency and organic objectivity seem primarily to be a consequence of the shared scale's demand for uniformity and thus standardization, compelling forensic scientists to make their practices transparent to themselves and each other.[13] Thus, other evaluation tools that necessitate similar standardization across specializations may bring about a similar type of organic objectivity.

Distributing responsibility

So far, I have focused on the laboratory's quantification practices. The Bayesian approach, however, encompassed more members of the judicial system than only the forensic scientists. As noted above, the approach implied a particular distribution of responsibility: the forensic scientists only gave an expert opinion about the results and not about the propositions. The effect of this restriction was that the decision on the propositions was left to the court:

> It's not for us to guarantee that they belonged together, because we can't through analysis arrive at the conclusion that they've belonged together. We can only through analysis achieve a result that would be extremely odd if they hadn't. That's why we should say that the result supports extremely strongly that they've belonged together compared to if they hadn't. And the court decides about this – whether they believe this or see something else. (Forensic scientist)

In practice, it appears that this distribution of responsibility sometimes required active maintenance, perhaps also because of the recent adoption of the Bayesian approach. A forensic scientist described testifying as an expert witness as follows:

> Both the court and the prosecutor would very much like to get us to say that we think that this very strongly shows that it happened like this, but we try to guard against that … and I guess there's a risk that they don't see a huge difference there, often they can ask how likely is it, roughly, that it should have happened like this … [but] that's not for us to say.

The interviewee described a demand for something beyond the authority of a forensic scientist to deliver: the court wanted a conclusion about what happened, but to the forensic scientist 'that's not for us to say' but for the court to decide. The same forensic scientist described defense lawyers in similar terms as trying to 'devalue' the forensic scientists' statements by asking about alternative courses of action.

It is to be expected that defense lawyers would want forensic scientists to emphasize uncertainty, not to resolve it. Still, the contentious issue is the same for both prosecution and defense. The forensic scientists felt that prosecutors and defense lawyers were trying – not necessarily consciously – to elicit an expert opinion on a course of events rather than on laboratory results. Perhaps this was an expression of a desire to make the expert testimonies understandable within the framework of the case they were making for or against the defendant – for example, Wells (1992) argues that 'naked' statistical evidence is much more difficult to interpret for jurors than a person's opinion.

By being scrupulous about how they answered such questions and how they worded their written statements, the forensic scientists actively maintained the distribution of responsibility. In this way, forensic scientists took responsibility for the result and its evaluation; they did not express an opinion on what 'really' happened and subsequently were not responsible for the consequences of what the court decided.[14]

This could be understood as passing the buck to the court. For the forensic scientists, however, the transferral of the decision about the propositions to the court was not simply an evasion of responsibility for a potentially significant decision. They emphasized that even very strong forensic evidence can become substantially weakened when it is placed in the context of other evidence: 'They see more than we do, so perhaps they can see that it doesn't necessarily have to be like this, because they have more information than we do'.

That is, the court can base its decision on the evidence as a whole, whereas forensic scientists only analyze disjunct traces and often know very little about other evidence in the case. Thus, an opinion on the propositions would overstep both their authority and their competence. Therefore, the forensic scientists strive to extend their practices of organic objectivity beyond the laboratory, dividing and sharing decisions not only with one another but also with the court.

Transferring responsibility for the propositions to the court also gave the forensic scientists the latitude to present statements that carried comparatively strong uncertainty:

> I think [forensic scientists] have refrained from strong statements sometimes, because they think you have to be so very certain that it's like this … You're not saying how certain you are that it's like this, what you're saying is that your result indicates it. But if you do another analysis, it may indicate something else, that's something you have no idea about. You don't have to take responsibility for how it is. (Forensic scientist)

Thus, when the forensic scientists maintain the distribution, they not only aim for clarity about what they can and cannot say but also maintain possibilities. As long as the responsibility for deciding on a proposition lies with the court, forensic scientists are able to deliver uncertain evidence; consequently, the body of evidence in a given case can become larger as well as more nuanced. However, the distribution of responsibility also entails that it is important for forensic scientists to accurately communicate the results of their evaluations.

Communicating accurately

The distribution of responsibility also made it important for the forensic scientists to accurately communicate the results of their analyses and evaluations. The court was not the only recipient of SKL's expert statements. Forensic analyses were typically commissioned by the pretrial investigation leader – with very few exceptions, a prosecutor (I discuss their use of forensic evidence in Kruse, 2012). Defense lawyers incorporated forensic evidence into their strategies, and the police investigators who carried out the pretrial investigations used forensic evidence in their work.

However, when talking about how they convey forensic evidence, the forensic scientists predominantly spoke about the court, first and foremost the judges, more rarely the lay assessors (the Swedish judicial system does not employ juries). This focus on the court as the key recipients for the expert statements may have to do with the distribution of responsibility for presenting results and deciding about the propositions inherent in the Bayesian approach. In other words, the courts are important participants in making forensic evidence, as they provide the final closure of each piece of evidence. Thus, this section will focus on the communication from the laboratory to the court.

The forensic scientists I interviewed emphasized that it was crucial when presenting evidence in court to accurately communicate the degree of support for the propositions:

> Even if you get a very good result, you can't be super, dead certain that it has to be this way. The result indicates extremely strongly that it's like this, *and the court is going to believe you if you say so*, but it could be in a different way. (emphasis added)

The concern this forensic scientist raises is not only about overselling forensic evidence but also about taking the court's trust in SKL into consideration. Accordingly, forensic scientists consciously and deliberately pointed out to me that even a result that very strongly supported one proposition over the other one would not eliminate all uncertainty and that circumstances that they did not know about might affect the result.

In Bruno Latour's (1987) terms, the forensic scientists did not close the black box of forensic evidence – they did not make all the '[u]ncertainty, people at work, decisions, controversies' (p. 4) that went into their expert statements invisible. They also did not expect the recipients in the court to take those uncertainties into consideration unless they were mentioned. However, neither did they expect their recipients to concern themselves with *all* the intricacies of forensic science – for example, with questions about contamination, mix-up, or accuracy. Instead of trying to convey all these details, the forensic scientists strove to 'package' their results and the details they deemed salient in the most comprehensible way. For example, SKL's forensic scientists considered the Bayesian approach superior to the alternative method of calculating the statistical frequency of a DNA profile:

> [W]e stopped doing that, because … it's easy to get it wrong in court, they get carried away. If they, for example, learn that this DNA profile occurs at a frequency of one in 10 bill- no, let me exaggerate even more – 10 quintillions, there aren't even that many people on earth, then they'll think there can't be anyone else. But there can, the frequency is not the same thing as there can't be anyone else. We don't know whether there's anyone else, what we know is, if this person has a[n identical] twin, there is another one with the same [profile], no matter the low frequency. This is easily misunderstood.

They also regarded the scale as more user-friendly and comprehensible than raw likelihood ratios for conveying the weight of results, especially since they only rarely were called upon as expert witnesses in court.[15] The written expert statements thus had to contain all relevant information. They also said that the Bayesian approach enabled forensic scientists to convey the probative value of comparatively weak results:

> If it's 90 times more likely that it looks like this if it is this way than if it is that way, that's something we should tell them, it still indicates that. Not terribly strongly, but it's not like it's a bad likelihood. (Forensic scientist)

The forensic scientists I interviewed often emphasized this point when discussing possible future developments for fingerprint evidence. A Bayesian approach, many forensic scientists argued, would make it possible to communicate fingerprint comparison results in a more nuanced fashion, and it also would allow forensic scientists to give the court, as well as the police investigators, potentially valuable information on partial fingerprints.

In the distribution of responsibility and the communication of results, the recipients of SKL's expert statements were thus expected to concern themselves with some, but not all, sources of uncertainty and decision making; in other words, the black box was neither fully opened nor closed. Forensic evidence is not the only 'box' that is neither fully transparent nor opaque; Kathleen Jordan and Michael Lynch (1992: 107), for example, use the term 'translucent box' for processes and technologies whose 'outlines … are not clearly resolved' and whose 'inner workings remain clouded by uncertainty and dispute'. However, unlike the molecular biology technique that they discuss, a forensic expert statement is a stable and standardized (and rarely disputed) artifact. Thus, to avoid confusion, I want to call the forensic evidence produced at SKL a semitransparent box – neither fully opaque nor fully transparent.

The uncertainties and decisions left visible by this semitransparency were not necessarily a focus of interest in court, however:

> The fact that your DNA has been recovered from a particular place … normally, you accept the fact that the point has been proven, so there won't be any discussion there, but, 'yes, sure, I was there'; or the question becomes instead, can the DNA – or the object from which the DNA has been recovered – have ended up at the place by mistake, even though the person wasn't there.

In this fairly typical description of a hypothetical scenario, this district court judge collapses the (for the forensic scientists) complicated issue of what a DNA match means into 'your DNA has been recovered', immediately moving on to the (for the judge) more interesting issue of what the presence of 'your DNA' at the crime scene means, thus simply making the decision on the propositions, namely, that the match between the DNA profiles from the crime scene and the suspect means that they both come from the suspect.

The judge's stance was not unique. The central question also for prosecutors and defense lawyers was how to interpret the forensic evidence delivered by the laboratory in terms of the defendant's actions and culpability. They focused primarily on what forensic scientists would have called the offense level and what Lynch et al. (2008: 345) call the field of possibilities.

Uncertainty can become a focus of attention in court, however. Later on in the interview, the same judge discussed differences in leeway in different grades:

> Well, if SKL say it's a +4, … that's good enough, so to speak, and what they're saying is that it's essentially out of the question that there is a different explanation. If it's a +2, they're saying that there is – an opening … for finding a different explanation, and then it's more a question of, then you can take into account the value of the evidence that supports the prosecutor's version in the form of a +2, and what other evidence there is, and the circumstances that support the opposite.

Such 'weaker' evidence, for example, graded +1 or +2, was seldom brought into the courtroom in the first place. On the occasions when it was, forensic scientists sometimes were summoned as expert witnesses and asked to explain how the analysis in question had been performed and how they had arrived at a particular grade. Both in the hearings I attended and in forensic scientists' descriptions, questions in such situations revolved around what was wrong with a result that had been graded lower than +4:

> They often don't understand the cause of the uncertainty … if it's a grade +2 or +3 they think it's a bad result, you know? Something's gone wrong with the machine or something, that's why it turned out a bit second-rate. But it's not like the result is wrong or something … it can be a terrific and neat result all the same. There may be a v-shaped cut in a plastic bag … but if you cut another bag, chances are that it looks about the same. But of course, if there's details, that changes … but … the fewer details, the – there's nothing wrong with the details, but there's too little information … When it comes to DNA, for example, they think there's something wrong, they think there's a deviation in the profile [i.e. a partial mismatch], that's why it's only a +2. So there they want to pressure us … what's wrong, where's the deviation, why is it only a +2? (Forensic scientist)

In other words, while the forensic scientists talked and wrote about how strongly the results support one of the propositions over the other – in other words, the *weight* of the results – judges, prosecutors, and defense lawyers on occasion understand the grade as being about the results themselves – their *quality* – for example, an imperfect match.

This also suggests that at least some recipients of SKL's expert statements expect a match to result in grade +4, as the expected explanation for a lower grade is a partial mismatch or some other fault. The forensic scientists appeared to encounter this attitude regularly:

> [W]hen we tell [the police and the prosecutors] that for glass cases, for example, we seldom get a higher grade than +2, they say, well, then, we'd like you to put that into your statement, say that it doesn't get any higher than +2. All right, we say, why do you want to know that, or why should we write that, won't you take that as, well, that's the same as certainty? Yes, they admit, that's how they'll interpret it in any case, this is as good as it gets, so it's certain. And that's not what we're saying, +2 is only +2.

Relatedly, the same forensic scientist regarded as quite problematic the practice of introducing a higher grade in order to be able to express stronger support: 'We see a danger in introducing another grade, because then they would automatically think that the other ones are less valuable'. The attitude the forensic scientists described indicates that the recipients of expert statements sometimes treated the scale as relative rather than absolute. They seemed to measure grades against what they believed should be possible, instead of focusing on – as the laboratory did – what the results supported. Consequently, for them to ask what was wrong with a result with a low grade seemed only logical.

The type of uncertainty that was not addressed in this context was uncertainty about the laboratory's practices; this seemed to be a matter of the court participants' trust in SKL. At least for defense lawyers, the practical difficulties involved with obtaining a second opinion may have been a consideration, as well, as SKL is the only forensic science laboratory in the country. While it is possible to commission a second opinion from abroad, tight time frames, along with language barriers that are difficult to breach even with highly qualified interpreters, make it difficult to do so in practice.[16] Johanne Yttri Dahl (2009) describes a similar situation for Norway at the time of her research and discusses Norwegian defense lawyers' perspectives on the situation.

The recipients of expert statements thus did not always look into the semitransparent box of forensic evidence. With 'strong' evidence, they did not seem to see any need to look into the background details, but with 'weak' evidence, they did not always see what the forensic scientists expected them to see. Trust, practical hurdles, and differences in focus or expertise could make the black box less transparent than intended and cause friction between producers and users. This friction could arise regarding the amount of uncertainty, the implications of what the laboratory result meant, and the source of the uncertainty.

Similar friction can be found in the British *R v T* judgment[17] and the forensic science community's reaction to it. The case was an appeal of a conviction for murder. In its judgment, the Court of Appeal focused on forensic comparisons of shoe marks at the crime scene with a pair of shoes seized from the defendant. The marks and the shoes matched in sole size and pattern, but the shoes – which had been seized well after the crime occurred – showed more wear than indicated by the marks. In addition, according to the forensic scientist who had analyzed the evidence, there were discrepancies that may or may not have been caused by features of the soles having been worn away or by small stones or other artifacts. The forensic scientist had concluded that there was 'a moderate degree of scientific support for the view' that the defendant's shoes had made the marks – moderate support being equivalent to a likelihood ratio of between 10 and 100 on the British Forensic Science Service verbal scale, which differs slightly from SKL's.

The court took issue with the forensic scientist's use of a 'mathematical formula' to characterize uncertain evidence. That is, the court objected that he had developed a likelihood ratio, which gave the court the impression of a precise result, when the result was based on estimates of, among other things, how common a specific sole pattern was and not on precise statistical data. The court deemed the use of a mathematical formula for shoe print evidence both nontransparent and unsuitable due to insufficient data, contrasting it to DNA evidence where, in their opinion, the statistical basis was sufficient.

The forensic science community responded that the court had misunderstood the Bayesian approach (Aitkin et al., 2011; Berger et al., 2011; Robertson et al., 2011). They argued that the Bayesian approach was precisely about 'reasoning in the face of uncertainty' (Aitkin et al., 2011: 1) and that the court had confused the uncertainty arising from limited knowledge about shoe sole patterns with uncertainty about the forensic scientist's conclusion (Robertson et al., 2011).

The judgment and the reactions to it mirror some of the friction in the Swedish judicial system: like SKL's forensic scientists, the forensic scientist in the *R v T* case used the Bayesian approach as a tool for managing limited knowledge and for communicating forensic results and their weight, and the British court did not evaluate the evidence in the intended way. Forensic scientists as well as management at SKL viewed the judgment as an illustration of how judges (as well as prosecutors and defense lawyers) sometimes misunderstand the Bayesian approach, and dangerously so, and they wondered whether it would cause difficulties in the long run for using that approach, both by SKL and internationally. This indicates that it can be difficult in practice to extend the laboratory, so to speak, to involve the court in the use of the Bayesian approach for evaluating the results of forensic analyses.

Conclusion: on keeping the black box semitransparent

I have discussed the Bayesian approach used by SKL to address a particular type of uncertainty in the judicial system, namely, the uncertainty inherent in forensic evidence due to limitations in knowledge. The forensic scientists I studied understood the approach primarily as a way of thinking when evaluating the weight of a match between a trace and a suspected source. Through its uniform scale, one of the Bayesian approach's effects on the laboratory as a whole was an organic objectivity based on division of labor and shared understandings and practices, resulting in increased transparency for laboratory members, though not necessarily for others.

The approach also entailed distributed responsibility between the laboratory and the court. In addition, as the court made the final decision on what the laboratory results meant, the forensic scientists' effort to communicate the results' value (or weight) to the court was crucial. This communication did not encompass all the aspects of the laboratory's 'internal' transparency. Expert statements were not meant to turn their users into forensic scientists, only to communicate laboratory results accurately, so that the court could understand the implications of those results and decide what they meant. The prosecutors, defense lawyers, judges, and lay assessors were expected not to 'get carried away', and to know that a low grade did not mean that '[s]omething's gone wrong with the machine'. The forensic scientists expected them to understand which alternatives had been taken into consideration in the laboratory's evaluation. They were, however, not expected to concern themselves with the uncertainty associated with the production of forensic evidence or all the details of analysis and evaluation. While some of these details could be exposed during informal inquiries or formal court hearings, others were regarded as falling exclusively in the domain and responsibility of forensic scientists. In other words, the forensic scientists kept the results they delivered semitransparent instead of black boxing them completely.

This use of the Bayesian approach straddled Latour's (1987) distinction between ready-made science and science in action: while not requiring expertise in forensic science, the Bayesian approach required court participants to engage in active and informed cooperation with the forensic scientists in order for the evidence to travel from the laboratory to the legal arena of the court without losing the accuracy the forensic scientists emphasized.

The occasional friction between expert witnesses and, mainly, prosecutors and defense lawyers described by the forensic scientists at SKL, or exhibited by the controversy surrounding the *R v T* judgment, suggests, however, that it may be difficult to keep the black box of forensic evidence semitransparent when producers and recipients operate with very different expectations and forms of expertise. For someone unfamiliar with the inner workings of forensic science, the numbers generated with the Bayesian approach may not be sufficient to communicate the extent of expertise and professional vision that are a vital part of producing forensic evidence. They also may be insufficient to communicate limitations in the knowledge on which evaluations are based.

At SKL, these aspects of forensic science formed part of the shared understandings that forensic scientists maintained through shared practices. Forensic scientists and other professions in the judicial system, however, did not have many points of contact where they could develop shared understandings. Expert statements – which might form such a connection between different professions – often traveled through the judicial system without further communication between the laboratory and the recipients. When further communication between them did take place, through informal conversations or expert testimony in court, it was typically instigated by prosecutors or defense lawyers, not by judges and lay assessors, who are expected to hear the evidence presented in court, not to conduct inquiries of their own. What is more, further communication only took place when the prosecutors or defense lawyers had additional questions or anticipated contention. In addition, Sweden's judicial system presently is not so fiercely adversarial that the defense would routinely question SKL's impartiality or reliability and thus reopen that part of the (semitransparent) black box of forensic analyses.

This made it difficult to maintain semitransparency when evidence moved from the laboratory to the court. It also made it difficult to extend organic objectivity to encompass the participants in trial. A further contributing factor may be the forensic scientists' reluctance to teach (potential) criminals how to avoid conviction: court proceedings are public documents, and therefore, the forensic scientists considered very carefully how much detail they could put into their expert statements without teaching criminals how to avoid leaving usable traces.

In other words, outside the laboratory, where forensic scientists and the recipients of their expert statements had not developed a shared framework, forensic practices were less transparent than their producers had intended. It appears that although it was important for forensic scientists to keep the black box semitransparent, such a degree of transparency takes work to sustain.

Acknowledgements

I would like to thank my collaborators within the Swedish judicial system who graciously opened their world to me and generously made time for me in their strained schedules. I am grateful to the *Technology, Practice, Identity* group for dissecting an early draft of this article and to Simon Cole, Claes-Fredrik Helgesson, Sergio Sismondo, Ebba Sjögren, Victor Toom, Robin Williams, and Steve Woolgar for helpful comments. Special thanks to Magnus Blondin for suggesting the term 'organic objectivity'. I am also indebted to Michael Lynch and the anonymous reviewers for their careful and insightful comments and questions that have improved this article.

Funding

This research was supported by a grant from the Swedish Research Council.

Notes

1. The Bayesian approach is not synonymous with Bayes' theorem (see Aitkin et al., 2011).

2. In this article, ‘forensic scientist’ refers very broadly to someone working in a forensic science laboratory. Swedish National Laboratory of Forensic Science’s (SKL) forensic scientists had varying backgrounds, many in life sciences, but none were police officers.
3. I use the term ‘Bayesian approach’ to describe the conceptual framework as a whole, including the courts’ role; SKL would call their part a likelihood ratio approach.
4. In this context, a ‘case’ is not necessarily a court case; in the laboratory, a ‘case’ can also be a single trace.
5. This is the city where SKL is located.
6. This analysis involves the question of whether the items have been in contact with each other, that is, the activity level. While the reasoning is the same as for the source level, the activity level requires more information about a case and its circumstances.
7. Alternatively, a grade of +3 could be given when finding fibers from three separate items worn together on a chair made of a fabric that does not release fibers.
8. Both Martin and Lynch (2009) and Cole (2009) discuss the knotty issue of sameness.
9. This applies to fiber analyses. In other types of analyses, forensic scientists do not require as much information and may prefer to remain ignorant about the case.
10. The suggestion of addressing the question of guilt in court mathematically (see, for example, Fienberg and Kadane, 1983) might also be seen as striving for impartial objectivity.
11. In the cases where the defense does not disclose the defendant’s version, or parts of it, until the court hearing, the defendant’s version cannot of course be taken into account in the propositions.
12. If that is impossible, a third person is consulted – apparently a very rare occurrence.
13. Lawless and Williams (2010) connect the Bayesian approach to economic values, showing how the transparency of following explicit rules enables the British Forensic Science Service to assess the expected value of an investigation and thus provide a basis for the customers’ decisions on whether to commission an investigation. As SKL’s ‘customers’ did not pay directly for their services, this aspect of the Bayesian approach’s transparency was not an issue there.
14. This distribution of responsibility echoes the displacement of uncertainties and ambiguities associated with technologies to users, which Brian Rappert (2001) discusses.
15. This may have to do with Sweden not employing a jury system. Judges and lay assessors can draw on experience with forensic evidence and may not need as much explanation of routine evidence as might be the case for a lay jury.
16. However, there was a high-profile trial in 2012 in which the defense brought in an expert witness from the United States in order to challenge SKL’s analysis.
17. *R v T* (2010) EWCA Crim 2439.

References

Aitkin C, Berger CEH, Buckleton JS, Chapod C, Curran J, Dawid AP, et al. (2011) Expressing evaluative opinions: A position statement. *Science & Justice* 51(1): 1–2.

Berger CEH, Buckleton J, Champod C, Evett IW and Jackson G (2011) Evidence evaluation: A response to the court of appeal judgment in *R v T. Science & Justice* 51(2): 43–49.

Cole SA (2001) *Suspect Identities: A History of Fingerprinting and Criminal Identification*. Cambridge, MA: Harvard University Press.

Cole SA (2009) Forensics without uniqueness, conclusions without individualization: The new epistemology of forensic identification. *Law, Probability and Risk* 8(3): 233–255.

Cook R, Evett IW, Jackson G, Jones PJ and Lambert JA (1998a) A hierarchy of propositions: Deciding which level to address in casework. *Science & Justice* 38(4): 231–239.

Cook R, Evett IW, Jackson G, Jones PJ and Lambert JA (1998b) A model for case assessment and interpretation. *Science & Justice* 38(3): 151–156.

Daemmrich A (1998) The evidence does not speak for itself: Expert witnesses and the organization of DNA-typing companies. *Social Studies of Science* 28(5/6): 741–772.

Dahl JY (2009) Another side of the story: Defence lawyers' views on DNA evidence. In: Aas KF, Gundhus HO and Lomell HM (eds) *Technologies of In Security: The Surveillance of Everyday Life*. Abingdon: Routledge-Cavendish, pp. 219–237.

Dahl JY and Sætnan AR (2009) 'It all happened so slowly': On controlling function creep in forensic DNA databases. *International Journal of Law, Crime and Justice* 37(3): 83–103.

Daston L (1992) Objectivity and the escape from perspective. *Social Studies of In Science* 22(4): 597–618.

Daston L (1995) The moral economy of science. *Osiris* 10: 2–24.

Daston L and Galison P (1992) The image of objectivity. *Representations* 40: 81–128.

Daston L and Galison P (2007) *Objectivity*. New York: Zone Books.

Durkheim E (1984[1893]) *The Division of Labour in Society*. Basingstoke: Macmillan Education.

Fienberg SE and Kadane JB (1983) The presentation of Bayesian statistical analyses in legal proceedings. *Journal of the Royal Statistical Society, Series D (The Statistician)* 32(1/2): 88–98.

Ghoshray S (2007) Untangling the CSI effect in criminal jurisprudence: Circumstantial evidence, reasonable doubt, and jury manipulation. *New England Law Review* 41(3): 533–562.

Goodwin C (1994) Professional vision. *American Anthropologist* 96(3): 606–633.

Halfon S (1998) Collecting, testing and convincing: Forensic DNA experts in the courts. *Social Studies of Science* 28(5/6): 801–828.

Hindmarsh R and Prainsack B (eds) (2010) *Genetic Suspects: Global Governance of Forensic DNA Profiling and Databasing*. Cambridge: Cambridge University Press.

Jackson G (2009) Understanding forensic science opinions. In: Fraser J and Williams R (eds) *Handbook of Forensic Science*. Uffculme: Willan Publishing, pp. 419–444.

Jasanoff S (2001) Judicial fictions: The Supreme Court's quest for good science. *Society* 38(4): 27–36.

Jasanoff S (2006) Just evidence: The limits of science in the legal process. *Journal of Law, Medicine & Ethics* 34(2): 328–341.

Jordan K and Lynch M (1992) The sociology of a genetic engineering technique: Ritual and rationality in the performance of the 'plasmid prep'. In: Clarke AE and Fujimura JH (eds) *The Right Tools for the Job: At Work in Twentieth-Century Life Sciences*. Princeton, NJ: Princeton University Press, pp. 77–114.

Kaye DH and Koehler JJ (1991) Can jurors understand probabilistic evidence? *Journal of the Royal Statistical Society, Series A (Statistics in Society)* 154(1): 75–81.

Knorr-Cetina K, Amann K, Hirschauer S and Schmidt KH (1988) Das naturwissenschaftliche Labor als Ort der 'Verdichtung' von Gesellschaft. *Zeitschrift für Soziologie* 17(2): 85–101.

Koehler JJ (1996) On conveying the probative value of DNA evidence: Frequencies, likelihood ratios, and error rates. *University of Colorado Law Review* 67: 859–886.

Kruse C (2010) Producing absolute truth: *CSI* science as wishful thinking. *American Anthropologist* 112(1): 79–91.

Kruse C (2012) Legal storytelling in pre-trial investigation: Arguing for a wider perspective on forensic evidence. *New Genetics and Society* 31(3): 299–309.

Latour B (1987) *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.

Lawless CJ and Williams R (2010) Helping with inquiries or helping with profits? The trials and tribulations of a technology of forensic reasoning. *Social Studies of Science* 40(5): 731–755.

Lindsey S, Hertwig R and Gigerenzer G (2003) Communicating statistical DNA evidence. *Jurimetrics* 43: 147–163.
Lynch M (1985) *Art and Artifact in Laboratory Science: A Study of Shop Work and Shop Talk in a Research Laboratory*. London: Routledge & Kegan Paul.
Lynch M, Cole SA, McNally R and Jordan K (2008) *Truth Machine: The Contentious History of DNA Fingerprinting*. Chicago, IL: University of Chicago Press.
McCartney C (2006) *Forensic Identification and Criminal Justice – Forensic Science, Justice and Risk*. Portland, OR: Willan Publishing.
Martin A and Lynch M (2009) Counting things and people: The practices and politics of counting. *Social Problems* 56(2): 243–266.
M'charek A (2000) Technologies of population: Forensic DNA testing practices and the making of differences and similarities. *Configurations* 8: 121–158.
M'charek A (2008) Silent witness, articulate collective: DNA evidence and the inference of visible traits. *Bioethics* 22(9): 519–528.
Nordgaard A, Ansell R, Drotz W and Jaeger L (2012) Scale of conclusions for the value of evidence. *Law, Probability and Risk* 11(1): 1–24.
Popper KR (1972[1959]) *The Logic of Scientific Discovery*. London: Hutchinson & Co Ltd.
Porter TM (1992a) Objectivity as standardization: The rhetoric of impersonality in measurement, statistics, and cost-benefit analysis. *Annals of Scholarship* 9: 19–61.
Porter TM (1992b) Quantification and the accounting ideal in science. *Social Studies of Science* 22(4): 633–651.
Porter TM (1995) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.
Prainsack B and Toom V (2010) The Prüm regime: Situated dis/empowerment in transnational DNA profile exchange. *British Journal of Criminology* 50(6): 1117–1135.
Rappert B (2001) The distribution and resolution of the ambiguities of technology, or why Bobby can't spray. *Social Studies of Science* 31(4): 557–591.
Robertson B and Vignaux GA (1995) *Interpreting Evidence – Evaluating Forensic Science in the Courtroom*. Chichester: John Wiley & Sons.
Robertson B, Vignaux GA and Berger CEH (2011) Extending the confusion about Bayes. *The Modern Law Review* 74(3): 444–455.
Schaffer S (1988) Astronomers mark time: Discipline and the personal equation. *Science in Context* 2(1): 115–145.
Schklar J and Diamond SS (1999) Juror reactions to DNA evidence: Errors and expectancies. *Law and Human Behavior* 23(2): 159–184.
Wells GL (1992) Naked statistical evidence of liability: Is subjective probability enough? *Journal of Personality and Social Psychology* 62(5): 739–752.

Author biography

Corinna Kruse is Assistant Professor at the Department of Thematic Studies – Technology and Social Change at Linköping University, Sweden. Her current research focuses on how forensic evidence is produced and transported across the Swedish judicial system's professions.