Synthese (2023) 201:134
https://doi.org/10.1007/s11229-023-04141-2

ORIGINAL RESEARCH

Bare statistical evidence and the legitimacy of software-based judicial decisions

Eva Schmidt[1] · **Andreas Sesing-Wagenpfeil**[2] · **Maximilian A. Köhl**[2]

Received: 26 September 2021 / Accepted: 23 March 2023 / Published online: 4 April 2023

Abstract
Can the evidence provided by software systems meet the standard of proof for civil or criminal cases, and is it individualized evidence? Or, to the contrary, do software systems exclusively provide bare statistical evidence? In this paper, we argue that there are cases in which evidence in the form of probabilities computed by software systems is not bare statistical evidence, and is thus able to meet the standard of proof. First, based on the case of State v. Loomis, we investigate recidivism predictions provided by software systems used in the courtroom. Here, we raise problems for software systems that provide predictions that are based on bare statistical evidence. Second, by examining the case of People v. Chubbs, we argue that the statistical evidence provided by software systems in cold hit DNA cases may in some cases suffice for individualized evidence, on a view on which individualized evidence is evidence that normically supports the relevant proposition (Smith, in Mind 127:1193–1218, 2018).

Keywords Bare statistical evidence · Normic support · COMPAS · Legal epistemology · Standard of proof · DNA evidence · Inference to the best explanation

✉ Eva Schmidt
eva.schmidt@tu-dortmund.de

Andreas Sesing-Wagenpfeil
andreas.sesing@uni-saarland.de

Maximilian A. Köhl
mkoehl@cs.uni-saarland.de

[1] Department of Philosophy and Political Science, TU Dortmund, Emil-Figge-Straße 50, 44227 Dortmund, Germany

[2] Saarland University, Saarbrücken, Germany

1 Introduction

Courts are obliged to apply the relevant legal rules to given facts. The central task of judges is therefore the application of the law. This is preceded, however, by the important task of establishing the actual facts of the case—according to certain rules—and further, making well-founded predictions where necessary. Courts are therefore traditionally supported by experts who have the necessary expertise to clarify the facts of the case. Current examples show that these experts are increasingly supplemented or even completely replaced by software systems. This can be illustrated by two recent legal cases from the US:

(1) In State v. Loomis (USA, 2016), Eric Loomis was sentenced to six years in prison. The decision was based on a high recidivism probability that was calculated by the software COMPAS. COMPAS is designed to support a judge's sentencing decisions (cf. State v. Loomis, 881 N.W.2d 749 (2016)).
(2) In another 2016 US case, Martell Chubbs was convicted for murder on the basis of the very high probability that he was the source of a pertinent DNA trace left at the crime scene, as calculated by the software TrueAllele. The software is designed to support forensic experts by providing such probabilities (cf. People v. Chubbs CA2/4, B258569 (Cal. Ct. App. 2015); Smith et al., 2018).

Both court decisions were based on probabilities computed by software systems. This raises the suspicion that they were founded on bare statistical evidence. Legal epistemologists claim that bare statistical evidence is not the right kind of evidence to convict a defendant of a crime; it cannot meet the standard of proof. It lacks an appropriate connection to the defendant's committing the crime and is thus not individualized (Thomson 1986).[1] For instance, we cannot convict a prisoner for participating in an attack on a prison guard, given that our only evidence is that she is a member of a group of 100 prisoners, out of which only 99 took part in the attack. Although the issue is rarely discussed, a similar worry may be raised for sentencing decisions on the basis of bare statistical evidence. In particular, it may well be that a more severe sentencing decision cannot be based on bare statistical evidence that the convict has a high recidivism risk (Jorgensen, 2021).[2]

Against this backdrop, it is of interest—both from an epistemological and from a legal perspective—whether the evidence provided by software systems can meet the standard of proof for civil or criminal cases, or in other words, whether such evidence ever suffices for conviction or for imposing a severe sentence. Otherwise, the use of such systems for judicial decisions might be severely limited. In this paper, we argue that there are cases in which evidence in the form of probabilities computed by software systems is not bare statistical evidence, and is thus able to meet the standard

[1] We use 'individualized evidence' as equivalent with *'not* bare statistical evidence' for the purposes of this paper. The term 'individualized' as used in the literature reflects the idea that evidence, to suffice for convicting an individual, has to be concerned with her directly, rather than—for instance—some statistical generalization under which the individual happens to fall. Cf. Gardiner (2019a) and Thomson (1986). The relevance of the connection to the individual will emerge more clearly when we discuss several competing accounts of individualized evidence.

[2] We will not be able to discuss an analogous worry regarding pretrial decisions, e.g. whether to issue a search warrant (Bambauer 2015).

of proof for conviction. Let us add right away that this is not to deny that the use of such systems in the courtroom may very well be problematic for other reasons—indeed, we intend to discuss further problems in future work.

But there is not only a lesson to be learned for the use of software systems in the courtroom; vice versa, the cases we explore allow us to gain a better understanding of the debate over bare statistical evidence. On the one hand, the case of Eric Loomis can be used to show that issues over bare statistical evidence extend from the conviction stage to the sentencing stage. On the other hand, we can illustrate with People v. Chubbs how the normic account of individualized evidence (Smith, 2018) can handle convictions in cold hit DNA cases. Note that both of these are US cases. Since we are familiar with the German system, which interestingly subscribes to the same assumption, we will draw on the German and Anglo-Saxon traditions for our discussion.

We begin by introducing the distinction between individualized and bare statistical evidence in the law and in legal epistemology, as well as the functioning of typical software systems used in the legal system. We then turn to our first case (recidivism risk assessment software), which we introduce with the help of State v. Loomis. We argue that even the use of software systems to provide evidence for sentencing rather than conviction is problematic, given that it is not individualized. We then turn to the second case (DNA comparison software), exemplified by People v. Chubbs. We argue that, despite appearances, the statistical evidence provided by software systems in cold hit cases may suffice for individualized evidence, on a view on which individualized evidence is evidence that normically supports the relevant proposition.

2 Background

2.1 Standards of proof and individualized vs. bare statistical evidence

A legal judgment finding a defendant guilty or liable needs to be based on sufficient evidence. This requirement is expressed by the term 'standard of proof', which describes the degree of evidence which is necessary to establish proof in a criminal or civil proceeding.

Under the German Code of Civil Procedure ('Zivilprozessordnung', 'ZPO'), the court is to decide, at its discretion and conviction, and taking account of the entire content of the hearings and the results obtained by evidence being taken, if any, whether an allegation as to fact is to be deemed true or untrue (sec. 286 para. 1 ZPO). The conviction of the truth of a fact does not presuppose an absolute or irrevocable certainty, since such a certainty cannot be achieved.[3] It is sufficient to have a degree of certainty which is useful for practical life and which excludes reasonable doubts.[4] The same standard of proof applies to criminal procedural law under the German

[3] *Miebach* in Münchener Kommentar StPO (2016), sec. 261 note 52.

[4] *Bundesgerichtshof* (Federal Court of Justice), NJW-RR 1994, 567, 568; *Greger* in Zöller, Zivilprozessordnung, 31st Ed. (2016), sec. 286 note 17 et seq.

Code of Criminal Procedure ('Strafprozessordnung', 'StPO').[5] Since under German laws, conviction and sentencing are two elements of a unitary criminal proceeding, this standard of proof applies to both conviction and sentencing decisions in criminal proceedings.

This standard of proof is not met merely if there is a high probability that a fact is true. This is justified by the fact that evidence based solely on a certain probability is not sufficient for the formation of a judge's conviction.[6] Such evidence is not strong enough to exclude reasonable doubt about the facts of a situation. The issue is also reflected in the legal debate on the question whether judges must be convinced of either alleged facts or merely of the (very high) probability of the existence of these facts.[7] Assuming that the latter is not sufficient, bare statistical evidence—even if it comes with high probabilities—cannot meet the standard of proof under German procedural laws.

The very same point has been extensively discussed in a recent debate in (English language) legal epistemology, although that debate is based on the American and British legal systems.

Let us focus our discussion with the help of two examples, both taken from Gardiner (2019a).

> Red Taxi. A vehicle hit Jeanette late one night. She could determine it was a taxi, but could not discern the color. The Red Taxi company operates 75% of taxis in town. The remaining 25% are operated by the Green Taxi company. Jeanette sues the Red Taxi company.

Prison Yard. One hundred prisoners exercise in the prison yard. Ninety-nine prisoners together initiate a premeditated attack on a guard. Security footage reveals one prisoner standing against the wall refusing to participate. There is no evidence indicating who refused to participate.

Intuitively—and also based on how courts (in the US at least) actually rule in such cases—the evidence in both cases does not meet the relevant standard of proof despite the fact that, given a probabilistic understanding of the American standards of proof, the relevant probabilities are high enough to do so.[8] Assume that both cases take place in the US, so that US civil and criminal law are relevant. In Jeanette's case, based on the evidence, the probability that it was a Red taxi that hit her is 75%. This should suffice for her to win a civil lawsuit finding Red Taxi liable for the accident, since in US civil law the standard of proof is 'preponderance of evidence', i.e., in probabilistic terms, greater than 50%. In Prison Yard, the probability that a certain prisoner P was involved in the attack is 99%; this should be sufficient to convict P, as this exceeds the

[5] *Bundesgerichtshof* (Federal Court of Justice), NJW 1980, 2423; *Miebach* in Münchener Kommentar StPO (2016), sec. 261 note 57.

[6] *Ott* in Karlsruher Kommentar zur StPO, 7th Ed. (2013), Sect. 261 note 4.

[7] See, for instance, *Wagner* in Münchener Kommentar ZPO, 5th Ed. 2016, § 286 notes 32 et seq.

[8] Note that a probabilistic understanding of standards of proof is problematic in part because of issues with bare statistical evidence as discussed here. For an overview over such a probabilistic understanding, some of its problems, and relevant references, see Urbaniak & Di Bello (2021), Sect. 5.

'beyond reasonable doubt' standard of proof for criminal cases of around 90–95%, on a probabilistic interpretation.[9]

It would be unjust to find Red Taxi liable for the accident if the only evidence in the case concerns the market share of the taxi companies in town, and equally unjust to randomly pick out *P* and convict her of a crime, given that our only evidence of her involvement is that 99 out of 100 prisoners participated. In other words, bare statistical evidence is an insufficient basis for a judicial decision against a defendant. This is borne out by actual legal cases such as Smith v. Rapid Transit, which inspired Red Taxi.[10] Similar results follow from the 'in dubio pro reo' principle: In the absence of further circumstantial evidence, bare statistical evidence usually leaves room for well-founded doubts as to the guilt of the defendant.

This raises the puzzling question: What is it about bare statistical evidence like this that prevents it from meeting the standard of proof even where its probabilistic strength meets or exceeds the standard? Further, under which conditions does evidence suffice for conviction? Without wanting to incur any metaphysical commitments, we here put the relevant question as: When is evidence 'individualized', so that it is—on the relevant dimension—sufficient for conviction? And when is it 'bare statistical evidence' and thus insufficient on this dimension?[11]

Two answers to these questions are backed by standard positions in epistemology, viz. the causal theory of knowledge (cf. Goldman, 1967) and sensitivity theories of knowledge (cf. Nozick, 1981).[12] According to Thomson's (1986) causal account, evidence counts as individualized iff it "is in an appropriate way causally connected with the (putative) fact that the defendant caused the harm." (Thomson, 1986, p. 203) Her diagnosis of Red Taxi is that, since there is no causal connection between the evidence (the fact that 75% of the taxis in town are owned by Red Taxi) and the alleged fact that a Red taxi hit Jeanette on the night in question, we are dealing with bare statistical evidence. Contrast this with a case in which a reliable witness *testifies* to having seen a Red taxi hit Jeanette. His testimony is individualized evidence—his seeing the accident is (causally) due to the fact that a Red taxi hit Jeanette, as is his

[9] Our characterizations of the US standards of proof are from Gardiner (2019a).

[10] Cf. 317 Mass. 469, 58 N.E.2d 754 (1945).

[11] Thanks to Marvin Backes for discussion on this point. Note that the 'relevant dimension' still needs to be clarified. At any rate, it is distinct from the probabilistic dimension. For on this dimension, bare statistical evidence may well be sufficient.

[12] The aim of this paper is not to compare, in depth, different accounts of individualized evidence, but to reflect on whether software systems provide only bare statistical evidence and are thus of limited use in the courtroom. For this purpose, we rely on the normic account as one recent and very plausible account. As we argue below, this account gets the individuation of evidence right across good and bad cases, unlike at least some of its competitors; it is further of special interest to us because of its issues with classifying DNA evidence, as discussed in Sect. 4 of this paper.

To provide readers with some relevant context, we will contrast the normic account with two prominent and classic alternatives, the causal and the sensitivity account. Due to lack of space, we cannot discuss the plethora of other accounts of individualized evidence. Here are some other important accounts: Some authors rely on further epistemological notions to mark out individualized evidence, such as knowledge (Blome-Tillmann, 2017; Littlejohn, 2020; Moss, 2018), relevant alternatives (Gardiner, 2019b), safety (Gardiner, 2020), or stakes and risk (Bolinger, 2021). Others tie individualized evidence to moral notions such as respect for a defendant's autonomy (Wassermann, 1992); finally, some appeal to the likelihood ratio to explain the puzzling examples (Cheng, 2013, Di Bello, 2019).

testimony. Correspondingly, the testimony of a reliable witness meets the standard of proof for civil cases; it suffices for finding Red Taxi liable for the accident.

On the other hand, there is Enoch et al.'s (2012) sensitivity account of individualized evidence. They suggest that individualized evidence is such that, had the relevant claim been false, the evidence wouldn't have obtained. That is to say, individualized evidence is counterfactually sensitive to the obtaining of the fact (or truth of the proposition) of which it is evidence: In the closest possible worlds in which the proposition is false, the evidence fails to obtain. Had the Red taxi not hit Jeanette (i.e. had the proposition the court needs to establish been false), the statistical evidence would have been unaffected (i.e. the evidence would still have obtained); it is not sensitive to whether it was a Red taxi that hit Jeanette. By contrast, had no Red taxi hit Jeanette, the witness would not have observed a Red taxi hit her, so that he would not have given the testimony that he saw a Red taxi hit Jeanette (i.e. the testimonial evidence would not have obtained). So this is individualized evidence, which is sensitive to the fact it is evidence of. So unsurprisingly, sufficiently reliable testimony meets the standard of proof.

We take a third, novel proposal to be more promising than either of these accounts. Martin Smith defends a normic account of individualized evidence—it is the kind of evidence that "normically supports a proposition" (Smith, 2018, p. 1211). A piece of evidence *e* normically supports a proposition *p* if and only if the obtaining of *e together* with the fact that *p* requires less explanation than the obtaining of *e without* the fact that *p*; *e* with *p* is more normal than *e* without *p*. Normalcy here is to be understood not as a matter of frequency (after all, it's supposed to provide an alternative to probabilistic support), but as a matter of whether a combination of facts would be in need of explanation. According to Smith, in cases of normic support, the explanation called for by *e* obtaining without *p* will typically involve some interfering factor, such as malfunctioning or unusual environmental conditions, which would explain how we have ended up with *e* despite *p*'s not obtaining (Smith, 2010, p. 17). In that sense, the combination of *e* but not *p* is in need of explanation.

Note that whether *e* with *p* requires less explanation than *e* without *p* cannot be relativized to whether a subject—say a judge—happens to find one or the other combination more normal. For this would open up the disturbing possibility that, e.g., a woman's testimony does not normically support the proposition to which she testifies because the biased judge believes that it is not abnormal for female witnesses to provide misleading testimony.[13] A more objective way to conceive of normalcy is this: A combination of facts is normal just in case it would be inappropriate to require further explanation, and it is abnormal just in case it would be appropriate to require further explanation. This is what we presuppose here.[14]

For example, in Prison Yard it would not be abnormal if prisoner *P*, despite the fact that 99% of the prisoners were involved in the attack, turned out to be the one prisoner who wasn't. This combination (*e* obtains, *p* doesn't) is not in need of explanation. Going by the statistical information about the case, it is going to be true of one of the prisoners that she was not involved in the attack, and so it can "just so happen" (Smith,

[13] To be clear: the same holds for *probabilistic* support. Whether a piece of evidence makes a proposition sufficiently likely to meet the standard of proof (probabilistically understood) cannot hinge on a subject's—potentially mistaken—beliefs about the relevant probabilities.

[14] We take the notion of appropriateness from discussions in the theory of normativity. Cf. Jacobson (2011).

2018, p. 16) that it was *P* who wasn't involved. As Enoch et al., (2012, p. 208) put it: "you win some, you lose some". By contrast, imagine that a video camera clearly shows *P* participating in the attack. Given this evidence, it *would* require extra explanation (such as that a prison employee tampered with the video—i.e., an interfering factor) if it wasn't true that *P* was involved in the attack; but it wouldn't require any explanation if *P* was involved in the attack, given the video evidence. So, the video camera evidence normically supports *P*'s involvement, but the statistical evidence does not.

The normic account of individualized evidence is more promising than its competitors. For, as Smith (2018) argues, on both the causal and the sensitivity accounts, misleading evidence is not individualized.[15] If the evidence obtains even though the claim evidenced by it is false, it doesn't meet these accounts' criteria for being individualized. If no Red taxi hit Jeanette, but still the witness testifies he saw a Red taxi hit her, there can be no causal relation between his testimony and a Red taxi hitting Jeanette. So, the evidence is not individualized. Further, in this case, the witness's testimony is not sensitive. For it's not true that if the claim evidenced were false, the evidence wouldn't obtain either. The actual world is among the closest possible worlds; and in it, no Red taxi hit Jeanette, but still we have the testimonial evidence that a Red taxi hit her. So this is not individualized evidence.

But this makes both accounts useless when it comes to giving judges or juries a handle on whether a piece of evidence they have available is individualized and thus able to meet the standard of proof. Judges and jury are *trying to figure out* whether the defendant committed the crime or was liable for the damage. For this, they need to know whether the relevant piece of evidence is individualized or bare statistical evidence. It doesn't help them to learn that *if* the defendant committed the crime or was liable for the damage, the evidence is individualized, and if not, not. Moreover, sufficiently reliable testimony or video recordings are individualized evidence—they meet the standard of proof—whether they are misleading or not. What we need, therefore, is a conception of individualized evidence that is neutral between the good and the bad case, i.e. that covers both the situation in which the evidence obtains, as does the state of affairs evidenced by it, and the situation in which the evidence obtains, but what it is evidence of does not.[16]

This condition is met by the normic account. No matter whether the defendant really committed the crime or caused the accident, a piece of evidence of these claims is either such that it would be more normal if, given the evidence, what the evidence indicates were the case—or it isn't. Because of this advantage of the normic account of individualized evidence, we will presuppose it here.[17]

[15] See also Gardiner (2019a).

[16] Admittedly, Enoch et al., (2012, p. 209) do address this issue, claiming that individualized evidence is evidence that is sensitive in the good case. We don't have the space to discuss this here; suffice it to say that an account that covers good and bad cases in its initial formulation seems to us more convincing. The same goes for a similar move that Thomson (1986) has available.

[17] Accordingly, our purpose here is to run with the normic account and see how far we can take it, and we will not call it into question.

2.2 Software systems

How do these issues connect to the use of software systems in the courtroom? The software-based systems that are increasingly used in the US legal system (cf. Piana, 2018)—though e.g. the German system is much more cautious in their adoption[18]—typically either generalize via correlations or provide probabilities instead of definite results. For instance, artificial neural networks internalize correlations from huge amounts of training data (cf. Hao, 2018). In short, software systems currently used in the legal system are largely based on statistics and not on deductive reasoning.

As e.g. Lipton (2018), Barocas (2014), and Mittelstadt et al. (2016) emphasize, software systems may not be sensitive to the individual case. In Mittelstadt et al.'s (2016, p. 5) words, "knowledge [about causality or correlations] may only concern populations while actions are directed towards individuals"; individuals may show up only as instances of simplified generalizations. Barocas (2014, p. 2) cites Ramirez (2013) as raising the worry that "big data [can] be used to make determinations about individuals, not based on concrete facts, but on inferences or correlations that may be unwarranted". All this indicates that such systems may provide evidence or predictions that are insensitive to the particular individuals they concern; that are reflections of general patterns into which they fit the individuals under consideration. This point holds not just for primarily data-driven AI. It is also true of systems with a more classical architecture that are nonetheless built to find and apply statistical correlations, such as those used for DNA mixture analysis. It is important to keep this in mind, for—as is the case both with TrueAllele and with COMPAS—the architecture of software systems currently in use in legal contexts is often a trade secret. What matters to our case is not their concrete architecture, but whether they are statistics-based.

It appears, then, that the evidence provided by software systems consists in statistics and is thus arguably bare statistical evidence. This raises the question whether it can ever be acceptable to base judicial decisions—such as finding a defendant guilty or liable, or issuing a harsher sentence—on the statistical evidence provided by such systems. We will address this question in the following by looking in more detail at the case of the recidivism prediction software COMPAS (Sect. 3) and at the case of the DNA comparison software TrueAllele (Sect. 4).[19]

3 Software systems and the need for individualized predictions

3.1 Predicting vs. establishing facts

Let us start by making explicit a distinction between the role of making predictions as compared to the role of establishing facts for judicial decisions.

[18] From a constitutional perspective, see Härtel 2019; see also the opinion of the German Federal Government, BT-Drs. 19/15131, p. 2 et seq.

[19] Other software systems used in US courtrooms include facial recognition systems, voice recognition systems, or fingerprint recognition systems. These raise issues similar to those raised by DNA comparison systems. Cf. Cino (2018) for discussion.

In some legal contexts, the law requires the court to make decisions based on predictions. For example, the question of whether a sentence is to be suspended depends on the prognosis of the offender's future behavior. Similarly, in custody disputes, it is necessary to decide to which parent custody should be granted in order to best serve the interests of the child. State v. Loomis is an example of the first kind: Eric Loomis was sentenced to 6 years in prison partly based on a high recidivism probability calculated by COMPAS. What is the role of a recidivism probability, as (for example) provided by a software system? The system output that a certain person is very likely to reoffend is not evidence used to establish past facts relevant for a decision whether to convict the person, but rather is evidence used to determine the correct prognosis for the person, which in its turn is used to decide the appropriate sentence. Put differently: In determining a sentence, the court takes into account how probable it is that the convict will reoffend—her prognosis. It needs to answer the question: How likely is it that the convict will reoffend in the next few years? To establish this recidivism probability,[20] courts can rely on different kinds of evidence, including statistical correlations between certain features of persons and recidivism risk levels, which give us general probabilities that persons with certain characteristics—shared by the convict—will reoffend. Such recidivism predictions can be provided by software systems like COMPAS; similarly to evidence provided to the court by experts, the outputs of such software systems can be thought of as evidence. Other types of evidence that might be used as the basis of a prediction are a convict's family ties, the prospect of a new job, or the convict's previous life.[21]

In many other cases, courts have to establish past facts. To come to a court decision, therefore, the questions, 'What has happened before? How did the crime/accident take place?' have to be answered. For example, criminal convictions are based on the (apparent) facts of the crime of which the court is convinced. In People v. Chubbs, Chubbs was convicted for murder on the basis of the very high probability that he was the source of the DNA trace found at the crime scene, as calculated by TrueAllele. The calculated probability in this case was used as evidence to determine the defendant's guilt.

Both when software systems are used to provide evidence for predictions and when they are used to provide evidence to establish past facts, legal decisions are at least partly based on their outputs; they have an impact on the judicial decision. We now turn to the worry that, because their outputs consist in bare statistics, it is illegitimate to use software systems to provide recidivism predictions.

[20] For the sake of the argument, and following the claims made by Equivant (see below), we assume that such a probability does indeed exist.

[21] According to Sec. 56, para. 1 of the German Criminal Code, the decision whether a sentence shall be suspended has to take into account the personality of the convict, her previous life, the circumstances of her offense, her behavior after the offense as well as her living conditions. This covers existing or emerging social ties, such as stable partnership or employment relationships, see *Kinzig*, in: Schönke/Schröder, Strafgesetzbuch [German Criminal Code], 30th Ed. 2019, Sec. 56 n. 31.

3.2 State v. Loomis

In State v. Loomis, Eric Loomis was sentenced to six years in prison. He had allegedly taken part in a drive-by shooting (though he only admitted to two lesser charges). The judge was motivated to saddle Loomis with this sentence in part by Loomis's high recidivism risk, as calculated by COMPAS. This software system, which is a trade secret of the private company Northpointe (now Equivant), takes data from a convict's criminal record and from an interview with him, and then provides a prognosis of how probable it is that a person with such features will re-offend. This functioning is based on COMPAS's model of how certain biographical and personal features are generally correlated with certain recidivism risk levels.

Loomis filed a motion for post-conviction relief, arguing, among other things, that the ruling violated his right to an individualized sentence. He held that the recidivism prognosis provided by COMPAS did not take into account his individual situation, and so did not give individualized evidence of his concrete recidivism risk; rather, it estimated the recidivism risk associated with people with certain features similar to his, and then applied that recidivism risk to him. In her assessment of the case, Justice Anne Rush Bradley concedes this point:

> … risk scores are intended to predict the general likelihood that those with a similar history of offending are either less likely or more likely to commit another crime following release from custody. However, the COMPAS risk assessment does not predict the specific likelihood that an individual offender will re-offend. Instead, it provides a prediction based on a comparison of information about the individual to a similar data group. (State v. Loomis, 881 N.W.2d 749 (2016).)

She insists, however, that the severity of the sentencing was not based on this bare statistical evidence provided by COMPAS all by itself, but that the judge had additional individualized information available. She suggests that the sentence was legitimate because it was based on a mix of statistical and individualized, and thus overall individualized, information. This leaves the worry that a judge, when faced with the recidivism prognosis of an (apparently) objective, reliable software system, may be unable to give the individualized information she has available its due weight. The worry is exacerbated by the fact that software systems like COMPAS are black boxes to those in court, so a judge cannot assess how close to or far from doing justice to the individual the system's 'reasoning process' is.

It appears that what is going on in State v. Loomis is very similar to what drives the debate about bare statistical evidence regarding conviction. Generally speaking, a just ruling has to be honed in on the individual judged by the ruling. This applies to both the conviction and the sentencing stage. It is illegitimate to find a defendant guilty or to sentence him based only on statistical or general features, as provided by systems like COMPAS. The judicial decision has to be based on evidence of *his* wrongdoing, or on evidence that *he* is likely to reoffend.[22]

[22] Another way to put this claim is to say that bare statistical evidence does not meet the standard of proof for sentencing. When a severe sentencing decision is based on a negative recidivism prognosis that is itself based on bare statistical evidence, that evidence is not sufficient to establish the claim that *the convict* will

But maybe this is going too fast. COMPAS is used to provide evidence during sentencing, not for a conviction. And even if the evidence on which a *conviction* is based cannot be bare statistical evidence, this does not imply that prognoses of recidivism relying on bare statistical evidence are an illegitimate basis for determining a *sentence*. One way to strengthen this objection is to point out that common law is more restrictive about the kinds of evidence that are allowed for conviction than for sentencing, e.g. regarding character evidence.[23] That bare statistical evidence does not suffice for conviction does not entail that such evidence is insufficient for sentencing, at least under common law. Still, we believe that it is possible to make a case against using a software system's recidivism prediction as a basis for a judge's sentencing decision if we consider common law.

We first wish to emphasize that apparently it *is* accepted in the US criminal justice system that the basis of a sentencing, including evidence of recidivism risks, has to be individualized (cf. the quote from Justice Bradley). Similarly, under the German Criminal Code ('Strafgesetzbuch'; 'StGB'), the sentencing of an offender has to take into account the effects which the sentence is expected to have 'on the offender's future life in society' (see Sec. 46 para. 1 StGB). Since it is the sentence's effect on the offender's life that is at stake, the sentence has to rely on an individualized prognosis, which should be based on individualized evidence.

Second, for the question of whether it is (or would be) *just* to convict and then sentence Loomis to six years in prison based on bare statistical information, it seems to make little difference whether this information takes the shape of bare statistical evidence at the conviction stage or of bare statistical evidence at the sentencing stage. At bottom, the call for individualized evidence is motivated by the idea that society ought to convict and punish a person for a crime only if the evidence can connect her to the crime, rather than merely show that there is a certain likelihood that someone with her features committed it. Convicting an individual is unjust if this connection is missing. But a similar point can be made for evidence used in support of recidivism predictions for sentencing. Where a conviction has to be based on sufficient (overall) individualized evidence of whether the defendant committed the crime, a sentencing decision has to be based on sufficient (overall) individualized evidence of whether a convict will re-offend.[24] A prediction that the convict will re-offend, which can be used to motivate a severe sentence, has to connect *her* individual back story and character to the expectation that she will re-offend. It has to be evidence that *she* will

Footnote 22 continued
likely reoffend. What is needed is individualized evidence. This is obvious under German laws, where the standard of proof for the sentencing is the same as for the trial decision (which are both part of a uniform proceeding). For the same reason, the objection we discuss in the following paragraph does not apply under German law.

Our line of thought leaves it open what the standard of proof for sentencing is on a *probabilistic* dimension. As our argument does not rely on settling this question, we do not answer it here, though see Lyons (1993) for discussion of this contentious issue in the US system. It is worth noting that algorithmic tools used to predict recidivism since the 1970s, including COMPAS, have had a predictive accuracy of no more than 65-70% (cf. Jorgensen 2021, p.12).

[23] We thank an anonymous reviewer for raising this problem.

[24] Obviously, this is given that we should base sentencing decisions on recidivism predictions at all. The sentencing decision needs to be based not only on such predictions, but also, for instance, on the severity of the crime committed.

re-offend, rather than evidence that someone who is like her in certain respects will re-offend. For instance, say that a close family member of the convict, who knows her character extremely well, is aware that the convict has been in a downward spiral for years. This is evidence about the individual and might be used as a basis of an individualized recidivism prognosis.[25] Without such individualized evidence, a severe sentence based on such evidence is unjust.

It is not acceptable to base a prediction and thus a sentencing decision on the fact that a convict belongs to a certain group (e.g. male, young, poor, African-American) whose members are statistically comparatively likely to recidivate. For it is unfair to punish an individual in response to his social or economic background which he cannot change; further, it is not true that *every single person* with such a background has a high recidivism risk, so the individual whose sentencing is at stake may very well get a severe sentence despite being low-risk himself (Starr, 2014). The worry is not alleviated if a system like COMPAS (which as a matter of fact uses 137 informational items as input) takes into account other factors, such as the convict's statements as to his social integration, employment situation, or aggression levels.[26] On the one hand, many such features will likely track features such as poverty; on the other, even if people with all these features *tend* to have a high recidivism probability, this is only a statistical correlation. It may still be that the particular individual under assessment is very unlikely to re-offend.[27]

This argument from justice can be further backed up by considerations put forth by Jorgensen (2021). As she argues, we need individualized evidence at the sentencing stage, since any person subjected to legal sanctions has the right to be treated as an individual. Jorgensen's core point is that each individual has a claim to an equal distribution of burdens and benefits of the rule of law. In particular, each individual has a claim not to bear any extra costs (beyond what she incurs by her own responsible action) so that others benefit, e.g., from having people from demographics with de facto higher general recidivism probabilities removed from the public for longer periods of time. But receiving a more severe sentence based on a negative recidivism prognosis, resulting from bare statistical evidence concerning the social group that one belongs to (as provided by a software system like COMPAS), violates this claim. So, by virtue of this claim, more severe sentences cannot be based on bare statistical evidence.

That recidivism predictions based on bare statistical evidence cannot support a severe sentence is respected by the normic account: Given the nature of recidivism predictions generated by extant software systems, as just described,[28] there is no need

[25] Note that individualization does not alleviate worries one might have regarding the *reliability* of such a prognosis, or regarding the use of recidivism predictions generally.

[26] Cf. https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE.html

[27] A further issue, which we cannot discuss here, is that systems like COMPAS apparently don't even do very well on the probabilistic dimension: COMPAS predicts recidivism with an accuracy of merely about 65%. This is not much better than chance. For suggestions that this is as good as prediction gets, see Yong (2018).

[28] We do not claim that software systems cannot even in principle provide individualized predictions, but merely that COMPAS and similar systems fail to do so. If individualized predictions are possible at all, then software systems should in theory be able to provide them (even if it may be difficult to build such systems in practice).

for extra explanation if the convict does not recidivate despite the negative prognosis. It would not be abnormal if the convict shared the features of a group with, on average, high recidivism rates, but did not re-offend. This suggests that, to be sufficient, a negative prediction provided by a software system would at least have to be such that in its light it would be more normal if the convict did commit another crime in the future, than if he did not. Evidence with this normic strength might be provided by someone who knows the convict very well, such as a friend, family member, or long-time classmate; or it might be provided by a psychological expert who has spent sufficient time to get to know and evaluate the convict's personality. Again, a family member may know that the convict is on a downward spiral, so that it would be abnormal if he didn't re-offend; or she may know that the convict is genuinely trying to be a better person, so that it would be more normal if he did not commit another crime. As Smith (2021) points out, it is not sufficient for normic support to merely fill in values in a list of certain relevant variables relating the convict to certain classes, thus treating him as an instance of such classes.

But can predictions ever be based on individualized evidence? Are they not, of necessity, based on statistical correlations between features of people and their behavior? How else are we going to make predictions about the future? Our response to this worry is that the normic account is not committed to the claim that all predictions are based on bare statistical evidence, even if we (also) rely on correlations when we make predictions: Individualized evidence is evidence that supports a proposition with normic force. Whether a piece of evidence does so is independent of whether it is evidence for a prediction or evidence of past facts. This can be made especially clear by thinking of normic support in terms of possible worlds, such that "a body of evidence E normically supports a proposition P just in case the most normal worlds in which E is true and P is false are less normal than the most normal worlds in which E is true and P is true." (Smith, 2010, pp. 16/17). This comparatively greater normalcy can hold with respect to E and a prediction as much as it can hold with respect to E and a past fact. To illustrate, say that convict *C*'s mother, who is very close to *C*, sincerely states that he has been on a downward spiral for the last few years, consuming harder and harder drugs, becoming less and less reliable and respectful, and getting increasingly into conflict with the law. Given this, the most normal worlds are ones in which *C* will indeed recidivate; by contrast, if *C* suddenly were to break the downward spiral and stop committing crimes, *this would call for an explanation*. For instance, we might appeal to an intervening factor such as *C* turning to religion or entering a stabilizing romantic relationship. This contrasts with the case involving a high recidivism probability provided by COMPAS. In line with our argument in the previous paragraph, the most normal worlds with this evidence from COMPAS, in which the convict doesn't recidivate, are no less normal than the most normal worlds with such evidence in which he does recidivate. So, evidence from COMPAS doesn't provide normic support that the convict will recidivate, and is bare statistical evidence.[29]

[29] Other accounts of individualized evidence can potentially give substance to a distinction between individualized and bare statistical evidence for predictions as well. Here is a suggestion for the sensitivity account: In the example, the fact that *C* is on a downward spiral may be counterfactually interdependent with whether *C* will indeed recidivate—in the closest possible worlds in which *C* does not recidivate, he is also not on a downward spiral. In other words, if *C* were not to commit another crime in the next two years,

A third issue with sentencing decisions based on bare statistical evidence relates to a point that Enoch et al. (2012) make with respect to individualized evidence for past facts: Part of the motivation behind convicting offenders is that this can be an incentive for people considering whether to commit a crime not to do so. (Their thought will be something like: 'If I don't commit the crime, I won't be convicted.') As to sentencing, it can plausibly also serve as an incentive—it is well-known that a more severe crime means a more severe sentence. But further, the prospect that if one is prone to commit another crime in the future, then one will get a harsher sentence, may be an incentive (to intend) not to re-offend after one has committed a crime. (The thought being: 'If I am/become the kind of person who doesn't re-offend, my sentence will be lighter.') To the extent that this incentive strategy can be successful, it presupposes that the evidence used for recidivism predictions hones in on the individual convict. For if a convict is given the impression that she will be put in a 'high recidivism risk' category no matter what she does now or how she intends to change, then she may rightly feel that she might as well *not* try to be a person who won't re-offend. That is to say, only sentences based on individualized evidence of recidivism risk can plausibly work as an incentive not to re-offend.

Again, this line of argument is supported by considerations from Jorgensen (2021). She emphasizes that laws ought to be "prospective" rather than retroactive (Jorgensen, 2021, p. 8). This is so that individuals can act intentionally to avoid violating these laws, but it also has implications for sentencing: decisions for a harsher sentence cannot be based on immutable features of a person, but should rely on features that are in principle under her control, for then she can act so as to avoid a harsher sentence.

Overall then, sentencing decisions should be based on individualized evidence and predictions, but systems like COMPAS fail to provide either one. This cautions against using their output as the (sole) basis of sentencing decisions. Our general suggestion is that evidence and predictions must be individualized on the normic account to provide a legitimate basis for severe sentences.[30]

We now turn to our core case of software systems that provide statistical evidence: to software used for DNA comparison. We begin with some general points.

Footnote 29 continued
it would not be the case that he is on a downward spiral, and the mother wouldn't testify to this (for this, we have to assume that c's being on a downward spiral couldn't have failed to lead him to recidivate, cf. Lewis 1973, pp. 565/66). So, the mother's testimony is individualized evidence. And a similar suggestion for the causal account: Say that *C* will indeed recidivate and that the fact that *C* is on a downward spiral is a cause of this. So this fact is individualized evidence for the prediction that he will recidivate. Contrast this with the evidence underlying the COMPAS prediction: Even if it is true that poor young Black male convicts in the US are comparatively likely to recidivate, this statistical fact does not *cause C* to commit another crime. So this is bare statistical evidence.

We do not have the space to investigate here whether these suggestions can be made to work. What is important to us is that a distinction between individualized and bare statistical evidence for predictions does not just make sense pre-theoretically, but can plausibly be borne out under different accounts of the distinction.

[30] Or more cautiously: If we are right that evidence at the sentencing stage ought to be individualized, then according to the normic account, the evidence provided by systems like COMPAS does not suffice as a basis for recidivism predictions used for sentencing.

4 Software systems and the need for individualized evidence

4.1 DNA evidence as a kind of statistical evidence

Despite the fact that both Anglo-Saxon and German traditions assume that bare statistical evidence does not meet the standard of proof, they allow that there are cases in which courts legitimately base their judgments on statistical statements.

Typical examples can be found in court decisions based on the analysis of DNA material. In criminal law, this concerns criminal convictions where the defendant is convicted on the basis of a 'match' of DNA traces. In civil law, such assessments are commonly used for paternity determination. In the first context, the DNA trace (e.g. found at the crime scene) is compared with the DNA of the accused person; in the second context, the DNA of the child is compared with that of a (presumed) father. Based on such a comparison, a statistical evaluation is applied, which results in a statistical statement about the agreement of the tested samples, typically expressed by a probability value.

The probability assessment is, on the one hand, based on the number of matches between trace and reference material determined in the respective procedure. On the other hand, the biostatistical probability statement depends on the frequency with which the individual DNA characteristics occur in the reference population. The decisive basis for the frequency statements on the individual traits are data published in population genetic studies. For the assessment of biostatistical probability, there are recognised standards in molecular genetic science which lead to reliable and equivalent results.

We focus on the interpretation of DNA evidence in the US criminal justice system, from which People v. Chubbs is taken. What a judge needs to know to determine whether a defendant is the source of a DNA trace is not only whether the defendant's DNA matches this DNA trace, but also how likely or unlikely it is that the DNA from some *unrelated* person also matches the DNA trace. This second probability is often stated as the 'Random Match Probability' (RMP), which measures "the probability of a matching DNA profile (not the person of interest's) within a chosen population." (Coble & Bright, 2019, p. 219) Another measure that is often used in court is the 'Likelihood Ratio' (LR), which is the ratio between (1) the probability of a match between the defendant's DNA and the crime scene DNA, given that the defendant is the source of the DNA, and (2) the probability of a match (again between the defendant's DNA and the crime scene DNA), given that some other person is the source of the DNA.

Neither the RMP nor the LR themselves provide what we really want: an estimate of how likely it is that the defendant is the source of the DNA trace, given that his DNA matches the trace. As Roth (2010, p. 1148) argues, in cold hit cases, we need to factor in the prior odds that the defendant committed the crime to get such an estimate. The prior odds can be extracted from the number of people who might have committed the crime, all told.[31]

[31] Roth takes the male population of California—18 million men—as her example, putting the prior odds that the defendant committed the crime at 1/18 million. If we multiply this with the reverse RMP, in Roth's

It is noteworthy that despite the ostensibly statistical nature of DNA evidence—after all, it is evidence as to the likelihood that the defendant (rather than someone else) is the source of a relevant DNA sample—DNA evidence is sometimes used in civil and criminal cases as the sole evidence on which a judicial decision is based, viz. in so-called cold-hit cases. We now turn to such a case, in which the evidence was provided by a DNA comparison system.

4.2 The use of software systems in cold hit cases

In 2012, Martell Chubbs was arrested as a suspect for having raped and murdered a young woman in her home in 1977. In 2016, he was convicted and sentenced for second-degree murder (Nguyen, 2016). This was a cold hit case: The only evidence connecting Chubbs—an African-American man—to the crime was the DNA found at the crime scene. The DNA trace included two to three distinct DNA profiles, according to Sorensen (the company which at first analyzed the DNA).[32] Sorenson found that Chubbs's DNA, which was stored in a national criminal database, matched the DNA trace, and that the RMP of the trace for a Black person was one in 10,000. Chubbs was convicted of the crime on the basis of a very high Likelihood Ratio, as (at that point) calculated by TrueAllele. TrueAllele, a software system owned by the company Cybergenetics, gave as the LR that a match between the DNA trace and Chubbs's DNA was 1.62 quintillion times more probable than a coincidental match to an unrelated Black person (cf. Smith et al., 2018, p. 295). Taking the LR to be just the reverse RMP, there is a great discrepancy between the probabilities calculated by Sorensen and by Cybergenetics. This relates to a different problematic feature of this particular case, the fact that the methods and codes of DNA comparison software systems are typically trade secrets, which US courts so far have not forced companies to disclose. This lack of transparency prevents adversarial testing of the DNA evidence provided by them, which presents a severe disadvantage for defendants (Chessman, 2017).

For the purposes of this paper, we have to leave this worry to one side. Recall that we are concerned only with cold-hit cases, and let us assume here that the LR calculated by TrueAllele is correct. Given this, does the evidence on which Chubbs

Footnote 31 continued
example 180 million/1, we get the posterior odds that the defendant is the source of the DNA at ten to one. The probability that the defendant is the source is at 91%.

Note that Roth's proposal is similar to, but distinct from the so-called 'defense attorney's fallacy' (cf. Mauro 2015). The fallacy is supposedly to assume that every other person in the relevant population had the same access to the crime scene (and thus to factor in the prior odds as proposed). Two points on this: First, in a cold hit case, we have as little evidence that the defendant had access to the crime scene as we have for any other person in the population. Second, the LR or RMP by themselves just are not very helpful, since they don't tell us how likely it is that the DNA trace is the defendant's. In any case, the probability is a lower bound for the probability, i.e., is at least as high as when choosing a smaller group, as decreasing the number of people with access to the crime scene will increase the probability. Hence, working with an over-approximation of the number of people with access to the crime scene yields a lower bound. If the probability that the defendant is the source of the crime scene DNA is still extremely high, as in our case, this just goes to show how strong the DNA evidence is.

[32] As we don't think that it is relevant to our argument whether we are dealing with a mixed DNA trace, we do not elaborate on this point. A further question is what the probability is that the DNA found at the crime scene is that of the perpetrator (whoever they may be). For our purposes here, we will assume that that probability is 1.

was convicted meet the standard of proof? Is it individualized evidence? TrueAllele's LR doesn't quite tell us what the probability is that Chubbs was the source of the DNA from the crime scene and thus the murderer. As explained above, we think that this probability is relevant for conviction. For the sake of our argument, we assume that Chubbs has been picked randomly from the population.[33] Assuming that in 1977 the male African-American population of California was at about 800,000[34] and that this gives us the pool of possible suspects, the prior odds that Chubbs was the perpetrator is 1/800,000. We can then calculate a posterior probability that it was Chubbs of more than 99.9999999999%.[35] This is a statistical statement, which can be used for an inductive inference to the conclusion that Chubbs committed the crime.

> *Inductive Picture*: It is more than 99.9999999999% likely that Chubbs is the source of the crime scene DNA. So, Chubbs committed the murder.

The inductive picture casts the posterior probability that Chubbs is the source of the DNA as bare statistical evidence, on the normic account. For if it turned out that someone else was the perpetrator and the source of the DNA, and that by some extremely remote chance Chubbs's DNA profile merely happened to match the perpetrator's DNA, this would be surprising, but not abnormal; there would be no need for a special explanation via an interfering factor. In the end, it would just be statistics—unlikely possibilities sometimes *are* actual, no explanation possible or needed. So apparently, the evidence on which Chubbs was convicted was bare statistical evidence. It did not meet the standard of proof for conviction, and the same holds for cold hit cases generally.

This would be a setback for the use of software systems in court cases—one standard use, as providers of DNA evidence in cold hit cases, would be ruled out.[36] And assuming that voice recognition systems, fingerprint recognition systems etc. function in the same way, this would also undermine their use in cold hit cases. However, it seems intuitively fine to base a conviction on a cold hit DNA match, at least given that the probability that the defendant is the source of the DNA trace is exceedingly high, as in our example.[37] Unlike in the standard examples, it is not initially obvious that the (alleged) bare statistical evidence in this case does not meet the standard of proof.

[33] Note that, in reality, Chubbs has not been drawn uniformly from the population. Instead his DNA was stored in a database with the help of which he was identified. One distorting factor is that Black men are greatly overrepresented in the US DNA databases (Murphy and Tong 2020).

[34] Cf. https://en.wikipedia.org/wiki/Demographics_of_California. Our motivation for taking the Black male population of California as the reference group for calculating the prior odds is that the LR provided by TrueAllele references the likelihood of a match with another Black person and that the DNA was extracted from a sperm sample. We were not able to find out why both Sorensen and Cybergenetics focused on probabilities with regard to the Black population.

[35] $1/800{,}000 \times 1.62e18 = 2.025e12$ as the posterior odds. Transformed into a posterior probability, this gives us approximately 99.99999999995%.

[36] Note that another way to address the issue would be to abandon the normic account. As stated above, we take this account for granted here.

[37] And given that the provided probability was calculated correctly, see above.

To put the abstract numbers into perspective: If every day 100 people are convicted in cold-hit cases where the probability is 99.9999999999% that they committed the crime, then, on average, approximately every 30 million years an innocent person is convicted. In comparison, the genus *Homo* first appeared with the *Homo habilis* just about 2 million years ago (Schrenk et al., 2007). This is not to say that it is just to

We suggest resolving this issue by providing an alternative picture of the relevant inference from evidence to conclusion, which is not inductive, but abductive. Contra Smith's own rather pessimistic view, we believe that DNA evidence calculated by software systems may come out as individualized after all, on the normic account, if we think of the pertinent inference as an inference to the best explanation (IBE). Our argument has two steps: First, we argue that an inference, from some evidence to the proposition that best explains it, entails that this evidence normically supports the proposition. Second, we argue that the evidence in People v. Chubbs abductively supports the verdict that he committed the murder.

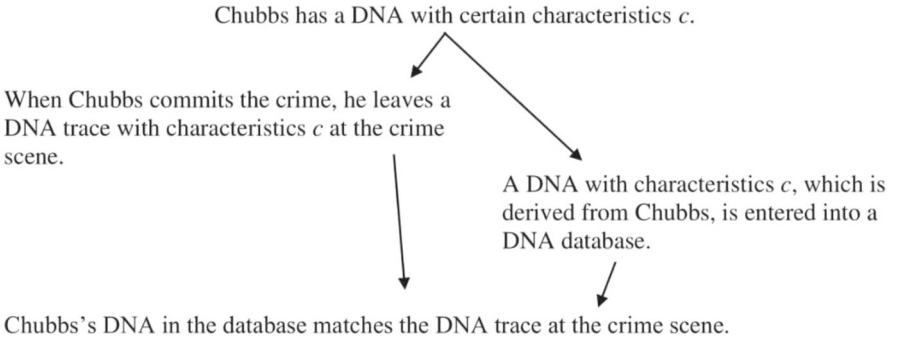
First, whenever some evidence *e* is best explained by a certain proposition *p*, there is normic support.[38] Consider any abductive inference. Given that *p* really best explains *e,* the combination of *e* and *p* will be more normal than combinations of *e* with competing (not-*p*) explanations. For, given that *p* best explains *e*, there is no room for an explanation, via an interfering factor, of why both occur together, and it's not appropriate to request a further explanation. That competing explanations are not as good means that e.g. they are more complicated or that they have less explanatory power. Either way, these explanations have to appeal to further circumstances fully to explain *e*—in other words, *e* and not-*p* is less normal than *e* and *p*. For example, comparing the heliocentric and geocentric worldviews, the heliocentric worldview best explains the movements of the planets. The geocentric worldview is more complicated and has to appeal to additional epicycles to make sense of the movements of the planets. So, this combination of evidence and explanatory hypothesis is less normal. In Smith's (2010, p. 15) terminology: A situation in which the evidence and its best explanation both obtain "is explanatorily privileged" over a situation in which the evidence obtains together with a sub-par explanation. So, abductive support guarantees normic support.

Second, a natural way of thinking about the reasoning used in People v. Chubbs is in terms of an inference to the best explanation. We are looking for an explanation of why Chubbs's DNA matches the DNA found at the crime scene. In light of the extremely low probability that another person is the source of the crime scene DNA, the best explanation of the evidence seems to be that Chubbs is the source of the crime scene DNA and thus the perpetrator. Call this explanation (H), as depicted in Table 1. This gives us the

Footnote 37 continued
convict this innocent person every 30 million years. However, it shows how unlikely it is to actually convict an innocent person.

[38] This is compatible with Smith's (ms.) statement that the normic account is distinct from explanationist accounts of justification. While abductive support entails normic support, normic support does not entail abductive support. See Smith (ms., 19). Here is an example involving individualized evidence that is covered by the normic account, but not by an explanationist account: Assume that it is relevant for a legal case whether Nancy was at the café at noon. Nancy was seen at the café engrossed in a book at 11:45 am. That she was seen at the café engrossed in a book at 11:45 is (at least some) individualized evidence that she was at the café at noon. But that she was seen at the café engrossed in a book at 11:45 is not best explained by the fact that she was at the café at noon; it is not best explained by something that takes place later in time. So, on an abductive account, this is not individualized evidence. By contrast, the normic account can maintain that this is a case of individualized evidence. It is more normal for Nancy to be seen at the café engrossed in a book at 11:45 and to still be at the café at noon, than for her to be seen at the café engrossed in a book at 11:45 and then not to be at the café at noon. It is because of problems with such cases that we prefer the normic account to an explanationist account of individualized evidence.

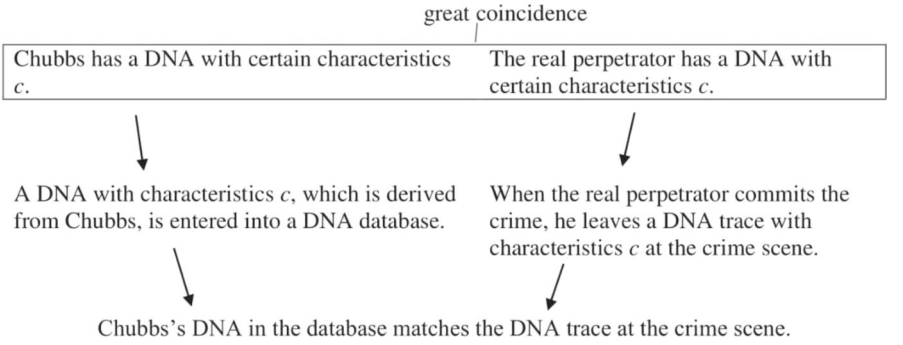
Table 1 Explanatory hypothesis (H)



IBE Picture: (1) The evidence consists in the fact that Chubbs's DNA matches the crime scene DNA. Given (2) the background information that it is exceedingly unlikely that someone else is the source of the DNA, (3) the simplest and best explanation of the evidence is provided by the hypothesis that Chubbs is the source of the crime scene DNA and committed the crime. So, (4) Chubbs committed the crime.

This inference to the best explanation is successful because the alternative explanations of the evidence are worse. The obvious alternative explanation is that—despite the exceedingly low probability that someone else is the source of the DNA—someone else with a matching DNA committed the crime and that Chubbs's DNA profile merely happened to match the perpetrator's DNA (as a false positive). Call this explanation (H'), as depicted in Table 2. It is worse than (H) on several interrelated counts. First, it is a worse fit with our background knowledge. The explanation that Chubbs committed the crime coheres better with what we know about how improbable it is that someone else is the source of the DNA, who committed the crime. Second, (H') is more complicated, in that it brings in a distinct causal factor (viz. another person who is the source of the crime scene DNA), while presenting the match between Chubbs's

Table 2 Explanatory hypothesis (H')



DNA and the crime scene DNA as causally disconnected. Third, the explanatory story it provides is internally less coherent and relies on an extreme coincidence: Instead of one coherent chain of events (Chubbs committing the crime and leaving DNA at the scene, which then matches his DNA), we are presented with an additional causal chain (unknown person X committed the crime and left a DNA trace) and a great coincidence (Chubbs's DNA accidentally matches X's DNA trace, i.e., the match is a false positive).

To make this more tangible, compare Tables 1 and 2. (H) involves one unified and coherent causal chain of events, in which the fact that Chubbs has a DNA with certain characteristics is behind both the facts about the DNA trace at the crime scene and about the DNA in the database. This fact thereby nicely explains the match. By contrast, (H') involves a great coincidence—although this is highly improbable, there is another person with a DNA with characteristics *c*—and then two distinct causal chains, which in the end explain the match between Chubbs's DNA and the crime scene DNA. Since, unlike (H), (H') relies on a huge coincidence, and then on two distinct causal chains, this explanation of why Chubbs's DNA matches the crime scene DNA is less coherent and simple than (H). Explanation (H) is better than its alternative, and an inference to the best explanation from the evidence to (H) goes through. So, assuming that we are right that abductive support entails normic support, the evidence (that Chubbs's DNA matches the crime scene DNA, against the background that it is exceedingly unlikely that someone else is the source of the DNA) also normically supports the proposition that Chubbs committed the crime.

Here is a worry about our proposal: Doesn't (H') explain the evidence just as well as (H) by definition?[39] Sure, it's extremely unlikely that Chubbs's match is a false positive and that there is another person who has a DNA with the same characteristics and who is the real source of the DNA trace at the crime scene. But, as elaborated above, sometimes extremely unlikely possibilities are actual. We cannot introduce an explanatory difference on the basis of a bare statistical difference.

In response to this, note first the difference between (H) and (H') that goes beyond the probabilities: the two hypotheses differ with respect to the causal relations that they introduce. Where (H) causally traces the match back to one causally relevant fact, (H') relies on two distinct causal factors and two distinct causal chains. These differences render (H) simpler and more coherent than (H'). Second, (H') but not (H) relies on the extreme coincidence that there is someone else with a matching DNA, so that Chubbs's match is a mere false positive. Regarding this second point, there is a similarity between the case under discussion and cases to which philosophers of science apply the so-called Common Cause Principle. Simply put, the Common Cause Principle states that "when apparent coincidences occur that are too improbable to be attributed to chance, they can be explained by reference to a common causal antecedent." (Salmon, 1988, p. 96) When the probability with which two event types occur together is greater than is to be expected from the probabilities of the event types occurring individually, we are allowed to infer to a common cause and thus a common explanation of both

[39] We thank an anonymous reviewer for raising this objection.

event types.[40] The Common Cause Principle is used to infer causal relations from merely probabilistic information (Hitchcock and Rédei 2021). What we want to take from this is that, in principle, there are ways to extract causal-explanatory information from statistical information. Moreover, we think that there is a shared motivation that backs both the Common Cause Principle and inferences to the best explanation more generally, which is that explanations that stipulate great coincidences suffer with respect to their theoretical virtues, specifically simplicity and coherence.

To provide one final motivation for our proposal that (H) is a better explanation of the evidence than (H') because it avoids postulating a great coincidence, we turn to the case of Sue, discussed by Smith (2010, p. 24): Sue's community hosts a hundred ticket lottery once a year and Sue buys a ticket every year. At the point in time where Sue participates in the lottery the first time, we apparently have bare statistical evidence that Sue will not win the lottery fifty years running. Now imagine that Sue has won the lottery every year for the last fifty years. According to Smith, this scenario raises the suspicion that the lottery was rigged, but this is due to a *bias* that humans have against patterned outcomes and coincidences.

Smith's analysis strikes us as implausible. The scenario indeed involves a great coincidence, and while, from a probabilistic standpoint, it may just be a cold hard fact of life that something highly improbable actually happens, we have more to go on if we apply abductive reasoning. Inference to the best explanation licenses us to accept a simpler, more coherent explanation that avoids extreme coincidences, viz. that the lottery is rigged. By contrast with what Smith says about this scenario, we submit that the belief that the lottery is rigged, given that Sue has won the last fifty years, is justified by way of inference to the best explanation, and then also normically justified.[41]

Overall then, our proposal is not to think of the statistical evidence provided by software systems in cold hit DNA cases involving extreme probabilities as evidence on whose basis we draw an inductive inference, but as a piece of the body of evidence which allows us to draw an abductive inference. We have further argued that the combination of a piece of evidence and its best explanation is more normal than the combination of the evidence and an explanatory hypothesis that's less good, and so that abductive inferences entail normic support. So, in cold hit cases like People v. Chubbs, there is normic support—the DNA evidence, though statistical, is not bare statistical evidence.

Our argument that cold hit DNA cases allow for an inference to the best explanation is restricted to cases of extremely high probabilities, in which an accidental match between the suspect's DNA and the crime scene DNA would be a great coincidence. To illustrate, imagine a case in which a suspect *S*'s DNA matches the DNA trace found

[40] As our speaking of event types indicates, this principle cannot be applied to our case directly, since we are dealing with the probabilities of particular event tokens. See Hitchcock and Rédei (2021), Hitchcock (1998, p. 428). It would be interesting to investigate what information would be needed to apply common cause reasoning to cold hit DNA cases, but this would lead us too far afield.

[41] Note, however, that the hypothesis that the lottery is rigged may be more complicated in another respect: It appeals to someone interfering with the lottery results, and so to complications that go beyond the simple account that every year, a person draws one number from the lottery. This distinguishes the scenario from cold hit DNA cases.

at the crime scene. Further, say that the posterior probability that *S* is the source of the crime scene DNA is 95%—there's a 5% posterior probability that the match between *S*'s DNA and the crime scene DNA is a false positive and someone else is the source of the DNA. Here it is unlikely, but not exceedingly unlikely, that someone else is the source of the crime scene DNA. It wouldn't be a great coincidence if *S*'s DNA merely happens to match the crime scene DNA, as a false positive. Given the numbers, it may well be that more than one person's DNA is a match. The DNA evidence is intuitively not individualized—since there may well be another person whose DNA matches the crime scene DNA, the fact that *S*'s DNA matches it doesn't tie *S* to the crime. Moreover, given the numbers, it seems that the match between *S*'s DNA and the crime scene DNA can be explained, but not *best* explained by *S* being the source of the crime scene DNA. The alternative explanation (someone else committed the crime and left the DNA trace, and the match between *S*'s DNA and the crime scene DNA is a false positive) seems as good as the explanation that it was *S*. It coheres equally well with our background knowledge about the probabilities; since we can safely assume that more than one person's DNA is a match, the overall explanation will not be more complicated; and that *S*'s DNA accidentally matches the crime scene DNA will not be much of a coincidence.

Similarly for Red Taxi. The fact that a taxi caused the accident can be explained, in part, by the hypothesis that a Red taxi caused it. (Imagine someone asking, "How come a taxi hit Jeanette?"—the answer, "Well, there was this Red taxi and it caused the accident," seems to be suitable.) However, the background information that 25% of taxis are owned by Green Taxi and 75% by Red Taxi doesn't suffice to make the explanation that a Red taxi caused the accident better than the explanation that it was a Green taxi. Again, that a Green taxi caused the accident coheres as well with our background knowledge about probabilities; since we know that there are Green taxis around anyway, it makes no difference in terms of simplicity whether it was a Red or a Green taxi that caused the accident; and it doesn't make sense to say that it is a coincidence that there are Red taxis in town that are also taxis. We conclude that cases like Red Taxi that have a different structure and lack extreme probabilities are not cases in which inference to the best explanation can get a grip, and we are not forced to say that they involve normic support.

Further, the structure of these cases is disanalogous from that of lottery cases. The proposition that my ticket will lose does nothing to explain that the probability of my ticket winning is, say, 0.0001%. What explains this probability is that there is only one winning ticket in a million tickets, not that my ticket will lose. DNA cold hit cases are also distinct from cases like Prison Yard. The proposition that prisoner *P* attacked the guard may be part of *an* explanation of the fact that 99% of the prisoners in the yard attacked the guard. (The suggestion is that what explains that 99% of the prisoners attacked the guard is that *these specific* 99 prisoners, *P* among them, were involved in the attack). But it is not part of the best explanation of this fact. For another explanation is equally good: That in which it was the other 99 prisoners who attacked the guard, and *P* stood by. So we don't have to worry that our argument will depict obvious cases of bare statistical evidence as involving individualized evidence.

But doesn't the IBE picture's reliance on (extreme) probabilities mean that our proposal reduces normic support to probabilistic support? Even worse, can't our alleged

IBE picture be better understood as probabilistic through and through? And if it's true that we rely on (mere) probabilistic support, doesn't this mean that our proposal on behalf of the normic account fails—that the evidence in the discussed cold-hit DNA cases is not individualized by our own lights?

In response, note first that relying on extreme probabilities in an inference to the best explanation doesn't mean that inference to the best explanation is, at bottom, a probabilistic inference pattern. Rather, we hold that background probabilities that a certain state of affairs obtains (such as the probability of there being a further matching DNA) can make a difference to what is the best explanation of a body of evidence. For they can affect how well explanatory hypotheses fit with one's background knowledge, the simplicity of these hypotheses, and their internal coherence. One can certainly wonder what the threshold of probability is at which background probabilities render one explanatory hypothesis better than another, enabling a successful inference to the best explanation. We don't need to give a definite number for this. This may have to be decided case by case. The threshold may be context-dependent and fuzzy. However, this does not mean that there are no clear cases in which statistical evidence provided by a software system meets the threshold and, hence, is individualized.

Second, what is important to our argument concerning cold-hit DNA cases is that they involve abductive and thus normic support, and so that the normic account can allow for the evidence in such cases to be individualized. We hope to have shown that they do by arguing for the superior quality of the explanation provided by (H) as compared to (H') above. By contrast, we can remain neutral regarding the nature of abductive (normic) support and in particular regarding the question whether it can be reduced to probabilistic support. It is a commonly shared assumption in epistemology, and in the debate over bare statistical evidence in particular, that explanatory or modal claims are distinct from claims about probability. While we find this assumption attractive, we can allow, one, that evidence which provides abductive (normic) support also supports the proposition in question on a probabilistic dimension; and two, that abductive (normic) support may be metaphysically reducible to probabilistic support—e.g. that it can be grounded in facts about Bayesian support.

Third, the concepts of probabilistic support and abductive support differ in important ways: Abductive support is a kind of support that works by maximizing intelligibility (of the evidence in light of the hypothesis), probabilistic support is a kind of support that works by raising the probability of the proposition in question. Our proposal reflects this difference by way of its differential treatment of People v. Chubbs as compared to the hypothetical case presented on p. 22, in which the posterior probability that the suspect is the source of the crime scene DNA is only 95%. While intuitively, the evidence in People v. Chubbs is individualized and suffices for conviction, the evidence in our hypothetical case is intuitively not individualized: The given probability doesn't sufficiently connect the suspect—rather than someone else with a matching DNA—to the DNA trace left at the crime scene. We can capture this by pointing to the missing extreme coincidence in the hypothetical case, which causes the inference to the best explanation, that succeeds in People v. Chubbs, to fail. By contrast, if we only had probabilistic support to go on to make sense of the two cases, then we would be unable to assess them differently with respect to individualization. For the evidence in both cases meets the (probabilistically interpreted) 'beyond reasonable

doubt' standard of proof. That a differential treatment is available to us indicates that our proposal respects the conceptual distinction between abductive and probabilistic support; the normic account can make use of this distinction to explain why DNA evidence provided by software systems is sometimes individualized.

Overall, then, our claim is that statistical evidence provided by software systems in cold hit cases is individualized evidence whenever it is best explained by the defendant's being the source of the crime scene DNA and thus the perpetrator. For then there is a normic relation between the evidence and what it is evidence of: Given the evidence, it *would* then be more abnormal if the worse explanation of the evidence were correct instead of the privileged explanation—it would be abnormal if the defendant had not committed the crime. It is compatible with this claim that there may be cases in which such evidence is not strong enough to support conviction, since abductive inferences can be weaker or stronger.

To conclude, the evidence provided by software systems like DNA comparison systems can be individualized, despite being generated by doing statistics, and thus meet the standard of proof for criminal cases.[42] Such evidence is individualized just in case it is best explained by the hypothesis that the defendant committed the crime. On this basis, it is less normal for the evidence to obtain, but for that of which it is evidence not to obtain, than for the evidence *and* that of which it is evidence to obtain. That is to say, there is then a normic connection between the evidence and what it is evidence of.

5 Results

The question we tried to answer in this paper was whether evidence provided by software systems can be a legitimate basis for judicial decisions. Our entry point was a widely accepted claim from legal epistemology—bare statistical evidence cannot meet the standard of proof for either civil or criminal cases. This, together with the fact that typical software systems used in the courtroom are statistics-based, raises the worry that the evidence they provide is unfit as the basis of judicial decisions.

We first focused on recidivism predictions based on bare statistical evidence such as that provided by COMPAS. Our argument was that bare statistical evidence, as a basis for sentencing decisions, is as bad as bare statistical evidence as a basis for conviction. For a sentence to be just, as well as for it to have a chance to serve as an incentive, it has to be based on features of the actual individual concerned, not just on general features of a group she belongs to which are correlated with high recidivism risk.

Our second case addressed software systems providing DNA evidence, such as TrueAllele. DNA analysis software employs complicated statistical analyses so as to estimate the probability that a defendant, rather than some other person, is the source

[42] Or, more cautiously: Assuming that convictions need to be based on individualized evidence, the normic account can allow that, in cold hit cases, convictions can be based on DNA evidence provided by software systems. Theorists denying the antecedent should still be interested to learn that the normic account can make this allowance—the intuitively legitimate use of DNA evidence in cold-hit cases is not a problem for the account.

of the DNA trace from the crime scene. We argued that such evidence, despite its statistical nature, can count as individualized evidence because of the role it plays in pertinent inferences to the best explanation. Our claim concerned cases of extreme probabilities in particular. This is consistent with the normic account of individualized evidence, for the best explanation of a piece of evidence is plausibly also the most normal explanation of that evidence.

If our proposal is on the right track, the typically statistical nature of the evidence provided by software systems does not pose an in principle problem for their use in the courtroom. On the one hand, it is legitimate to use statistical evidence, as long as it is supplemented by individualized information. On the other hand, to the extent that the evidence they provide is best explained by the hypothesis that the defendant committed the crime, such evidence will be individualized.

This, however, leaves much to worry about the use of software systems to support judicial decision-making. On the one hand, there is the worry that judges may give undue weight to the apparently objective evidence provided by them, while downplaying the central importance of individualized evidence. On the other hand, a whole host of issues is raised by the fact that the used software is often a trade secret, so that it is not transparent to those in court. If a defendant cannot examine the functioning of a software system, she cannot know whether it was faulty, nor can she appropriately defend herself. Further, it is hard to see how a judge can reach a well-founded decision if she cannot access the source code and other artifacts such as training data; she can really only take the provided evidence or leave it (cf. Baum et al., 2022). These issues relate to discussions of the explainability of software-based systems in interesting ways. Mere access to the relevant artifacts may be insufficient. Addressing these issues will have to wait for another occasion.

Acknowledgements We are grateful to audiences in Kraków and Hamburg for helpful comments and discussion. We would like to thank members of the research project *Explainable Intelligent Systems* for their input. Finally, we'd like to thank two anonymous reviewers of this journal for their challenging comments on the paper.

Funding Open Access funding enabled and organized by Projekt DEAL. This paper was supported by the Volkswagen Foundation, as part of the project Explainable Intelligent Systems (EIS), Project Numbers AZ 98510, AZ 98511, and AZ 98514, and by the German Research Foundation (DFG) under Grant No. 389792660, as part of TRR 248, see https://perspicuous-computing.science.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest (financial or non-financial).

References

Bambauer, J. (2015). Hassle. *Michigan Law Review, 113*(4), 461–511.

Barocas, S. (2014). Data mining and the discourse on discrimination. *Proceedings of the data ethics workshop, conference on knowledge discovery and data mining (KDD).*

Baum K, Mantel, S., Schmidt, E., & Speith, T. (2022). From responsibility to reason-giving explainable artificial intelligence. *Philosophy and Technology* 35, Topical Collection on AI and Responsibility, In N. Conradie, H. Kempt, and P. Königs (eds) pp. 1–30.

Bolinger, R. J. (2021). Explaining the justificatory asymmetry between statistical and individualized evidence. In Z. Hoskins & J. Robson (Eds.), *The social epistemology of legal trials* (pp. 60–76). Routledge.

Blome-Tillmann, M. (2017). 'More Likely Than Not'—knowledge first and the role of bare statistical evidence in courts of law. In A. Carter, E. Gordon, & B. Jarvi (Eds.), *Knowledge first—approaches in epistemology and mind* (pp. 278–292). Oxford: Oxford University Press.

Cheng, E. K. (2013). Reconceptualizing the burden of proof. *Yale Law Journal, 122*(5), 1254–1279.

Chessman, C. (2017). A 'Source' of error: Computer code, criminal defendants, and the constitution. *California Law Review, 105*, 179–228.

Cino, J. G. (2018). Deploying the secret police: The use of algorithms in the criminal justice system. *Georgia State University Law Review, 34*, 1072–1102.

Coble, M. D., & Bright, J.-A. (2019). Probabilistic genotyping software: An overview. *Forensic Science International, 38*, 219–224.

Di Bello M. (2019). Trial by statistics: Is a high probability of guilt enough to convict? *Mind, 128*(512), 1045–1084. https://doi.org/10.1093/mind/fzy026

Enoch, D., Spectre, L., & Fisher, T. (2012). Statistical evidence, sensitivity, and the legal value of knowledge. *Philosophy & Public Affairs, 40*, 197–224.

Gardiner, G. (2019a). Legal burdens of proof and statistical evidence. In D. Coady & J. Chase (Eds.), *Routledge handbook of applied epistemology* (pp. 179–195). Routledge.

Gardiner, G. (2019b). The reasonable and the relevant: Legal standards of proof. *Philosophy & Public Affairs, 47*(3), 288–318. https://doi.org/10.1111/papa.12149

Gardiner, G. (2020). Profiling and proof: Are statistics safe? *Philosophy, 95*, 161–183.

Goldman, A. (1967). A causal theory of knowing. *Journal of Philosophy, 64*, 357–372.

Härtel, I. (2019). Digitalisierung im Lichte des Verfassungsrechts—Algorithmen, Predictive Policing, autonomes Fahren. *LKV, 25*, 49–60.

Hao, K. (2018). What is machine learning? We drew you another flowchart. *MIT Technology Review*: https://www.technologyreview.com/s/612404/is-this-ai-we-drew-you-a-flowchart-to-work-it-out/, Accessed 22 March 2019.

Hitchcock, C. (1998). The common cause principle in historical linguistics. *Philosophy of Science, 65*(3), 425–447. https://doi.org/10.1086/392655

Hitchcock, C., & Rédei, M. (2021). Reichenbach's common cause principle. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), https://plato.stanford.edu/archives/sum2021/entries/physics-Rpcc/.

Jacobson, D. (2011). Fitting attitude theories of value. In E. N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition), URL: https://plato.stanford.edu/archives/spr2011/entries/fitting-attitude-theories/.

Jorgensen, R. (2021). Algorithms and the individual in criminal law. *Canadian Journal of Philosophy, 51*, 1–17. https://doi.org/10.1017/can.2021.28

Knauer, C., Kudlich, H., & Schneider, H. (eds.) (2016). *Münchener Kommentar zur Strafprozessordnung*, Vol. 2 (Sec. 151-332 StPO). Beck.

Krüger, W., & Rauscher, T. (Eds.), (2016). *Münchener Kommentar zur Zivilprozessordnung*, 5th Edn, Vol 1 (Sec.1–354 ZPO). Beck.

Lewis, D. (1973). Causation. *The Journal of Philosophy, 70*, 556–567.

Lipton, Z. (2018). The mythos of model interpretability. *Queue Machine Learning, 16*, 1–27.

Littlejohn, C. (2020). Truth, knowledge, and the standard of proof in criminal law. *Synthese, 197*(12), 5253–5286. https://doi.org/10.1007/s11229-017-1608-4

Lyons, D. J. (1993). Federal sentencing guidelines: Retaining the preponderance standard of Proof. *St. John's Law Review, 67*, 639–653.

Mauro, D. (2015). Fourteen Zeros. Blog Entry on the blog *Ipsa Loquitur.* https://blog.ipsaloquitur.org/post/fourteen-zeros/, Accessed 25 March 2019.

Mittelstadt, B., et al. (2016). The ethics of algorithms: Mapping the Debate. *Big Data and Society, 12*, 1–21.
Moss, S. (2018). *Probabilistic knowledge*. Oxford University Press.
Murphy, E., & Tong, J. H. (2020). The racial composition of forensic DNA databases. *California Law Review, 108*, 1847–1911.
Nguyen, A. (2016). Man pleads no contest to 1977 murder of teen mom. *Patch*. https://patch.com/california/longbeach-ca/man-pleads-no-contest-1977-murder-teen-mom, Accessed 25 March 2019.
Nozick, R. (1981). *Philosophical explanations*. Harvard University Press.
Piana, D. (2018). Predicting justice: what if algorithms entered the courthouse?. World Economic Forum, https://www.weforum.org/agenda/2018/11/predicting-justice-what-if-algorithms-entered-the-courthouse/, Accessed 22 March 2019.
Ramirez, E. (2013). The privacy challenges for big data: A view from the lifeguard's chair. Speech at *Technology Policy Institute Aspen Forum.*
Roth, A. (2010). Safety in numbers—deciding when DNA alone is enough to convict. 85 N.Y.U. L. Rev.: 1130–1185.
Salmon, W. (1988). Statistical explanation and causality. In J. C. Pitt (Ed.), *Theories of explanation* (pp. 75–118). Oxford University Press.
Schönke, A. & Schröder, H. (Eds.). (2019). *Strafgesetzbuch, Kommentar.* Beck.
Schrenk, F., Kullmer, O., & Bromage, T. (2007). The earliest putative homo fossils. In W. Henke & I. Tattersall (Eds.), *Handbook of paleoanthropology* (pp. 1611–1631). Springer.
Smith, J. C., Larson, E. J., & Nagle, J. C. (Eds.). (2018). *Property: Cases and materials.* Wolters Kluwer.
Smith, M. (2010). What else justification could be. *Noûs, 44*, 10–31.
Smith, M. (2018). When does evidence suffice for conviction. *Mind, 127*, 1193–1218.
Smith, M. (2021). More on normic support and the criminal standard of proof. *Mind, 130*, 943–960.
Smith, M. (ms). *Justification, normalcy and evidential probability*. http://philpapers.org/rec/SMIJNA-2.
Starr, S. B. (2014). Evidence-Based Sentencing and the Scientific Rationalization of Discrimination. *Stanford Law Review, 66*, 803–873.
Thomson, J. J. (1986). Liability and individualized evidence. *Law and Contemporary Problems, 49*, 199–219.
Urbaniak, R., & Di Bello, M. (2021). Legal Probabilism. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Fall 2021 Edition), https://plato.stanford.edu/archives/fall2021/entries/legal-probabilism/.
Wasserman, D. T. (1992). The morality of statistical proof and the risk of mistaken liability. *Cardozo Law Review, 13*, 935.
Yong, E. (2018). A popular algorithm is no better at predicting crimes than random people. *The Atlantic*. https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/, Accessed 04 Oct 2019.
Zöller, R. (or.) (2016). ZPO. Zivilprozessordnung. 31st Edition.