Presumed Prior, Contextual Prior, and Bizarre Consequences

Dahlman, Christian; Kolflaath, Eivind

Letter to the Editor

Presumed prior, contextual prior, and bizarre consequences—a reply to Ronald Meester and Lonneke Stevens

Christian Dahlman[1],* and Eivind Kolflaath[2]

[1]Lund University, Sweden
[2]University of Bergen, Norway
*Corresponding author. E-mail: christian.dahlman@jur.lu.se

The problem of the prior is a hotly debated issue in the literature on legal evidence. In a recent contribution to this debate, Ronald Meester and Lonneke Stevens argue that the prior must take 'context' into account (Meester and Stevens 2024: 8). They do not explain what this means or how it should be done, but their paper offers some examples that give a rough idea what they are after. Taking account of context could, for example, mean that the prior probability is higher when someone is accused of a crime in a small village compared to someone accused of crime in a large city with a greater number of alternative perpetrators, and with regard to two crimes both committed in the same large city, context can be taken into account for by assigning a higher prior to a suspect living in the part of the city where the crime took place than a suspect living in another part of the city. Meester and Stevens recognize that this approach is 'very problematic' (Meester and Stevens 2024: 4), but then go on to use it themselves in their examples of 'applying context'.[1]

A problem with this *contextual prior* is that it becomes highly arbitrary what should be regarded as the geographical area where the crime was committed. What should count as the 'same part' of a city and what should count as a different part? The same street? The same neighborhood? The same borough? And why should the small village be regarded as the geographical area where the crime was committed rather than the county where the village is located? This problem is well known in the literature on the problem of the prior (Dahlman and Kolflaath 2021: 292–5, with further references). A contextual prior that leaves it open to the individual fact-finder in each case to decide what should be regarded as the context of the case at hand is extremely arbitrary, and will result in arbitrary and highly diverging priors. There is no guarantee that cases with very similar geographical circumstances will not be decided on substantially different priors, due to different decisions on context by the fact-finders, and lead to different outcomes under the same evidence (acquittal with a low prior and conviction with a higher prior).

As a way to avoid this unacceptable arbitrariness, we have explored the possibility that a legal system could operate with a *presumed prior* that is equal for all cases (Dahlman and Kolflaath 2021: 295–7). In such a system, the prior probability would be the same whether the crime has

---

[1] For example, 'there are almost 9,000 people in The Netherlands with this partial profile' (p. 6) and 'the suspect regularly visited the boat where the victim was staying' (p. 11).

been committed in a small village or a large city, and the fact-finder would be deprived of the arbitrary choice of geographical context. In our paper, we explore the advantages and disadvantages of this way to solve the problem of the prior. The upside, as already mentioned, is that it escapes the severe arbitrariness that follows in the footsteps of approaches to the prior that rely on 'context'. The downside of the presumed prior is that it is detached from the fact-finders degree of belief in the prosecutor's hypothesis based on background knowledge about the world (Dahlman and Kolflaath 2021: 297). Certainly, a presumed prior is not an ideal solution to the problem of the prior, but no ideal solution exists. In our view, the presumed prior is the least problematic solution that has been proposed in the literature so far.[2]

   In their recent paper, Meester and Stevens argue against the presumed prior. They claim that we have overlooked the consequences of disconnecting the prior from the geographical context (Meester and Stevens 2024: 8). This is not true. In our paper, we discuss the consequences of this disconnection as an argument against the presumed prior (Dahlman and Kolflaath 2021: 297). More importantly, Meester and Stevens claim that the presumed prior leads to 'bizarre' and 'untenable' consequences that are avoided if we instead operate with a contextual prior (Meester and Stevens 2024: 6–8). What they fail to see is that these consequences follow from the contextual prior as well. Their examples are constructed so that it appears that the alleged problems arise from the presumed prior and go away if we use the contextual prior. Meester and Stevens present the following example as an argument against the presumed prior.

> Suppose a partial DNA profile is found with a likelihood ratio of 2,000 for the hypothesis that it comes from suspect, versus it comes from an unknown, unrelated person. Suppose that there are no further pieces of evidence in the case, apart from the fact that the suspect has no alibi: he claims to have been at home, alone, when the crime was committed. According to the procedure sketched by [Dahlman and Kolflaath], this should lead to a conviction, since having no alibi will not lead to a decrease in the likelihood ratio. However, there are almost 9,000 people in The Netherlands with this partial profile, and many of them would not have alibi's either. Should we convict each and every single person in that group? Obviously, this would be a bizarre conclusion. (Meester and Stevens 2024: 6)

In this example, a presumed prior combined with a likelihood ratio of 2,000 leads to a conviction. The example does not say what the prior is. Since a presumed prior of 1% is mentioned in the previous section of the paper and the evidence is sufficient for a conviction, we assume that what Meester and Stevens have in mind is a prior of 1% resulting in a posterior probability of 95.3% ($1/99 \times 2{,}000 \approx 0.953/0.047$), which they take to be sufficient for conviction. According to Meester and Stevens, such a conviction would be 'bizarre' since the population of the Netherlands is 17 million. There are $17{,}000{,}000/2{,}000 \approx 9{,}000$ people who fit the evidence just as well as the suspect, and that means that if we start with a prior of 1% in a hypothetical trial against any of them, we would always end up with a posterior of 95.3% leading to a conviction. Does this mean that we could convict anyone of them? Or convict all of them? The argument against the presumed prior, presented by means of these rhetorical questions, is that it leads to the bizarre result that the joint posterior probability for 9,000 people where each has a posterior probability of 95.3% grossly exceeds 100%, which does not make sense mathematically. As Meester and Stevens point out, the total posterior exceeds 100% in this example because the total prior exceeds 100%. They argue that if 9,000 people each has a presumed prior of 1%, the joint prior probability already exceeds 100%.

> This means that the total probability mass of all priors together might very well exceed 1. Obviously, this means that we are no longer in a well-defined mathematical framework, in which the total probability mass will always be exactly equal to 1. In that situation, "probability" is no longer a well-defined concept anymore, and the calculus of probability no longer applies. (Meester and Stevens 2024: 7)

---

[2]   In our paper, we also discuss the possibility of reconceptualizing the standard of proof in a way that replaces the probability threshold for the posterior probability of the prosecutor's hypothesis with a threshold for the combined strength of the evidence, that is, the Bayes factor for all the evidence in the case (p. 298–9).

This is the argument against the presumed prior, and according to Meester and Stevens these bizarre consequences go away if we use a contextual prior instead. If we use a contextual prior based on the Dutch population the prior is only 1/17,000,000 for each of the 9,000 people that fit the evidence just as good as the suspect, and the posterior probability for each of them only amounts to 0.012% ($1/17,000,000 \times 2,000 \approx 0.00012/0.99988$), which is not sufficient for a conviction against any of them. This contextual prior avoids a result that begs the rhetorical questions above ('convict any of them?', 'convict all of them?'). But it does not avoid the 'bizarre' consequence that the joint posterior exceeds 100%, since $9,000 \times 0.012 = 106\%$.

Meester and Stevens have chosen numbers that beg the rhetorical questions if we use a presumed prior of 1% as opposed to a contextual prior based on the Dutch population. This may leave the impression that the 'bizarre' consequences follow from the presumed prior and are avoided by the contextual prior. But this is not true, which becomes obvious if we change the numbers in the example. Let us increase the likelihood ratio of the partial DNA match to 350 million, all other things remaining equal. Now, the same contextual prior based on the Dutch population will result in a posterior of 95.4% ($1/17,000,000 \times 350,000,000 \approx 0.954/0.046$), which is sufficient for a conviction in the given example. The problem with such a conviction is that the perpetrator does not have to belong to the Dutch population. It could have been a visitor from France, or any other country. In fact, if we are thinking about all the individuals who could have committed the crime and could potentially be prosecuted for it, we must widen our scope to include the population of every country around the globe. The world population is 8 billion people. This means that there are expectedly $8,000,000,000/350,000,000 \approx 23$ people who fit the evidence just as well as the suspect, and since the European population is 740 million, two of them can be expected to live in Europe ($740,000,000/8,000,000,000 \times 23 \approx 2.1$). One of these two is our Dutch suspect with a posterior probability of 95%, and the other can be expected to live somewhere else in Europe outside the Netherlands. What is the posterior probability for this non-Dutch person? A contextual prior based on the European population leads to a posterior of 32% ($1/740,000,000 \times 350,000,000 \approx 0.32/0.68$), which means that the joint posterior of these two individuals is 95% + 32% = 127%. This shows that the 'bizarre' consequence is not avoided by using a contextual prior instead of the presumed prior. It only transports the 'bizarre' consequence to a different likelihood ratio.

We get the 'bizarre' consequence because the prior based on the Dutch population is too high with regard to the expected probability that one of the twenty-three people in the world matching the DNA profile will be living in the Netherlands. The expected number of Dutch people matching the DNA profile is less than one ($17,000,000/8,000,000,000 \times 23 \approx 0.05$), but by using the Dutch population to form the contextual prior for the Dutch suspect, we calculate as if there was one expected perpetrator within the Dutch population.

We could avoid the 'bizarre' consequence if we base the prior for the Dutch suspect on the whole world population instead of the Dutch population (setting the prior at 1/8,000,000,000 instead of 1/17,000,000), but this is not to 'take context into account' as Meester and Stevens advocate. In fact, it would be the exact opposite of taking context into account to base the prior on the entire world population regardless of where the crime has been committed and where the suspect lives.

We have now seen that things are not as simple as they appear. The root of the problem is not the presumed prior. As we start to understand that things are not what they appear to be at a first glance, we should continue this line of thought and ask ourselves if the consequences are really as 'bizarre' as they appear. On closer scrutiny, it turns out that they are not. As we have seen, the joint probabilities exceed 100% when we add up the probability for person X in an indictment against X with the probability for Y in an indictment against Y. But this only shows that we cannot treat probabilities in different indictments as belonging to the same probability distribution. In a trial against X the joint probability of X and alternative perpetrators will always add up to 100%, and in a trial against Y the joint probability of Y and alternative perpetrators will also add up to 100%. But we cannot mix these distributions. The two trials are two separate worlds with regard to probability distribution.

The need to keep the two probability distributions apart is a consequence of the normative legal framework in which these probabilities are situated. The normative framework that lawyers refer to as 'fair trial' is constructed with safeguards that protect the interests of the defendant, and when we say 'defendant' we mean the person accused in the trial at hand, not hypothetical defendants in other trials. One of these safeguards is the presumption of innocence. In a trial against X, the presumption of innocence is in effect with regard to X, but not with regard to Y and other alternative perpetrators. In a trial against Y, the presumption of innocence is instead in effect with regard to Y, but not with regard to X and other alternative perpetrators. A fact-finder who uses a contextual prior in the trial against X, and sets the prior at 1/17,000,000 because the Netherlands is taken as the geographical 'context' of the trial, is hereby interpreting and applying the presumption of innocence in this contextual way. As we all know, the actual class of possible perpetrators is not limited to the Netherlands, but the fact-finder is over-ruling this fact with a normative presumption that adequately protects the interest of the defendant (the defendant's interests could have been protected even more by setting a prior based on the entire world population, but this would have made it too hard for the prosecution to produce sufficient evidence). And the same happens with regard to Y in a trial against Y. The 'bizarre' consequence that the joint probability of X and Y exceeds 100% only occurs if we make the mistake of treating 'the probability of X as seen under the presumption of innocence' and 'the probability of Y as seen under the presumption of innocence' as existing in the same probability space. To see how this mistake leads to bizarre consequences, we can simply imagine a case where a crime as has been committed in a locked room and the perpetrator must be either X or Y. If X is accused, the presumption of innocence prescribes that in the eyes of the law X was not perpetrator until proven guilty. And if Y is accused, the same goes for Y. So, if we combine these two presumptions, the bizarre consequence follows that in the eyes of the law one of them must have done it, but none of them did. The error here is, of course, to combine the presumptions. One presumption only exists in the trial against X, and the other only exists in the trial against Y.

This shows that it has consequences for the prior how the presumption of innocence is interpreted and applied. Meester and Stevens incorrectly believe that the presumption of innocence has nothing to do with the prior (Meester and Stevens 2024: 8–9), and therefore incorrectly think that the consequences they regard as 'bizarre' are created by the presumed prior. As we have shown in this reply, the problem of the prior is much deeper than they realize.

The paper by Meester and Steven has several other analytic shortcomings, that will not be discussed in detail here, but should be mentioned briefly. They never define what they mean by 'context', and their examples of 'context' include a mixture of background knowledge and case specific evidence, without distinguishing between the two.[3] Also, they do not explain what they mean by 'prior', mixing examples about the prior probability of the prosecutor's hypothesis with examples about the prior probability of a source-level hypothesis that is not the ultimate hypothesis in the case, without distinguishing between these categories of hypotheses. The presumed prior is applicable only to the prosecutor's hypothesis.

There is a fundamental mistake in the approach that Meester and Stevens have to the problem of the prior, and it is that they fail to see that legal fact-finding takes place within a normative framework that constrains how hypotheses and evidence are treated in that framework. There are ways to formulate hypotheses and evaluate evidence that are perfectly fine outside the law but not allowed in a criminal trial. For example, because the burden of proof is on the prosecutor's hypothesis, the alternative hypothesis that cannot leave room for reasonable doubt is always the negation of the prosecutor's hypothesis. It is not sufficient proof for the prosecutor's hypothesis 'X did it' that the alternative hypothesis 'Y dit it' is not probable enough to leave room for doubt. The standard of proof requires that the negation 'X did not do it' does not leave room for reasonable doubt. Meester and Stevens incorrectly think that the fact-finder is free to define the alternative hypothesis at issue, choosing between the negation of the prosecutor's

---

[3] Failing to make this distinction, Meester and Stevens seem to assume that the case-specific evidence that they treat as 'context' (e.g. 'the victim was in a circuit of homeless people and (alcohol) addicts and, according to witnesses, he was often beaten by people from that circuit', p. 11) would not be considered by a fact-finder applying a presumed prior, and use this as an argument against the presumed prior. This is, of course, incorrect. To the extent that the case-specific evidence is deemed relevant, it will be included in the updating of the presumed prior.

hypothesis and other formulations of the alternative hypothesis. They suggest that a fact-finder in the Sally Clark case would be free to choose between 'Sally Clark did not murder her children' (the negation of the prosecutor's hypotheses) and 'Sally Clark's children both died of SIDS' as the alternative hypothesis (Meester and Stevens 2024: 6–7), and they continue to make the same error in other examples.

Meester and Stevens maintain that when they are 'taking account of context' in their reasoning about the prior they are not being 'normative'. They talk as if they are doing something purely epistemic that does not involve the application of the presumption of innocence or other legal norms (Meester and Stevens 2024: 8–9). What they do not acknowledge is that their own reasoning is shaped by some interpretation of the presumption of innocence. From a purely epistemic point of view there is no reason to frame the context of a crime as the population of the Netherlands. No one would think of the context in such a way if it was not for the presumption of innocence. To conceptualize a contextual prior in terms of a geographical area in which everyone is a possible perpetrator, but no one outside it, makes no sense from a purely epistemic point of view. The only thing that Meester and Stevens achieve by describing their method as 'taking account of context' is to obscure to the reader what norm they are actually applying.

Funding

No funding to declare.

Conflicts of interest

No conflict of interest to declare.

References

Dahlman, C., and Kolflaath, E. (2021) 'The Problem of the Prior in Criminal Trials', in Dahlman, Tuzet and Stein (eds) *Philosophical Foundations of Evidence Law*, pp. 287–300. Oxford: Oxford University Press.

Meester, R., and Stevens, L. (2024) 'Bayesian Reasoning and the Prior in Court—Not Legally Normative but Unavoidable', *Law, Probability & Risk*, **23**: 1–14.