ARTICLE

'The plausible, the possible and the probable'

Philosophical comments on plausibility in relation to possibility and probability[1]

*Prof. dr. A.R. Mackor**

1. Not certainty, but probability

In law, as in everyday life and in science, facts cannot be established with certainty, but only with a degree of probability. Probability judgements and plausibility judgements are closely related. Both express chances.[2]

If we consider it more likely that a fire was caused by arson than by lightning, we can say that arson is more probable, but also that arson is more plausible than lightning. In law, different standards of proof place different requirements on the degree of probability with which alleged facts must be proven. For example, the standard of 'beyond a reasonable doubt' requires a higher degree of probability than the standard of 'preponderance of probabilities'. Similarly, plausibility standards seem to require an - otherwise rather indefinite - chance that an alleged fact occurs.

This applies, for example, to the plausibility of danger as a condition for the imposition of preventive measures in criminal law; the plausibility of danger as a condition for the imposition of a muzzle order in administrative law and the plausibility of urgent own use of residential premises as a condition for the termination of the rental agreement in civil law.

However, it is not these standards of plausibility but the concept of plausibility that is the focus of this contribution. I aim to provide an explication in the sense of the philosopher Rudolf Carnap. Carnap understood explication to mean the replacement of a vague and imprecise concept, the *explicandum*, with a more precise concept, the *explicatum*. Carnap placed four demands on explications: precision, similarity, fruitfulness and simplicity.[3] I try, on the one hand, to tie in with the actual use of the term plausible (similarity), but on the other hand I also aim to give a specification that is as simple and as fruitful as possible for legal practice.

In philosophy, more has been written about probability than about plausibility, and also in more precise terms.[4] Therefore, it seems useful to first discuss which interpretations of probability are distinguished (section 2) and which of these is central in law (section 3). I then examine the meaning of plausibility (section 4). I end with a conclusion (section 5). In Section 4, I will argue that plausibility judgements are a specific type of chance judgements with the main characteristic that the person making a plausibility judgement not only gives a chance estimate, but also expresses that she has little confidence in her estimate.

If my analysis also applies to Dutch law, then plausibility standards differ from the two previously mentioned standards of proof of 'preponderance of probabilities'

* Anne Ruth Mackor is a professor of professional ethics, in particular of legal professions, at the Faculty of Law of the University of Groningen and editor of *Rechtsgeleerd Magazijn Themis*. The research for this article is part of her research project Preventing Miscarriages of Justice, NWO open competition, grant no. 406.21.RB.004. For information on the project, see preventingmiscarriagesofjustice.wordpress.com/.
This paper is a translation of a Dutch article that appeared in *RMThemis* 2023-5, pp 291-299, in a special issue on the plausibility standards in law.

1. I thank Christian Dahlman, Hylke Jellema, Henry Prakken and editors of *Rechtsgeleerd Magazijn Themis* for their comments on earlier versions of this article.
2. There appear to be large differences between people's interpretations of probability terms. See, for example, S.J.W. Willems, C.J. Albers & I. Smeets, 'Variability in the interpretation of probability phrases used in Dutch news articles - a risk for miscommunication', *JCOM* 19 (2020) 02, A03, doi.org/10.22323/2.19020203. Incidentally, the authors discuss probable and possible, but not plausible. Respondents estimated probable as a 67% probability and possible as a 47% probability on average, but individual estimates vary widely. See p. 22.
3. Rudolf Carnap, 'The Philosopher Replies', in: Paul Arthur Schilpp (ed.), *The Philosophy of Rudolf Carnap*, LaSalle, IL: Open Court 1963, pp. 859-1013, at pp. 936-937. I borrow the reference to Carnap's article from Hannes Leitgeb & André Carus, 'Rudolf Carnap', Supplement, D Methodology, *The Stanford Encyclopedia of Philosophy* (Summer 2023 Edition), Edward N. Zalta & Uri Nodelman (ed.), plato.stanford.edu/archives/sum2023/entries/carnap/.
4. For example, the Stanford Encyclopedia of Philosophy has no lemma on plausibility but has several lemmas on probability. See, for example, Alan Hájek, 'Interpretations of Probability', *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), Edward N. Zalta (ed.), plato.stanford.edu/archives/fall2019/entries/probability-interpret/. Another important difference is that far more formal analyses of probability than of plausibility have appeared. See, however, Joseph Y. Halpern, *Reasoning about uncertainty*, Cambridge: MIT Press 2003 and John R. Welch, *A Plea for Plausibility. Toward a Comparative Decision Theory*, New Yorkand London: Routledge 2023 for partly formal analyses of plausibility. Both argue for considering plausibility as a more fundamental and comprehensive concept than probability. Also see John R. Josephson, 'Appendix B Plausibility', in: John R. Josephson & Susan G. Josephson (ed.), *Abductive inference. Computation, philosophy, technology*, Cambridge: Cambridge University Press 1996, pp. 266- 272, especially p. 268.

and ‘beyond a reasonable doubt' by the relatively low requirement they place on confidence in the chance estimate. My impression is that the reverse is true for the standard of 'beyond a reasonable doubt': that standard requires not only that the chance estimate is very high, but also that confidence in the chance estimate is high.[5]

2. Three interpretations of probability

I have just pointed out that more has been written about probability than about plausibility. I therefore begin with an analysis of probability on the assumption that this analysis can be helpful in clarifying the concept of plausibility. Five interpretations of probability are usually distinguished: the classical interpretation, the logical interpretation, the subjective interpretation, the frequency interpretation and the disposition interpretation.[6] I will briefly discuss the three that are best known: the classical, the frequency and the subjective interpretation. I will discuss their meaning, as well as the question how probability can be determined according to the respective view.

2.1. Classical interpretation

Classical interpretations of probability are often used in games of chance. A characteristic of games of chance is that there is a finite number of outcomes and that the probabilities of the different outcomes, such as heads or tails, a 1 or a 6, are equal. When tossing a coin, there are two options, so the probability of both is 1/2. With a die, there are six options, so the probability of each outcome is 1/6. Formulated more generally and more abstractly: since it is assumed that the probability of events is equally distributed over all *possible* outcomes, the probability of a single event is the fraction of the total number of possible outcomes. The probability of an outcome is thus determined from the total number of possible outcomes.

2.2. Frequency interpretation

According to the second interpretation, probability is not - asin the classical interpretation - the number of possible outcomes, but the number of *actual* outcomes. Probability is thus a property of ‘real-world’ events, namely the frequency with which they occur. According to this interpretation, the probability of an event cannot be assumed a priori, but must be investigated empirically. For example, if we want to know the probability of a die falling on 6, we need to examine the die. One way to do that is to roll it often, say 6000 times. If the die falls on 6 about 750 times (1/8), then we not only know that it is impure, but then we will also know the frequency at which it falls on 6.

2.3. Subjective interpretation

According to the third interpretation, probability is not a frequency or some other property[7] of reality 'outside us', but a property of beliefs about reality 'inside us'. Probability expresses the degree of uncertainty of our beliefs. The degrees can range from the belief that something is impossible and therefore definitely not the case to the belief that something is certainly the case. In the middle lies the belief that something is as likely as unlikely.

This interpretation is called subjective or personal. However, an estimate of probability can be subjective in two different senses. The first is the sense just mentioned that probability does not exist outside of our minds, but in our minds as a property of our beliefs. The second meaning of subjective is arbitrary or irrational. Taste is subjective in this second sense: it is often said that there is no arguing about taste. The belief that the probability of the impure die that was discussed in section 2.1 falling on 6 is 1/10 isboth subjective in the first sense and subjective in the sense of arbitrary and irrational if we know that thefrequency actually is 1/8. A belief is not subjective in the sense of arbitrary if it is based as much as possible on reliable information, e.g., on information about frequencies. The estimate that the probability is 1/8 that the die falls on 6 is rational or at least reasonable and open to discussion because and insofar as it is empirically informed.

To distinguish rational or reasonable estimates from irrational or unreasonable estimates, rational or reasonable estimates are sometimes referred to as epistemically informed probability estimates.[8]

3. Subjective interpretation in law: Bayes

Which interpretations can be used in law? Classical interpretations seem inappropriate because court cases relate to events in the 'real' world. We do notknow in advance how many and which options are possible, nor whether the options we can foresee are equally likely.

Interpretations in terms of frequencies also do not seem suitable. A first problem is that we usually do not know how many outcomes there actually are.[9] A second problem is that while the frequency approach tells us something about patterns in series of events, it does not tell us anything about individual cases. An example: the frequency approach does tell us somethingabout the frequency with which a die falls on 6, but not about the probability of that same die falling on 6, on say,the 81st or 376th throw.

---

5. Unfortunately, this - in my view essential - point is rarely explicitly stated, even when the standard of 'beyond reasonable doubt' is quantified as a requirement of, say, a minimum of 90% or 95%.
6. Hájek 2019.
7. A disposition, one of the five interpretations of probability, is also a property of phenomena in reality.
8. In this context, also compare the classical legal distinction between the 'conviction raisonnée' (a conviction that is subjective only in thefirst sense) and the 'conviction intime' (a conviction that can, in addition, be subjective in the second sense).
9. However, see Anne Ruth Mackor, Veroordelen met ‘naakte’ statistieken? Editorial comment', *RMThemis* 2019, vol. 3, pp. 93-96 for a criminal case where it is practically certain that there are only two options.

A third problem is that, except for some forensic evidence, 'objective' i.e. scientifically based frequencies are usually not available in court cases. For subjective theory, the lack of 'objective' frequencies is not a problem now that probabilities are understood as degrees of beliefs. If our beliefs can be based on 'objective' frequencies, that is a good thing, but in their absence, they can also be based on frequencies estimated on the basis of general non-scientific background knowledge or personal experiences.[10] A second advantage of the subjective approach is that it is possible to talk about probabilities in individual cases. Therefore, the subjective approach is central in law. This ideally involves epistemically informed probabilities.

There are several subjective theories. The Bayesian theory, named after Thomas Bayes, who lived from 1701 to 1761, is the best known of these. This theory is dominant in law. For example, Bayesian probability theory is used in NFI reports.[11] At the heart of Bayesian theory is Bayes' rule, which can be used to calculate probabilities. In Bayes' rule, three different probabilities play a role: the prior, the likelihood and the posterior. It would take us too far to discuss Bayes' rule here. Suffice it to give a simple example to clarify the distinction and relationship between these three probabilities.

3.1. Posterior

In a drunk driving case, the most important piece of evidence is a positive alcohol test result.[12] The judge must determine how likely it is that the driver was under the influence given the positive test result. This probability is what Bayesians call the posterior. The question is how Bayesians determine this probability. A first important question is how reliable the alcohol test is. If the test is 100% reliable, it will give a positive result in all cases where the driver has been drinking and a negative result in all cases where the driver has not been drinking. We could then confidently conclude that we know with certainty that the driver was under the influence. In reality, tests, like other evidence, are never 100 per cent reliable. The reliability of tests can sometimes be examined empirically. Just as we can examine the frequencies with which dice fall on different sides, alcohol tests can be examined to find out how often test results contain errors.

By way of example, we assume that in 1% of cases the alcohol test gives a **false positive result** (the test result that driver *has been* drinking whereas in fact he has *not*) and in 10% of cases a **false negative result** (the test result that the driver has *not been* drinking while he *has* in fact been drinking). With this information, can we now answer the question whether the driver was under the influence of alcohol. Many people will be inclined to conclude that there is a 99% chance (= 100% - 1% false negative) that the driver in question has been drinking. Conversely, with a negative test result, many will estimate that the probability that the driver did not drink is 90% (100% - 10% false negatives). Both ways of reasoning are incorrect. They rest, according to Bayesian probability theory, on two fallacies.

3.2. Likelihood

The first error is that the test does not state that it follows from the test result that the probability of the driver having drunk is 99%. The test gives the opposite conditional probability: it says that if the driver did drink, the probability of a true-positive test result is 90% and it also says that if the driver did not drink, the probability of a false-positive test result is 1%. Unfortunately, in court cases both probabilities (the probability of A given B versus the probability of B given A) are so often mixed up that this error has received its own name: the prosecutor's fallacy.[13] Intuitively, it is not immediately obvious that these are different probabilities and even less so that the two probabilities can be very different. Therefore, I will first give an example. The probability that an animal has four legs if it is a cow is very high. The inverse probability that an animal is a cow if it has four legs is very small. The example makes it clear that it matters quite a bit which of the two questions must be answered.

We have seen that the probability we look for in court cases is called the posterior. More precisely, it is the posterior probability ratio or odds ratio. The posterior ratio is the ratio between the probability that the alleged fact ***did*** occur **when the evidence** is found and the probability that the alleged fact ***did not*** happen when

10. In these cases, there is a risk that the estimation is not only subjective in the first sense (a property of a belief), but also in the second sense (arbitrary or unreasonable). Daniel Kahneman, among others, has argued that personal judgements are often 'biased'. Availability bias, representation bias and anchor bias are some of the biases that can play a role. See Amos Tversky & Daniel Kahneman, 'Judgment under Uncertainty: Heuristics and Biases', *Science*, New Series, Vol. 185, No. 4157 (Sep. 27, 1974), pp. 1124-1131. See Daniel Kahneman, *Thinking, Fast and Slow*, New York, Farrar, Straus and Giroux, 2011 for an accessible discussion of these biases.
11. See NFI, *Vakbijlage waarschijnlijkheidstermen*, https://www.forensischinstituut.nl/over-het-nfi/publicaties/publicaties/2017/10/18/vakbijlage-waarschijnlijkheidstermen
12. I take this example from a teaching module developed by Christian Dahlman as part of our research project Preventing Miscarriages of Justice. See star note.
13. W.C. Thompson & E.L. Schumann, 'Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy', *Law and Human Behavior* (11) 1987, vol. 3, pp. 167-187. Also see H. Prakken, 'Kansoordelen door deskundigen over "logisch" rapporteren en wat daarbij mis kan gaan', *AA* (67) 2018, vol. 9, pp. 740-747. Prakken examined 31 court decisions. In 22 of them, he found the prosecutor's fallacy.

the evidence is found.[14] The inverse probability is called the likelihood; the inverse probability ratio is called the likelihood ratio.[15] The likelihood ratio is the ratio between the probability of a true-positive (finding the **evidence if the alleged fact *did*** occur) and the probability of a false-positive (finding the **evidence if the alleged fact *did not*** occur).[16] The posterior and likelihood are thus mirrored conditional probabilities.

Back to the result of the alcohol test. The likelihood ratio of the test is the ratio between the probability of a true-positive result, i.e. if the driver has been drinking, in this example 90% (100% - 10% false negatives), and the probability of a false positive result, if the driver has not been drinking, in this example 1% (100% - 99% true positives). The likelihood ratio is therefore 90:1 = 90.

3.3. Prior

With the likelihood ratio, do we now have the necessary information to estimate the probability that the driver who tested positive has been drinking? Again, the answer is negative. The second mistake many people make is that they do not realize that they still do not have enough information to determine the posterior. The information that is missing is the initial probability estimate we make before we have the test result.[17] This probability is called the prior.[18]

To illustrate the relevance of the prior, I will again use the example of cows and animals with four legs. We have already seen that the probability is very high that an animal has four legs if it is a cow, while the inverse probability that an animal with four legs is a cow is very low. This is partly explained by the fact that, prior to the concrete evidence we have, we already know (or believe) that the total number of cows in the world (the prior probability of finding a cow) is much smaller than the total number of animals that have four legs (the prior probability of finding an animal with four legs).

Back to the alcohol test example. Let's assume that frequencies have been established and that we estimate that 1% of tested drivers actually drank alcohol.[19] Suppose 1000 drivers are tested. We know that 1% of them, i.e., 10 drivers, have been drinking. If we combine this information with the evidence of the positive test result, we get the following two outcomes. These outcomes are probably very surprising to many people.[20]

First, we know that of the 99%, i.e., 990, drivers who have not been drinking, 1% get a false-positive test result. That is 9.9 - rounded 10 - people. On the other hand, we also know that 10% of the 10 drunk drivers, get a false-negative test result. So, 1 driver gets a false-negative result. The first surprising result is that the probability of a false-positive result (10) is much higher than the probability of a false-negative result (1) and even higher than the probability of a true positive result (9). This outcome is partly due to the low prior which makes the number of true-positives very low. I will explain this below.

The second surprising result is that at a 1% prior, the posterior probability that a driver testing positive has been drinking is not 99%, but only 48%.[21] This is related to the fact that out of 1,000 drivers, more drivers get a false-positive (10) than a true-positive (9) test result. Would the prior not have been 1% but for example 10%, we get a different posterior. With a prior of 10%, not a mere 10, but 100 out of 1000 drivers have been drinking and so 900 are sober. The false-positives do not change muchas a result, but the true-positives do. Of the 900 sober drivers, 9 (1%) get a false-positive. Of the 100 drunk drivers, 10 (10%) get a false-negative result and 90 get a true-positive result. The difference between a prior of 1% and 10% is that not a mere 9, but 90 drivers get a true-positive result. The posterior is therefore not 48%, but 91%.[22]

I summarize. The first insight that Bayesian probability theory gives us is that the probability that someone has been drinking when they test positive (the posterior probability) depends not only on the ratio of true-positive to false-positive test results (the likelihood ratio), but also on the probability estimate prior to the test result (the prior).

The second insight is that the prior and likelihood can sometimes be estimated on the basis of scientific research, but much more often they are based on non-scientifically verified general background knowledge or on personal experiences which increases the risk of subjectivity in the sense of arbitrariness that we

14. This probability is symbolically represented as p (H|E) : p (-H|E). Here, p represents the probability, E represents the evidence and H represents the fact to has to be proven and that is in that sense still hypothetical.
15. It is also referred to as the diagnostic value of evidence.
16. In symbols: p (E|H) : p (E|-H).
17. The question of whether the use of the prior is compatible with the presumption of innocence has led to much debate. I will not discuss this question here.
18. This probability is represented as p (H).
19. As with the likelihood ratio, the prior can sometimes be based on empirical research. Much more often, however, the determination of the prior is also an estimate based on general background knowledge or personal experience. Even if estimates of the prior and likelihood are scientifically based, they are still subject to much debate. If we want to estimate the prior in alcohol testing, we cannot 'just' examine 'the' frequencies. Among other things, we must decide whether we want to estimate that probability for the whole of the Netherlands, or for a certain day around a certain time (e.g., Saturday night or Tuesday morning), in a certain region, or even on a certain street, for a certain type of driver, etc.
20. The probability that the individual driver has been drinking is determined by using frequency information. This does not mean thatI am secretly using the frequency interpretation of probability. Frequencies are not the *meaning* of probability here; they are only the information on which the subjective belief is *based*.
21. Since I am not discussing Bayes' rule, I omit the explanation. The posterior probability at a prior of 1% is 9 true positives divided by (9 true positives + 9.9 false positives) = 9 : 18.9 = 0.476 = 48%.
22. With a prior of 10%, the posterior probability is 90 true positives divided by (90 true positives + 9 false positives) = 90 : 99 = 0.909 = 91%.

discussed earlier (see section 2.3).

4. Plausibility

I have discussed three interpretations of probability (classical, frequency and subjective interpretation) and subsequently the three probabilities that are central in Bayesian probability theory (prior, likelihood and posterior). The term plausibility plays no role in Bayesian probability theory; it only uses the term probability.[23] The question, therefore, is how the preceding account of probability can help clarify the meaning of plausibility. The analysis in this section starts from the assumption that plausibility and probability are both used to expresschances in the sense of degrees of beliefs.[24]

In this section, I discuss four characteristics of plausibility judgements that show that plausibility judgements are a specific type of chance judgements. I do not claim that my discussion is exhaustive, however. I only aim to clarify some aspects relevantto legal practice.[25] Nor does my analysis imply thatplausibility judgements and probability judgements differ in principle[26] or that plausibility judgements cannot be formulated as probability judgements.[27]

In section 4.1, I discuss the most important feature of plausibility judgements, namely that they do not only express a chance estimate, but also limited confidence in that same chance estimate. Section 4.2 discusses a first reason for limited confidence: the imprecise nature of plausibility judgements. In section 4.3, I examine a second reason for limited confidence: the lack of a sufficient amount of concrete evidence. Finally, in section 4.4, Idiscuss various functions that judgements of acceptability have on the one hand in the investigation phase and on the other hand in the final phase of judgment and decision.

4.1. A double and uncertain chance estimate

The first and most important characteristic of plausibility judgements concerns the nature of the belief. If we express our belief in terms of plausibility, we are willing to offer a chance estimate. After all, otherwise we would refrain from judging and say 'no idea'.

With the term plausibility, however, we do not only express a chance estimate, but also that we are not,or not yet,[28] very confident about our estimate.[29] In other words, a plausibility judgement does not express one chance estimate, but combines two chance estimates. The first estimate, that the alleged fact occurred, can range from high to low, from very plausibleto very implausible. The second estimate is the more or less constant relatively low estimate of confidence in the first estimate.

Low confidence in chance estimates can have different reasons. It may be due to the fact that the estimate is not made by using Bayes' rule, but by using rather imprecise criteria. This explanation is discussed in section 4.2. Low confidence may additionally be due tothe fact that the chance estimate is only based on a limited amount of evidence and is therefore not robust. This explanation is discussed in section 4.3.

4.2. Inaccurate chance estimates

4.2.1. Qualitative estimates

Plausibility judgements are only expressed in qualitative terms. We can formulate plausibility judgements in different ways. We can rank our beliefs, from least plausible to most plausible, and we can indicate their position in the ranking with numbers (1st, 2nd, 3rd, etc.). We can also make a categorical judgement and say, for example, that a stated fact is somewhat plausible or, on the contrary, very implausible. In the literature, however, it is argued that we cannot give a numerical weight to individual plausibility judgements.[30]

---

23. In the aforementioned subject annex, the NFI translates the term likelihood as 'aannemelijkheid', i.e., plausibility. This is confusing. Plausibility in the sense that is central to this paper is not likelihood. We have just seen that prior, likelihood and posterior are three different chances that we need to keep well apart. The term plausibility can be used for all three chances. In a sense, likelihood is at odds with plausibility in that likelihood refers mainly to concrete evidence; after all, it is the probability of finding the evidence if the alleged fact occurred. Plausibility, on the other hand, is often used in situations where little concrete evidence is (yet) available. More on this in section 4.3.
24. Plausibility is used exclusively in the subjective sense. We have seen in Section 2 that probability also has other interpretations, including the classical and frequency interpretation.
25. However, I leave aside the well-known 'relative plausibility theory'. According to this theory, evidentiary judgements are made bycomparing the relative plausibility of the parties' statements. See Ron J. Allen & Michael S. Pardo, 'Relative plausibility and its critics',*The International Journal of Evidence G Proof* 2019, pp. 5-59. The reason for excluding this theory is that Allen and Pardo do not provide an analysis of the meaning of plausibility and only argue that plausibility is determined by qualitative criteria. More on qualitative criteria in section 4.2.2.
26. However, this thesis is not uncontroversial. See, for example, Douglas Walton, *Abductive Reasoning*, Tuscaloosa: University Alabama Press 2004, Chapter 1 Abductive, Presumptive, and Plausible Arguments, pp. 1-50, p. 28 and Josephson 1996, pp. 268-271. See also Welch 2023 who argues that plausibility is a more fundamental concept than probability.
27. In terms of Carnap's requirements for explication (see the introduction), the question is whether the precision that can be gained by doing so always outweighs the increase in complexity. In this contribution, I will not discuss possible translations of plausibility in terms of probability.
28. See section 4.4.1.
29. In my earlier editorial comment on plausibility, I used the term 'credence' for the chance estimate and 'confidence' to express confidence in that estimate. Here I use the term 'belief' instead of 'credence'. A.R. Mackor, 'Evidence lessons from the surcharge affair: when is a fact "plausible"?', *RMThemis* 2022, vol. 2, pp. 37-40.

For example, we cannot say that we find something 60% plausible. This seems to be related tothe fact discussed in section 4.1 that a plausibility judgment expresses not a single but a double chance estimate.

Unlike plausibility judgements, probability judgements can be formulated both in qualitative and in quantitative terms. As with plausibility judgements, we can make both categorical probability judgements and rank them. The difference is that we cannot only express the ranking, but also the categorical probability judgement in quantitative terms. In Bayesian probability theory, the scale runs from 0 (certainly not the case, impossible) to 1 (certain). Probability judgements can therefore be expressed as 0.11 or 11% (very improbable), 0.5 or 50% (it is just as likely that the fact occurred as that it did not occur), 0.86 or 86% (very likely), 0.88 or 88% (almost certain), etc.[31]

4.2.2. Qualitative criteria

In Section 3, I indicated that we can make quantitative chance estimates in terms of probability using Bayes'rule. Plausibility estimates are not made by applying Bayes' rule. They are usually estimated intuitively usingrather imprecise qualitative criteria. The use of qualitative criteria seems to explain why plausibilityjudgements are only formulated in qualitative terms. Theuse of imprecise criteria seems to explain why we donot use the term probability, but plausibility: because the estimate is imprecise, we have relatively low confidence in our chance judgements.

Pennington and Hastie's Story Model is the best-known theory on the use of qualitative criteria in law. In the 1980s and 1990s, Pennington and Hastie empirically investigated how jurors arrive at an evidentiary verdict. They found that they do so, by creating a causal story to explain the evidence and then judge that story using qualitative criteria. According to Pennington and Hastie, jurors use three criteria in their assessment. The first assessment criterion is 'coverage'. It refers to whether the story can explain the evidence. The second criterion is 'coherence'. This criterion consists of three sub-criteria: the consistency of the story, the completeness of the story and the acceptability of the story in the light of the background knowledge of the juror in question. The third criterion is 'uniqueness'. If there are several stories that meet the first two criteria, this will reduce the level of confidence in all of them. [32,33]

I have two comments on the Story Model. First, we see that plausibility is only one of the criteria by which confidence in and acceptability of a story is assessed. Pennington and Hastie use plausibility in the specific sense of the chance that a story is true in the light of general background knowledge. This estimate seems to correspond with the estimation of the prior we discussed in section 3.3. According to other theories, plausibility hasa broader meaning. On these views plausibility is estimated not only on the basis of background knowledge, but also on the basis ofthe other criteria of coverage, coherence and completeness. On this interpretation, plausibility is not a sub-criterion but the overarching criterion.[34]

A second comment on the Story Model is that Pennington and Hastie do not mention the term probability at all; they only speak of plausibility and the confidence in and the acceptability of the story. However, this does not alter the fact that a juror will ultimately have to determine whether she considers the alleged facts proven according to the appropriate standard of proof, for instance 'beyond a reasonable doubt' or with a 'preponderance of probabilities'. Based on the 'coverage', 'coherence' and 'uniqueness' of the story, we can - in my opinion - not only assess the confidence in and the acceptance of a story, but also estimate, in terms of plausibility, the posterior probability that was discussed in section 3.1.

Thus, the use of imprecise qualitative criteria seems to be a first explanation of the fact that plausibility judgements express a relatively low confidence in the chance estimate. In the next section, we discussanother explanation for the low confidence in the estimate.

4.3. Few robust chance estimates

4.3.1. Lack of concrete evidence

In the previous section, I argued that the use of rather imprecise qualitative criteria may be grounds for making a plausibility judgment rather than a probability judgment. A lack of 'direct' concrete evidence is a second reason to make judgements in terms of plausibility.[35]

---

30. Doug Lombardi, E. Michael Nussbaum & Gale M. Sinatra, 'Plausibility judgements in conceptual change and epistemic cognition', *Educational Psychologist* (51) 2016, issue 1, pp. 35-56, at p. 37. They compare plausibility not only with probability, but also with other related terms such as possibility and coherence, see pp. 36-39.
31. However, see Willems, Albers & Smeets 2020 on the widely varying ways in which people quantify different terms.
32. Nancy Pennington & Reid Hastie, 'Reasoning in explanation-based decision making', *Cognition* 49 (1993), pp. 123-163.
33. In particular in the philosophy of science, much literature has appeared on the qualitative criteria (virtues) by which scientific theories and hypotheses are judged. There is no space here to go into them. See, for example, Paul R. Thagard, 'The Best Explanation: Criteria for Theory Choice', *The Journal of Philosophy*, Vol. 75, No. 2 (Feb. 1978), pp. 76-92 and Frank Cabrera, 'Can there be a Bayesian explanationism? On the prospects of a productive partnership', *Synthesis* (194) 2017, pp. 1245-1272, doi.org/10.1007/s11229-015- 0990-z.
34. Similarly, Allen and Pardo 2019.

This is in line with Dutch legal literature on the plausibility standard in which a lack of concrete evidence is also mentioned as a distinguishing feature of plausibility judgements.[36]

We have seen that the main distinguishing feature of plausibility estimates is that, in addition to a chance estimate, they also express a lack of confidence in the chance estimate. This raises the question of what concrete evidence contributes to our confidence in a chance estimate.

Concrete evidence does two different things. First, in discussing the likelihood ratio in section 3.2, we have seen that evidence can increase or decrease our estimate.[37] A positive alcohol test result increases our estimate of the probability that the driver had been drinking. A negative result lowers that estimate.

A second consequence of the increase in concrete evidence is that our confidence in, and hence acceptance of, the chance estimate increases. First, after carefully collecting as much evidence as possible, we will lower our estimate of the chance of discovering further new evidence. Secondly, we will also offer a lower estimate of the chance that any new evidence will drastically change our chance estimate. Both result in increasing confidence that our chance estimate will become more robust in the sense of more stable.[38]

First, if there are several pieces of evidence of good quality that all point in the same direction, this leads us to estimate the chance that something has happened or will happen ever higher (or lower). Second, it leads to us becoming more confident in our estimate. Evidence pointing in different directions only has the second effect. Because one piece of evidence increases the chance but the other decreases it, they do not increase the chance estimate, or at least not much. They do increase robustness and hence confidence in the chance estimate. If good investigations have been done and there is sufficient concrete evidence making confidence sufficiently high, then it seems natural to formulate the chance estimate no longer in terms of plausibility, but in terms of probability.

4.3.2. Hypotheses

The above leads to the question of what plausibility judgements are based on when there is little concrete evidence. In short, such judgements rely on general background knowledge and speculation. In section 4.2.2, we introduced Pennington and Hastie's Story Model. They argue that jurors construct a causal story that can explain the available evidence. Causal stories consist of different hypotheses that are temporally and causally connected.

An example. Suppose I walk into the living room, and I see the vase that was on the table five minutes ago now lying on the floor. The first question that comes to mind is: why is the vase on the floor? In other words, what caused the vase to be on the floor? A gust of wind and the neighbour's cat are the first plausible explanations that come to mind; whereas I would immediately reject the explanation that an elephant or a bluebottle was the cause as utterly implausible (assuming that such implausible explanations would come to my mind in the first place).

The plausibility of the various explanations is partly determined by the estimation of the prior probability of wind gusts, cats, elephants, and bluebottles. However, these probability estimates are not just about the probability of the presence of wind gusts and the species mentioned. It is also about answering whether and, if so, how well the hypothesis can causally explain the evidence. Elephants fall off because I estimate the chance of their presence negligible; bluebottles fall off because I consider the probability of their being able to knock over a vase to be nil.[39] As long as I have no concrete evidence (yet) about the cause of the fallen vase, I will judge that gusts of wind and cats are plausible explanations in the light of my background knowledge, wheras elephants and bluebottlesare not.

35. This situation occurs, among other things, when estimating the chance of future events. In 'future studies', a distinction is made between 'forecasting' and 'foresight'. Forecasting' involves predicting fairly specific events, often in the short term, in terms of probability. Foresight' concerns the exploration of broader and more long-term possibilities in terms of plausibility. Legal cases also regularly involve the estimation of future events, including when assessing danger and risk of recidivism.
36. See R.H. de Bock, *Tussen waarheid en onzekerheid: over het vaststellen van feiten in de civiele procedure* (PhD thesis, Tilburg), Deventer: Kluwer 2011, pp. 221- 225.
37. Reliable evidence such as a reliable test or witness statement increases the probability more than unreliable evidence such as an unreliable test or witness statement. Evidence from different sources such as two witnesses testifying independently also affects the probability more than evidence that is not independent, such as statements by witnesses who have talked to each other about what they saw.
38. On robustness, see, for example, Dale Nance, *The Burdens of Proof: Discriminatory Power, Weight of Evidence, and Tenacity of Belief*, Cambridge: Cambridge University Press 2016; Christian Dahlman & Anders Nordgaard, 'Information economics in the criminal standard of proof', *Law, Probability and Risk* 2023, p. 1-26, doi.org/10.1093/lpr/mgad004 and Hylke Jellema, 'Reasonable Doubt, Robust Evidential Probability and the Unknown', *Criminal Law and Philosophy* June 2023, doi.org/10.1007/s11572-023-09685-5.
39. Plausibility judgements play a crucial role in forms of reasoning called abduction and Inference to the Best Explanation (IBE). P. Lipton, Inference *to the best explanation*, London: Routledge 2004 (2nd ed.) is a classic on IBE. Also see Josephson 1996; Walton 2004and D. Walton, C.W. Tindale & T.F. Gordon, 'Applying recent argumentation methods to some ancient examples of plausible reasoning', *Argumentation* (28), 2014, vol. 1, pp. 85-119. On the role of causal explanations in law, see D. Lagnado, *Explaining the Evidence*, Cambridge: Cambridge University Press 2021 and A.R. Mackor, H. Jellema & P.J. van Koppen, 'Explanation-Based Approaches to Reasoning about Evidence and Proof in Criminal Trials', in: B. Brozek, J. Hage, & N. Vincent (ed.), *Law and Mind: A Survey of Law and the Cognitive Sciences*, Cambridge: Cambridge University Press 2021, pp. 431-470. See also A.R. Mackor & P.J. van Koppen, 'The scenario theory

When I find sufficiently concrete evidence for one (gust of wind) or another (cat) hypothesis, then I can also formulate my estimation as a probability judgement.This brings us to the fourth and final feature of plausibility judgements: their functions.

4.4. Functions of plausibility judgements

Plausibility judgements perform different functions in the investigative phase and the final phase of judgment and decision.[40]

Often, the term plausible is used to give an initial and preliminary assessment of what may have happened or what may be about to happen.

In some cases, we consider such a tentative assessment sufficient for a final judgement and decision. If I consider it plausible that it will rain (judgement), in most cases I will not investigate further, but take a foldable umbrella with me (decision). The inconvenience of a foldable umbrella is much less than the disadvantage of getting soaked and the inconvenience outweighs the time it takes to look at the weather forecast.

In many cases, however, an estimate of plausibility is insufficient to reach a final judgement and decision and further investigation is necessary.

4.4.1. Investigative phase

In the investigative phase, hypotheses and causal stories are still speculations. They answer the question of what *may have* happened or will happen, not the question of what *has* happened or will happen.[41] At this stage, plausibility judgements are still tentative or prima facie judgements, to which we do not yet want to (fully) commit ourselves.[42]

An important function of stories and hypotheses in the investigative phase is that they not only generate possible explanations for the already known evidence but, more importantly, that they also predict new evidence. If we investigate the hypothesis that the vase fell off the table because of a gust of wind, we predict other evidence than if we investigate the hypothesis that a cat knocked the vase over. Further investigation may subsequently lead to finding - or not finding - the predicted evidence. That new evidence may then lead to sharpening and improvement, but also to rejection of the initial hypotheses. Improved or new hypotheses can in turn lead to new predictions and subsequent discoveries of yet more new evidence that may or may not be discovered by further investigations, etc. If we find new evidence then our chance estimate changes and confidence in the estimate also grows.

However, whether a hypothesis is good enough to guide investigations cannot be determined by estimating the likelihood of the hypothesis. This is related to the fact that in the investigative phase, precise and sometimes bold hypotheses are usually formulated. They have, for the time being, viz., in the absence of concrete evidence a much lower probability than less precise and less bold predictions.[43] However, hypotheses must have the potential to become probable, namely after the discovery of the predicted evidence. This assessment can be made relatively quickly using the aforementioned qualitative criteria, including the plausibility criterion.

In the investigative phase, plausibility estimates thus have an exploratory and heuristic function. They play an important role in generating, selecting and improving hypotheses and stories.[44] However, the advantage of the speed of selection and improvement of stories using qualitative criteria is accompanied by the aforementioned risks of biases in chance estimates and of fallacies such as the prosecutor's fallacy.[45]

4.4.2. Final judgment and decision phase

In the final decision or judgment phase, plausibility judgements play a different role. They are no longer chance estimates of what *might* have been or what *might* be the case, but chance estimates of what *was, is, or will be the* case. At this stage, plausibility judgements no longer have a heuristic function. They only have the function discussed in section 4.1 that they express both a chance, and at the same time a limited confidence in that chance estimate.

5. Conclusion

In the previous sections, I described three different interpretations of probability and discussed the role of Bayesian probability theory in law. The analysis of plausibility started from the idea that probability and plausibility judgements both express chance judgements and that both can be understood as subjective, but in principle epistemically informed degrees of belief. I then discussed four distinctive features of plausibility judgements.

The first and most important feature is that the person making a plausibility judgment not only makes a chance estimate, but also expresses that she hasrelatively low confidence in that chance estimate.

---

about Evidence in Criminal Law', in: A. Stein, C. Dahlman & G. Tuzet (ed.), *Philosophical Foundations of Evidence Law*, Oxford: Oxford University Press 2021, pp. 213-228.

40. In the terminology taken from the philosophy of science: plausibility has a different function in the 'context of discovery', than in the 'context of justification'.
41. Similarly, Josephson 1996, p. 271.
42. For a discussion of the tentative nature of plausibility estimates, see Nicholas Rescher, *Epistemology. An introduction to the Theory of Knowledge*, Albany, NY: SUNY Press 2003, Chapter 5, pp. 81-100. Also see Josephson 1996.
43. More on this in Cabrera 2017, especially section 3.3.
44. On the role of plausibility judgements in hypothesis generation, selection and improvement, see G. Klein et al, 'The Plausibility Gap: A Model of Sensemaking', Technical Report, DARPA Explainable AI Program, 2021, pp. 1-18, especially pp. 12-13. More generally on the psychological process of generating, selecting and improving hypotheses, see Gary A. Klein, *Sources of power*, Cambridge: The MIT Press2017 (20th ed.).
45. See note 10 on biases and section 3.2 on the prosecutor's fallacy.

One reason for low confidence is that plausibility judgements are imprecise. This is because they are judgements not made by using Bayes' rule, but by using rather imprecise qualitative criteria.

Another reason for low confidence is that plausibility judgements are often made in situations where there is (still) little, little reliable and/or little diverse concrete evidence available, making the estimates little robust.

Finally, the fourth feature concerns the functions of plausibility judgements. In the investigative phase, causal stories are generated, selected and improved. In this phase, they serve not only to explain evidence already known, but especially to predict new evidence. In this phase, plausibility judgements are not yet judgements about what actually happened; they are possibility judgements. In the investigative phase, it is not a problem that we do not yet have much confidence in our chance judgements. After all, they are only preliminary estimates by which we anticipate on expected, but not yet discovered, evidence that may not only increase the posterior chance estimate but also make the estimate more robust.

In contrast, in the final judgment and decision phase, the only distinctive function of plausibility judgement is that they express that we have relatively little confidence in the chance estimate. Unlike in the investigative phase, in the final judgment and decision phase, the lack of precision of, and lack of confidence in, our chance estimates can be a major problem, especially when far-reaching decisions are taken in cases in which there are major interests at stake.[46] A lack of confidence in the chance estimation is – as far as I am concerned – among others a major problem when it comes to the plausibility of danger as a condition for imposing preventive measures. The problem is all the more pressing because the aforementioned risk of biases and errors in thinking when making plausibility estimates also exists in the final judgment and decision phase.

All in all, it is time to examine how high the risks of biases and thinking errors actually are when we make plausibility judgments, and to discuss the question of what risks are considered acceptable in what type of decision.[47]

46. Another problem is that in some cases the question of evidence seems to be subordinated to a weighing of interests.
47. See Anne Ruth Mackor, 'Risks of Incorrect Use of Probabilities in Court and What to Do about Them', in: Adriana Placani & Stearns Broadhead (ed.), *Risk and Responsibility in Context*, New York and London: Routledge 2023, pp. 94-108.