



HAL
open science

Expert Advice for Decision Making: The Subtle Boundary Between Informing and Prescribing

Marion Vorms

► **To cite this version:**

Marion Vorms. Expert Advice for Decision Making: The Subtle Boundary Between Informing and Prescribing. Epistemic Paternalism Conceptions, Justifications and Implications, Rowman & Littlefield Publishers, 2020. hal-03251754

HAL Id: hal-03251754

<https://hal.science/hal-03251754v1>

Submitted on 14 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Expert Advice for Decision-Making: The Subtle Boundary Between Informing and Prescribing

Marion Vorms

Published in G. Axtell and A. Bernal (eds), *Epistemic Paternalism, Conceptions, Justifications and Implications*, 2020, Rowman & Littlefield Publishers, chapter 3

A difficulty of experts' task in advising decision-makers is to provide the best and most relevant information, without infringing on their recipient's decision. The boundary between informing and prescribing is a subtle one, especially when the science at stake is complex and pervaded with uncertainty. As studied by the philosophical literature on 'inductive risk' and the role of values in science, what may appear as purely informative is often imprinted with value judgements. I will however show that there is another way experts are likely to exceed their duty, which is not reducible to imprinting their own values upon their audience. This other threat of undue paternalism lurking on experts' function is more specifically epistemic, and potentially harder to deal with, as it is more surreptitious and likely to go unnoticed, including to the experts themselves. It is related to the very structure of inference networks in complex evidential reasoning, which is often overlooked in the literature on scientific confirmation.

After highlighting some central issues arising from consideration of the requirement for experts reports to be 'value-free', I will draw from analyses of evidential reasoning in criminal inquiry and the courtroom, as well as from instructions given to forensic experts in England and Wales, so as to shed light on that other, subtler difficulty experts must overcome in reporting their results. This consists in fine-tuning between inferences that are actually within the scope of their competences, and which they can therefore legitimately endorse, and those which exceed such competences. As we will see, such fine-tuning is made particularly tricky by the fact that, in complex decisional tasks, expert and common knowledge are often intertwined in their contribution to evaluating the inferential bearing of the available evidence for a given issue.

Epistemic Deference and Inductive Risk

The Role of Expert Advice in Decision-Making: Informing Without Prescribing

Many of our decisions, either at the individual or group level, have to be grounded on scientific knowledge that is beyond our reach. Even though one may feel reluctant to consider beliefs acquired through the words of others as fully justified — hence to count them as genuine knowledge —, when it comes to making practical decisions, especially those that are highly consequential, we have no choice but to defer to experts' advice. Patients having to decide whether or not to undergo a surgery, jurors in criminal trials in charge of deciding whether to convict or not, as well as policy-makers having to take action regarding greenhouse gases emission regulations, need to rely on (medical, forensic, climate) experts to make the best-informed decisions.

Decision, however, is supposed to be the subject's own. A commonly shared ideal of individual freedom relies on a *prima facie* clear-cut distinction between informing and prescribing: experts must guide the subjects' decision by providing them with the best available knowledge regarding their case, but the resulting decision should depend on the subjects' own preferences. At the societal level, the democratic ideal dictates that the values guiding policies should be the people's own, rather than the experts'. We expect experts to neutrally tell us what it is rational to believe, not to prescribe our actions.ⁱ

Inductive Risk as a Challenge to the Neutrality of Expert Advice

As clear-cut as it may seem, such boundary between informing and prescribing is blurred by what Hempel (1965) first coined 'inductive risk': since empirical confirmation of any claim is doomed to be relative, accepting, as well as rejecting it, involves a risk of error. Which risk is worth taking cannot be objectively determined on empirical grounds. Some decision has to be made, based on an appraisal of the relative costs of accepting the claim if it is wrong (false positive), and of rejecting it if it is right (false negative). In other words, some consequence-

dependent confirmation threshold has to be set, beyond which a claim is considered warranted enough to be accepted, and publicized.ⁱⁱ But, as Rudner (1953) had already argued, appraisal of the relative costs and benefits of the various outcomes (accepting the claim if it is true / false / rejecting it if it is false / true) depends on values.

One could argue that, in scientific research, such values are shared by the whole community, and result in very high, and subject-independent, standards of proof (think of the *p*-value, as well as all institutional practices governing scientific publication). Those warrant that no claim can be scientifically published unless maximally confirmed, thus somehow saving the ‘value-free ideal of science’, according to which scientific claims should be justified on the basis of epistemic criteria only (Betz 2013, John 2015b, Kitcher 2011).ⁱⁱⁱ But, as soon as we are dealing with practically consequential claims, and even more so with policy-advising and decision-informing contexts, the question arises as to whether such an ideal can, and should be saved.^{iv} Indeed, in such cases, scientists’ decision to assert something will ground their recipient’s decision to take such or such consequential action. Douglas (2000, 2009) argues that scientists are morally obliged to take into account the practical consequences of their claims — hence, for issues of public interest (such as, for example, the toxicity of a given substance), they should factor in the expected costs and benefits of a policy based on each type of possible error (false positive and false negative) in setting their own standards of assertion. Those would therefore be ‘floating standards’ rather than fixed ones (John 2015a). Moreover, as Douglas has argued, such risk assessment does not only impact scientists’ final decision whether to publicize a claim, but their whole methodology: according to her, it is both ethically desirable, and practically unavoidable for scientists to make value judgements throughout the process of hypothesis testing (e.g., through test sensitivity setting, see Douglas 2000). From such a perspective, the distinction between informing and prescribing is far from clear: as scientific conclusions are

already shaped by the ensuing decisions, they must (both normatively and descriptively) contain prescriptions.

How to Escape Moral Paternalism?

The “floating standards” conception, as coined by John (2015a) raises several problems. As I endorse most of John’s criticisms, I will not recast them here, and will rather highlight a few further points, which are relevant for the rest of this paper.

First, unless assuming that she shares the same values as her intended audience, the scientist’s setting her standard of assertion according to her own appraisal of the relative costs of false positive and false negative would just be a way to impose her own values onto the decision-maker. This would be, so to speak, disguised moral paternalism under the pretense of legitimate epistemic authority.

One could therefore be tempted to prescribe that experts should take their audience’s values into account when setting their own standards of assertion.^v This however seems hardly workable. Even if we assume that experts have some secure access to such values (which is a strong, and doubtful assumption), it is by no means clear what they should do. Should they adopt the standards they assume their recipients would have adopted in their place? Consider a case for which the recipient’s values imply that false negatives would be much worse than false positives: scientists should adopt a low standard, and assert the corresponding hypothesis as soon as there is some tiny evidence in its favor. But what would the recipient’s own understanding of the scientist’s report consist in? If she is not aware that the scientist herself has already applied a low standard, she might understand her claim as ascertaining something that is in fact of very low probability. Hence, there might be some kind of ‘double counting’ of her own values, lowering even more her (actual) standard of acceptance. This is only one example of the pragmatic puzzles lurking in such situations. There is no room in this paper to explore the endless difficulties contained here, but this should be sufficient to claiming that the

idea of a standard tuned to the intended audience's expected interpretation is by no means trivial, to say the least.

What then should experts do in order to report their results so as to help the recipient's making the decision that corresponds best to her values, without presuming anything about those values? The most tempting answer may seem to lie in an attempt to state the results as precisely as possible, hence to report degrees of evidential support, and, where applicable, to be explicit about probabilities.^{vi} This again raises several problems, one of which being that, even to reach a probabilistic conclusion, one has to make outright decisions in the course of the inquiry.^{vii} Moreover, such a strategy is more akin to a way of protecting oneself by not committing too far, than to an attempt at offering full understanding and information to the recipient.^{viii} It might be the least bad solution in many situations, but it is far from being the panacea.

Even setting aside the question of values, and hence of the threshold that has to be reached to consider a claim acceptable, how a given claim (be it outright or probabilistic) grounds our decisions is far from trivial, and is not reducible to a cost-benefit calculus. As will appear in the following, the relevant distinction to understand the boundary within which scientists should remain might not be best thought in terms of value judgements. There are indeed other ways for experts to infringe on their audience's decisions, which are more specifically epistemic in kind. In order to highlight them, I will now concentrate on the duty of expert witnesses in criminal trials.

The Credibility and Relevance of Expert Evidence in the Courtroom

Judges and jurors in criminal trials are expected to make highly consequential decisions based on the evidence presented to them in court. But the evidential basis, of which they must appraise the probative force so as to decide whether or not the suspect is guilty of the alleged crime, is far from being a single piece of evidence. It is rather a complex, incomplete, and

partially contradictory set of evidence of heterogeneous types. Among those, forensic expert reports may be a key component, but they cannot be conclusive: experts are asked to report on specific issues, but they should not give their opinion on the final one, namely whether or not the suspect has indeed committed the facts alleged (not to speak about their legal qualification).

To be sure, there is a role for values as described in the inductive risk discussions here. At each step, whether a claim is proven enough to be accepted surely depends on what is at stake.^{ix} But here, I will highlight other aspects of the difficulty of setting the boundary within which expert reports should stay, which are not reducible to values issues, as I will argue.

In order to do so, I will first introduce some tools of analysis from the so-called ‘New Evidence Scholarship’ (see Twining and Stein 1992), and more precisely from David Schum’s (1994) study of the properties and uses of evidence in probabilistic reasoning — which was aimed at applying to any domain and discipline, but whose core object was criminal inquiry and judicial reasoning.

Complex Evidential Reasoning: Inference Networks and ‘Cascaded’ Inference

Confirmation theories in the philosophy of science (Hajek and Joyce 2008, Hartmann and Sprenger 2010, Crupi 2016) study the relation between a given piece of evidence E , and the hypothesis H on which E may or may not have confirmatory bearing. In particular, Bayesian theories usually take E as unproblematically given, and attempt at formalizing the quantitative relationships between E and H — the crucial quantity for assessing the probative value of E for H being the likelihood ratio $p(E|H) / p(E|-H)$.

However, considering the function and uses of evidence in the criminal context draws attention to the strikingly complex inference networks fact finders must construct so as to marshal a mass of evidence, which may or may not cohere. As Schum (1994) has shown, the study of the probative value of a given piece of evidence for a hypothesis cannot be exhausted by a quantitative analysis. In addition, and as a preliminary to such analysis, the *structural*

relationships among evidence items, and the subtleties of the inferential routes that fact finders may take, are worth studying in detail.

One major point is that the probative value of any evidential item is highly dependent on the rest of the available evidence. There are many different ways a piece of evidence may have inferential bearing on a given hypothesis, either directly (*e.g.* a threatening letter addressed by the suspect to the victim is directly relevant — though not conclusive — on whether she killed him), or indirectly (*e.g.* a report questioning the authenticity of such letter). In some cases, one single, *prima facie* neutral piece of information may turn some non-significant information into key evidence, or dramatically change the meaning of another piece of evidence, turning some exonerating evidence into incriminating one, or the other way around (see Schum 1994, chap. 3, for a detailed analysis, and examples).

Another important point is that the inferential path that may lead from a given E to a given H is potentially infinitely decomposable. Even the most ‘direct’ piece of evidence does not automatically and unquestionably lead to the final conclusion; any (inductive) inferential step is in principle defeasible, as resting on some implicit assumption or generalization that can be put in question. Consider a mobile phone geolocation report. In most cases, it seems reasonable to somehow automatically infer that the mobile’s owner was herself geolocated. However, where the stakes are high, and if other pieces of evidence suggest that the suspect might have been in the area where the crime took place at the alleged time of the crime, while his mobile was located somewhere else, it sounds highly reasonable to put into question the implicit generalization that mobile phones are where their owners are.^x Depending on what level of granularity one chooses — which clearly has to depend on the stakes — one can put in doubt every little step, which may appear as obvious and secure in most contexts.

It is worth noting that the generalizations warranting inferences may themselves be of different kinds, from common knowledge (which assumedly is shared among people of the

same community) to personal life experience (less likely to be widely shared) and expert knowledge. As forcefully highlighted by Schum, each agent constructs her own inference network from her very standpoint. To be sure, some generalizations may appear well warranted, while others are clearly prejudicial, and of course there are wrong reasoning paths (whose graphic representation, as proposed by Schum in the line of Wigmore 1913, helps figure out). But even though there might be better routes than others, there is not *the* ‘good’ route; probabilistic, evidential reasoning is doomed to being uncertain.^{xi}

To summarize, another way to cast the idea of inductive risk is to say that each inferential step is grounded on a defeasible generalization, which different agents may be more or less willing to accept. Whether one takes it or not has to depend on the stakes of the inferential process, but also on how much the underlying generalization seems grounded. To grasp the subtleties of the complex inferential relations between different nodes of a network, and how those may shed some light on the role of experts, let me now turn to the distinction, proposed by Schum, between the credibility and the relevance of evidence.

The Credibility and Relevance of Evidence

In order to be of some probative value for a given conclusion (e.g., the hypothesis that the suspect is guilty of the alleged crime), a piece of evidence (e.g., the victim’s neighbor’s report that she saw the suspect in the staircase one hour before the alleged crime) has to be both *credible* and *relevant*.

Credibility

In order to clarify the notion of credibility, it is useful to follow Schum in noting E^* the evidence item itself, as it is received by the subject (the neighbor’s oral or written report; a threatening letter; a CCTV recording; an expert report on fingerprints, etc.), and E the fact that this item attests to (the presence of the suspect in the staircase one hour before the alleged crime; the suspect’s writing of the letter; the happening of the recorded scene; the match between

fingerprints collected on the crime scene and the suspect's fingerprints, etc.). The credibility of E^* warrants inference from E^* to E . Credibility of a given piece of evidence is composed of several elements, depending on the type of evidence at stake. The credibility of an eyewitness report, for instance, can be decomposed at least into the witness's truthfulness (willingness to report what she actually believes she saw), objectivity (ability to correctly interpret what she perceived) and observational accuracy (see Schum 1994).

Clearly, credibility is a matter of degree. A maximally credible E^* would make it certain that E is the case (if E^* is maximally credible, then it is conclusive with regard to E). But, unless E is the final issue, one must also assess its relevance for this issue (or any intermediary hypothesis that might be of some relevance for it).

Relevance

E is relevant for H when it matters in some way for evaluating H ; it is also a matter of degree. A piece of evidence can be highly relevant for a given hypothesis while being minimally credible; on the other hand, one can have a highly credible report on a matter that is of very little relevance for the question at stake. Relevance itself can be assessed along many different lines and, as already emphasized, it is strongly dependent on the rest of the available evidence. Moreover, it can and is often diversely evaluated by different people with different background knowledge and life experience, without necessarily one of them being 'right' and others being 'wrong'.

Credibility and Relevance of Testimonial Evidence: Pragmatic Issues

Although the distinction may seem clear-cut with regard to (lay) testimonial reports, it soon blurs when one considers some pragmatic subtleties of linguistic communication. Even if we assume that a witness' report is perfectly accurate — hence maximally credible —, it is worth acknowledging that a human witness is not a mere tape-recorder, and that she chooses to report some details, and ignores others. Even if we assume that she is entirely willing to be

cooperative, such choice depends on her own appraisal of the relevance of those items. Hence, she may keep silent on some aspects of what she saw that she finds insignificant, but which might have turned out to be highly relevant from another standpoint. Not telling them might thus be misleading. And this is just one example of how the subtleties of human communication may turn a perfectly accurate report into a misleading inferential basis for its recipient. Should one say that such a report is not credible after all? Or is it up to the recipient's to correctly interpret it and evaluate its relevance?^{xii} It would be beyond the scope of this paper to further explore those pragmatic issues. Suffice it to say that even what could *prima facie* be considered as a maximally credible report may be misleading in some cases (for reasons independent to confirmation threshold and values).^{xiii} This has to do with the very *relevance* of the content conveyed — or not conveyed, and with the inferences the recipient will feel entitled to draw on this basis. Let us now consider the specific case of expert testimony.

Scientific Reports: Navigating Between Credibility and Relevance

The pragmatic issues highlighted above may seem less likely to arise in the case of expert reports, at least in the court context: expert witnesses are expected to provide their conclusion on a given issue — they are, so to speak, more constrained in choosing the information they deliver. Hence, there seems to be less room for a diverging interpretation of the implicatures of what she says, or does not say. As soon as the issue on which she is expected to report — call it *E* — is clearly defined, an expert's task can be described as follows: she should give a maximally credible report on *E*, without infringing on her audience's appraisal of the relevance of *E* to *H* (the final issue on which decision has to be made). If a judge asks a forensic expert to tell whether or not the fingerprints found on the crime scene match the suspect's own, her expertise does not go until assessing whether and how much this speaks in favor of guilt. Let me now quickly articulate such a view, and highlight its limits.

What does a maximally credible report on E consist in? Consider fingerprints analyses. First of all, the expert engages her responsibility with regards to the whole chain of events, processes, and analyses which led from the collection of the samples from the crime scene and the suspect, to the conclusion about a match between them. Given inductive risk, such conclusion is doomed to be uncertain: there is a risk of error at each step of the process. But it seems to belong to the expert's duty to take responsibility on all the decisions that lead to a conclusion on E . Up to E , she chooses the relevant standard of proof at each step, independent from considerations of her audience's values. Here, the standards that apply seem to be defined by the relevant scientific discipline, and to be independent from what is at stake in this particular context. This certainly does not solve the problems arising from the fact that most issues experts are asked to report on, are not an all or nothing matter, and are probabilistic in character (this is obviously the case for fingerprints and DNA match, but could be generalized to most cases). As highlighted above, the most precise scientific report might not be the most intelligible and usable one, and may even be misleading in some cases (see footnote 8). But providing an outright conclusion might override the expert's remit, by taking one more step towards the final conclusion. This problem however can be reduced to a value question, as studied in the first part of the paper. Another kind of problem arises when one considers the question of the relevance of the expert's (supposedly maximally credible) conclusion to the final issue at stake.

Following the image proposed above, the expert should not go beyond expressing her conclusions on E — she should not commit on the *relevance* of E for H . The reason for that is not specifically related to an attempt at preventing her own values from impacting her inferential process. Rather, being an expert in, let's say, fingerprints analysis, there is no reason to assume that she is better placed than the judge or jurors to assess the relevance of a match to the final issue at stake in the trial. Or is she?

It is worth noting that finding a match does not exclude the possibility that the fingerprints come from someone else with similar fingerprints (this again is generalizable to any kind of ‘trace’ found on a crime scene). This is a separate issue from the appreciation of the quality of the match itself; this rather seems to concern the relevance of a match for the final issue of finding the defendant guilty or innocent. However, this is clearly a point the expert has something to tell about. She is definitely better placed than non-experts to formulate both the probability of finding such traces if the suspect was on the crime scene, and the probability of finding them if they came from someone else. In other terms, her expertise encompasses both the analysis of the trace, and more general knowledge about fingerprints in the general population. One apparently straightforward solution is to actually define E (the issue on which the expert has to bring a conclusion) as a likelihood ratio encapsulating all relevant information, rather than some conclusion on the quality and precision of a match. After all, the expert should give a maximally credible report on what is on her remit. As we will see in the last part of the paper, this solution, though sensible, does not entirely solve the crucial and tricky problem of defining the exact level at which experts should give their report, which, I will finally argue, should ideally be done through concertation between experts and judges.

Insights from the ‘Case Assessment and Interpretation’ Framework

‘Case Assessment and Interpretation’ (CAI) is a model originally developed in 1998 by an interdisciplinary team of forensic practitioners and statisticians employed by the Forensic Science Service in England and Wales (Cook *et al.* 1998a, 1998b), which has been continuously developed since, and is at the core of forensic scientists’ training. It is aimed at providing them with a conceptual framework and practical protocols for structuring the investigation process, and most importantly for us, for producing reliable and useful expert reports of their results, which are most often pervaded with probabilistic reasoning and statistics.

The first and fundamental prescription of the model, which is based on the principles of Bayesian inference, is that scientists should never concentrate on a particular proposition, without considering at least one alternative. Hence, one crucial task is to precisely frame a pair of propositions to be addressed. In the forensic sphere, those correspond to a prosecution proposition and a defense proposition. At the end of the process, the interpretation of the evidence — the assessment of its probative value — should take the form of the likelihood ratio $p(E|Hp) / p(E|Hd)$, where Hp and Hd are the prosecution and defense hypotheses respectively.

However, as Cook *et al.* (1998a) highlight, the specification of the (pair of) propositions to be addressed “is rarely a trivial matter” (153). Following Cook *et al.* (1998b), one can distinguish between three levels of propositions, called *Source*, *Activity*, and *Offence*. An example of an Offence level pair of propositions would be “Mr A committed the burglary / Another person committed the burglary”; corresponding Activity level propositions could be “Mr A is the man who smashed window X / Mr A was not present when window X was smashed”; and Source level propositions would be “The glass fragment came from window X / They came from some other broken glass object”. As appears here, the propositions of a given pair must be exclusive; they must also be exhaustive “within the framework of circumstances as they are perceived by the scientist at the time of examination.” (1998b, 234) This means that they need not exhaust all conceivable scenarios, but only those that are possible, given other hypotheses one takes for granted (e.g., here, the facts that there exist actually someone who committed the burglary, and that he was acting alone, with no one else around — which could be established by a CCTV footage).^{xiv}

At which level should experts be asked to bring their conclusions? As Cook *et al.* (1998b, 231) note, “the higher the level of the propositions, the greater the assistance that will be given to the court but, of course, the scientist may not stray outside the bounds of his/her expertise nor, indeed, into the realms of advocacy”. So, where exactly do those bounds lie?

It is clear that offence level propositions are “completely outside the domain of the scientist” (*ibid.*, 233), not only because they imply specifically legal notions in order to (legally) qualify the found facts, but also because legal fact finding seldom relies on one and only piece of evidence. Even when one expert report is key to the case, other pieces of evidence must be relevant, which impact the expert’s report relevance — for reasons that may rely on common, or any specific knowledge that is not within the expert’s competences. It is also clear that, at the lowest end of the spectrum, source level propositions should be addressed by the scientist, whose expertise provides her with all legitimacy to draw conclusions on them. Hence, the problem of adjusting the exact level of propositions, which the scientist can, and should address, rather arises when one considers activity level propositions.

[...] level I [source] propositions generally require little in the way of circumstantial information, but level II [activity] propositions cannot be addressed without a framework of circumstances. For example, the transfer/persistence issues relevant to the question “what is the probability of finding this quantity of matching glass if Mr A is truly the man who smashed the window X?” require information about how the window was smashed and the time interval between the incident and the taking of the clothing. Likewise, to address the question “what is the probability of finding this quantity of matching glass if Mr A were unconnected with the smashing of the window?” then it is necessary to know if there is something about Mr A which would predispose him to have glass on his clothing. In the absence of such information, assumptions will need to be made about the relevance of [background surveys about glass transfer and persistence]. The more information that is available to the scientist, then the more effective he/she can be in exercising judgement in relation to level II propositions. In this regard

there needs to be a degree of interaction between scientist, investigator and/or advocate which is probably not required for level I propositions. (Cook *et al.*, 1998b, 233)

Moreover, because more circumstantial information is required for higher level propositions, “the interpretation is more vulnerable to change if, later, that information fails to become evidence” (*ibid.*, 234). More generally, the higher the proposition, the riskier it becomes to draw conclusions, and the more ‘mixed’ (*i.e.* expert and common) is the background knowledge needed to draw inferences. In the absence of enough information for a scientist to conclude on a higher-level proposition, prosecutors and advocates may ask him to “contribute to the argument by answering questions about hypothetical situations such as, for example, ‘is this the quantity of glass you would have expected if Mr X had smashed the window ten minutes before he was picked up?’ [...] There is a blurring of the roles of scientist and advocate here”. (*ibid.*, 236). As a consequence, the expert herself might be the “victim” of the questions asked.

Hence, expert and common knowledge are often intertwined in their contribution to evidential reasoning. Ideally, defining the exact level at which she can actually commit — which, as we have seen, is highly case-dependent, should be interactive. One could consider that it is the judge’s job to find the right question on which she expects the expert to answer; but experts themselves should know until where they can go, which might actually be part of their credibility as experts. It is their duty to fine-tune between giving the best warranted information, while guiding agents in their evaluation of the relevance of such information to the issue at stake.

Conclusion

Experts are expected to deliver reports that are intelligible and usable, without going beyond evidential support. One main obstacle to meeting this requirement can be formulated in terms of ‘inductive risk’: at each and every step of their inquiry, scientists have to make choices.

Discussions about inductive risk, and how this impacts experts' task in advising decision-makers, usually focus on the problem of value judgements: whether or not one advocates the desirability and feasibility of value-free scientific expertise, one acknowledges that, if scientific conclusions are already shaped by the expected costs and benefits of the ensuing decisions, they already contain prescriptions.

I hope to have shown that the question of value-judgements, and the associated decision-theoretic image of a costs and benefits dependent confirmation threshold, might not be the best — and definitely not the only — way to think of the difficult boundary for experts not to infringe. As soon as one takes seriously the extreme subtleness of inference networks in complex evidential reasoning, as the example of criminal inquiry and trials shows, one realizes that there is another way expert advice runs the risk of becoming (unduly) paternalistic, and one which is at least partially independent from any value judgement. Whereas they are definitely expected to commit on all the aspects of their conclusions within their domain of expertise — the 'credibility' side —, their tasks complexifies when it comes to finding the right level at which their expertise allows them to express their conclusions — how far they can and must commit on 'relevance'. As we have seen, whether, and how some scientific conclusion allows for one more inferential step may depend on both expert, and common knowledge.

Bibliography

Betz, Gregor. 2013. "In defence of the value free ideal." *European Journal for Philosophy of Science* 3: 207-220.

Cook, Robert, Evett, Ian W., Jackson, Graham, Jones, Peter J.H., and Jim A. Lambert. 1998a.

"A model for case assessment and interpretation." *Science and Justice* 38 (3): 151-156.

Cook, Robert, Evett, Ian W., Jackson, Graham, Jones, Peter J.H., and Jim A. Lambert. 1998b.

"A hierarchy of propositions: deciding which level to address in casework." *Science and Justice* 8 (4): 231-239.

- Crupi, Vincenzo. 2016. "Confirmation". In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*.
- Douglas, Heather E. 2000. "Inductive risk and values in science." *Philosophy of Science* 67 (4): 559-579.
- Douglas, Heather E. 2009. *Science, policy, and the value-free ideal*. Pittsburgh, PA: University of Pittsburgh Press.
- Fenton, Norman E., Berger, Daniel, Lagnado, David, Neil, Martin, and Anne Hsu. 2014. "When 'neutral' evidence still has probative value (with implications from the Barry George Case)." *Science and Justice* 54 (4): 274-287.
- Hajek, Alan, and James Joyce. 2008. "Confirmation." In Psillos, Stathis, and Martin Curd, eds., *Routledge Companion to the Philosophy of Science*, 115–29.
- Hartmann, Stephan, and Jan Sprenger. 2010. "Bayesian Epistemology." In Bernecker, Sven, and Duncan Pritchard, eds., *Routledge Companion to Epistemology*, London, 609–620.
- Hempel, Carl G. 1965. "Science and human values." *Aspects of scientific explanation*. New York: Free Press, 81-96.
- Jeffrey, Richard. 1956. "Valuation and acceptance of scientific hypotheses." *Philosophy of Science* 23 (3): 237-246.
- John, Stephen, 2015a. "Inductive risk and the contexts of communication." *Synthese*, 192 (1): 79–96.
- John, Stephen. 2015b. "The example of the IPCC does not vindicate the Value Free Ideal: a reply to Gregor Betz." *European Journal for Philosophy of Science* 5 (1):1-13.
- John, Stephen. 2018. "Epistemic trust and the ethics of science communication: against transparency, openness, sincerity and honesty." *Social Epistemology* 32 (2): 72-87.
- Kitcher, Philip. 2011. *Science in a democratic society*, New York: Prometheus Books.

- Lagnado, David, Fenton, Norman, and Martin Neil. 2012. “Legal idioms: a framework for evidential reasoning.” *Argument & Computation*, DOI:10.1080/19462166.2012.682656
- Levi, Isaac. 1960. “Must the scientist make value judgments?” *The Journal of Philosophy* 57 (11): 345-357.
- McCready, Eric. 2015. *Reliability in Pragmatics*, Oxford University Press.
- Picinali, Federico. 2013. “Two Meanings of ‘Reasonableness’: Dispelling the ‘Floating’ Reasonable Doubt.” *The Modern Law Review* 76 (5): 845-875.
- Saul, Jennifer. 2013. *Lying, Misleading, and The Role of What is Said*. Oxford: Oxford University Press.
- Rudner, Richard. 1953. “The scientist *qua* scientist makes value judgments.” *Philosophy of Science* 20 (1): 1-6.
- Schum, David. 1994. *The Evidential Foundations of Probabilistic Reasoning*, Northwestern University Press.
- Twining, William, and Alex Stein, eds. 1992. *Evidence and Proof*, New York University Press.
- Vorms, Marion, and Ulriek Hahn. 2019. “In the space of reasonable doubt.” *Synthese*, forthcoming.
- Wigmore, John H.. 1913. “The problem of proof.” *Illinois Law Review* 8 (2): 77-103.

Notes

ⁱ Of course, there are cases where the expert is expected to prescribe us some particular course of action, as a way of achieving some goal — which we have explicitly or implicitly acknowledged as desirable. But by no means are they expected to tell us what our ultimate goals should be.

ⁱⁱ In the context of an analysis of the norms of scientific experts' communication, John (2015b) insists that one should distinguish between standards of acceptance (which might be immune to inductive risk) and standards of assertion. However, whether there is something like 'acceptance' or 'outright belief', independent from any particular action to be taken (assertion in a given context being one of them), is a difficult issue, which is discussed in (Vorms and Hahn 2019).

ⁱⁱⁱ One could indeed consider that the value according to which admitting a false hypothesis within collective scientific knowledge is far worse than not accepting one which in fact is true, and wait until it is better confirmed, is just a warrant of the good functioning of the epistemic values that govern scientific inquiry (see Levi 1960, John 2015a). Moreover, this can be described as a way to regulating scientists' epistemic paternalism: a scientist's making a claim (within her field of expertise) imposes to any rational being the epistemic duty to accept it (in the absence of specific reasons not to), but such duty is itself grounded in the existence of epistemic warrants intrinsic to scientific practice and institutions. To designate such a maximal confirmation, many authors borrow the phrase 'beyond a reasonable doubt' (see Hempel 1965, Betz 2013, John 2015a) from the judicial domain. However, whether this legal standard of proof should be considered fixed or moving is itself a discussable issue (see Picinali 2013, Vorms & Hahn 2019).

^{iv} It is not even clear that such a distinction between 'pure', and practically orientated science, exists. Any scientific claim, even the most remote from any practical considerations, is likely to have some practical consequences at some point. However, one can follow John (2015a) in distinguishing between different contexts of communications: scientific publications are in principle not directed to a specific audience, but rather to the 'world at large', whereas reports made in private contexts (including public bodies policy-advising), are addressed to a specific

audience, in a specific decisional context. See John (2015a) for a criticism of the idea that standards should be different in those two contexts.

^v John (2015a) temporarily considers, before rejecting it, such a reformulation of Douglas' proposal: "scientists should consider their audience's proper epistemic standards for acceptance when setting their own epistemic standards for assertion."

^{vi} This is Betz's (2013) suggestion to save the 'value-free ideal', in the line of Jeffrey's (1956) point that scientists do not have to accept outright hypotheses.

^{vii} Even if scientific acceptance is graded rather than outright, one has to accept a series of intermediary hypotheses at each and every step (*e.g.* about the good functioning of instruments, or about the reliability of other scientists' results). Moreover, the very structure of intellectual inquiry relies on decisions to be made as to what hypotheses are still worth investigating, what hypotheses can be relied on for the rest of the inquiry, and what hypotheses can be safely rejected. To this regard, even the most theoretical inquiry does not mark off so clearly from a consequential-goal orientated one such as criminal inquiry (see Vorms & Hahn 2019).

^{viii} As John (2015a) notes, "Claims like 'given the evidence, it is extremely likely that neonicotinoids deplete bee populations' or 'given the evidence, it is unclear that neonicotinoids deplete bee populations' do not go beyond the available evidence. However, we know that the former is likely to be heard as 'neonicotinoids deplete bee populations' and the latter as 'neonicotinoids do not deplete bee populations'. The moral status of making a claim turns not only on what we say, but on how others (foreseeably) interpret what we say (Saul, 2013)." See (McCready 2015) about the pragmatic role of evidential language as being primarily plausible deniability.

^{ix} In fact, the very standard of proof in Common Law criminal trials, according to which guilt must be proved *beyond a reasonable doubt* for the suspect to be actually convicted can be

interpreted as embodying the very ratio between the relative costs of sending an innocent man to jail and letting a guilty man free.

^x Here is where values get into the picture again: if the stakes are high, you may scrutinize each and every step. But my point here is that the resulting acceptance or rejection of the inference are not entirely accountable in terms of confirmation thresholds.

^{xi} More recently, Bayes nets have been proposed and developed as a reasoning aid for criminal inquiry, forensic and judicial reasoning (see Lagnado *et al.* 2012).

^{xii} In a work in progress with Anouk Barberousse, we discuss the consequences of such a blurring between credibility and relevance of testimony for the reductionism/anti-reductionism debate in the epistemology of testimony.

^{xiii} Stephen John's (2018) criticism of 'honesty' as a virtue in scientists' communication may shed a new light on such analysis of credibility — understood as the property of yielding the right epistemic states in the recipient (and, assumedly prompting her to make the right decision).

^{xiv} See Fenton *et al.* (2013) for a criticism of the common uses of likelihood ratios in forensic inquiries and reports. Quite notably, a piece of evidence may be neutral at one level, and become diagnostic at another one, which requires to manipulate resulting likelihood ratios with great care.