

Context of Communication: What Philosophers can Contribute

Wayne C. Myrvold
Department of Philosophy
The University of Western Ontario

Draft: Please do not cite.

1. Introduction. The following is a call for philosophers to contribute to the task of developing a subfield of philosophy of science that at present barely exists into a flourishing subfield. It is a field that has the potential to combine contributions from the science and values literature and from formal epistemology, and it is a field in which philosophers have an opportunity to be of use to scientists. It has a rich history, some of which will be mentioned below, but it deserves a richer future.

The topic is *context of scientific communication*, a context distinction from our familiar contexts of justification and discovery, and from the context of pursuit. Its central questions are: Once an experiment is done, and the data has been analyzed, what should a scientist report, and how?

This is not a simple question. The scientist wants (or should want) to convey what has been learned. This often involves a very complex and nuanced epistemic state. It is not clear that we have, as yet, the tools available for concisely conveying a nuanced state of this type.

Our language lends itself most naturally to communicating our epistemic states via statements that one has an appropriate epistemic attitude towards: knowledge, or belief, or acceptance. This is far too crude, and can sow confusion. The debate over the suitability of representing scientific results as a list of statements accepted or rejected has an extensive history, with parallel instantiations among statisticians and philosophers (albeit a history that has been misunderstood in recent years). I discuss this in Section 3, below.

If acceptance is to be rejected, in favour of a more nuanced representation that captures gradations of uncertainty, what is to take its place? An obvious choice is via credence functions, or subjective probabilities, assigning numerical degrees of belief to propositions. This can be a useful tool, but it also has its limitations.

The appropriate attitude to take, is pluralist: there is no single technique that is appropriate for all situations. In Section 4 I will sketch a framework that is friendly to such pluralism, at the expense of being somewhat cumbersome and artificial. I do so with two motivations. One is to motivate philosophers who find the framework attractive to work on elaborating it. The other is to motivate philosophers who are repelled by it to develop something better.

First, some examples to help the reader get a sense of the sorts of problems to be dealt with.

2. Examples

2.1. Global temperature records. Consider the following, apparently similar, and apparently simple questions:

- 1) As of this writing, in mid-2016, in what year, since the beginning of the instrumental temperature record, was the average global surface temperature highest?
- 2) In what year prior to that was the average global surface temperature highest?

These seem like simple questions, with simple answers, to be answered straightforwardly by appeal to readily available data. However, as we shall see, things are a bit more complicated, and the bearing of the data on those two questions was interestingly different.

Let's start with question (1). In January 2016, most major news sources published stories to the effect that 2015 shattered all previous temperature records; that is, that the global mean surface temperature for that year was the highest since the beginning of the instrumental record. These reports followed upon analyses of temperature data from four major sources: the Japan Meteorological Association (JMA), the NASA's Goddard Institute for Space Studies (GISS), the U.S. National Oceanic and Atmospheric Administration (NOAA), and the Climatic Research Unit (CRU) at the University of East Anglia. On January 20, 2016, NASA and NOAA issued a joint press release, whose first sentence reads,

Earth's 2015 surface temperatures were the warmest since modern record keeping began in 1880, according to independent analyses by NASA and the National Oceanic and Atmospheric Administration (NOAA).

This was accompanied by a press conference in which Gavin Schmidt, Director of NASA's Goddard Institute for Space Studies, and Thomas Karl, Director of NOAA's National Centers for Environmental Agencies, presented the findings of their two agencies. On the same day, the UK Met Office issued a press release, beginning.

Provisional full-year figures for global average temperatures reveal that 2015 was the warmest year in a record dating back to 1850.

These were followed by a release from the World Meteorological Organization (WMO), on January 25, 2016, synthesizing results of the GISS, NOAA, and CRU datasets.

On to the second question. A year earlier, in January 2015 most major news sources had reported that 2014 was the hottest year on record, following reports from the same agencies. The joint press release from NASA and NOAA that year, issued on January 16, 2015, NASA and NOAA, opened with,

The year 2014 ranks as Earth's warmest since 1880, according to two separate analyses by NASA and National Oceanic and Atmospheric Administration (NOAA) scientists.

This was also accompanied by a press conference by Schmidt and Karl.

Given the near-identity of the opening sentences of the NASA/NOAA press releases regarding the two years 2014 and 2015, readers may be wondering why I say that these two years are interestingly different. The similarity between these two statements obscures an interesting nuance regarding the 2014 record, which came out in the more detailed presentation by Schmidt and Karl at the press conference, and which is reflected in the more cautious statement from the Met Office, on January 26, 2015, announcing the results of analysis of the CRU data for 2014.

The HadCRUT4 dataset (compiled by the Met Office and the University of East Anglia's Climatic Research Unit) shows last year was 0.56C ($\pm 0.1C^*$) above the long-term (1961-1990) average.

Nominally this ranks 2014 as the joint warmest year in the record, tied with 2010, but the uncertainty ranges mean it's not possible to definitively say which of several recent years was the warmest.

What was the difference? All four data sets agreed that the best estimate of global mean surface temperature for 2014 was higher than the best estimate for any previous year in the instrumental record. But such estimates are surrounded by some uncertainty. Moreover, the top candidates, prior to 2015, for warmest year, are very close in global mean temperature, and the differences are within measurement uncertainty. Figure 1 is a slide from the Schmidt/Karl press briefing of January 20 2015, depicting probabilities arrived at by NOAA and NASA for the truth of claims that this or that year was the warmest.

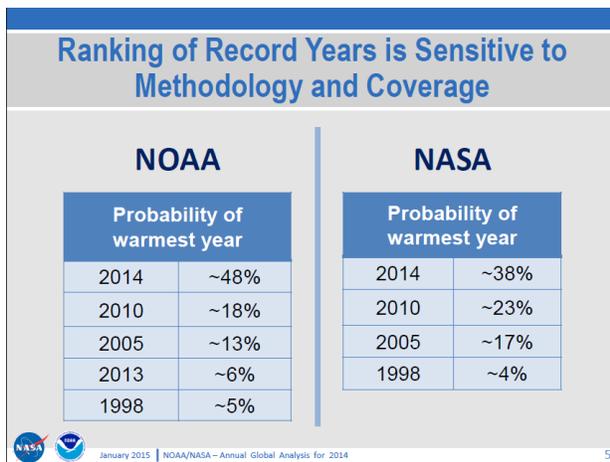


Figure 1

Figure 2 shows probability density functions for temperature of various years (all the record-brakers), provided by Schmidt in a follow-up blog post (Schmidt 2015).

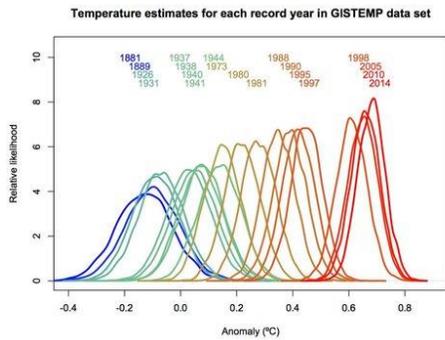


Figure 2

A few things are worth noting. First is that, on both NASA’s and NOAA’s probability estimates, 2014 is more likely than any other year to be the hottest year. Second, on both probability estimates, it is more likely that 2014 is *not* the hottest year than that it is. Third, though the probability rankings produced by NASA and NOAA agree on the ordering of the top three years, the probabilities assigned to these years differs significantly, in some case by as much as 0.10.

In spite of the fact that, on both their analyses, this statement is less probable than its negation, NOAA/NASA decided to make in their joint press release a categorical statement that 2014 was the hottest year on record, saving the more nuanced message for the more detailed briefing. I do not wish to question their decision, which may have been the right one; as Schmidt put it, “Records are bigger stories than trends.” As he noted in his follow-up blog post, the question of whether 2014, 2010, 2005, or some other years was the warmest (by a slim margin) to date is of little importance, certainly less important than the unequivocal fact (on which all the analyses agreed) that the warming trend was continuing, with no signs of abating.

As we have seen, the Met Office adopted a more cautious strategy in its press release. The press release from the WMO, summarizing the results from the three datasets, adopted a mixed strategy:

The representatives of the scientific teams that collected and analyzed the data found themselves with a complex epistemic situation to communicate. The question of what to communicate involved a trade-off: a categorical statement to the effect that 2014 broke a record is easy to grasp and lends itself to headlines but does not accurately the epistemic nuances of the situation; as we have seen, different groups made different choices.

I would like to highlight a few things. One is that a categorical statement that 2014 was, at the time, the hottest year on record was misleading, though it could be argued that, since a mistaken belief about this matter was of little consequence, it was not harmfully misleading (if, on the other hand, some decision of great import would be taken whose consequences would depend on whether or not it was true that 2014 was the warmest years, then, presumably, the statement would have been considerably more hedged). Another is that, if gradations of belief are to be expressed, there was not available a unique epistemic probability function; the analyses of NOAA and NASA differed considerably in the values of the probabilities assigned to various statements. What they opted to report, instead, was both probability distributions.

2.2. Reporting the outcomes of controlled trials. There are news reports almost daily of the form: Scientists have found a link between *A* and *B*. What this typically means is that a controlled trial was performed. Subjects were divided randomly into a test group and a control group. The test group was subjected to intervention *A* and the control group was not. Test and control groups were then evaluated as to presence of *B*. What is being reported is that the experimenters found a difference, with respect to *B*, between test and control group, that exceeds some conventional level of statistical significance, usually either 0.05 or 0.01. What this means is: on the assumption of the null hypothesis that *A* is irrelevant to *B*, the probability of finding a difference with respect to *B* that is at least as large as the difference found is less than 0.05 (or 0.01, or whatever significance level is used).

The standard way of reporting trials of this sort is misleading, in that it suggests that one can conclude, on the basis of such a test, that there is a link, causal or otherwise, between *A* and *B*. What the evidence actually warrants is something much weaker. To see this, consider that using a significance level of 0.05 as the threshold for reporting a positive result is to employ a test that has a false-positive rate of 5%, or 1 in twenty. If a large number of such trials is being done, and if the vast majority of interventions tested are ineffectual, then *most* experiments that yield a statistically significant result will be false positives. This rather elementary point seems difficult to convey, and, even if the conclusions of the papers in scientific journals are appropriately hedged, the way they are reported in the media. For a detailed analysis, see John Ioannidis, “Why Most Published Research Findings are False” (Ioannidis 2005).

If results are reported in a way that suggests that a statistically significant result proves, or makes it reasonable to believe that there is a connection between the variables investigated, then, indeed, most claims of this sort are false. But how should the results of such studies be reported? If there are well-defined priors—that is, if there were a clear answer to how likely one should regard it, before testing, that, say, a given drug is effective—then one could update such priors in light of the evidence and report a posterior probability that the drug is effective.

These two examples illustrate a point that could abundantly be illustrated by other examples, namely, that, in a host of cases, present the state of scientific knowledge on a given topic in the form of a list of propositions accepted and rejected is too crude. Further, representing shades of credibility via probability assignments will also distort the state of things, insofar as it suggests there is a unique correct degree of belief in a given proposition, given the evidence.

3. Is Acceptance acceptable? The issues that our examples bring to the fore were the topic of discussion among the statisticians who were the founders of current practices of statistical hypothesis testing, with Neyman and Pearson, and Wald, on the one hand, formulating decision procedures for accepting and rejecting hypotheses, and Fisher roundly criticizing them for doing so. The debate migrated into the philosophy of science literature in the 1950s, with Richard Jeffrey (1956) arguing, contra Churchman (1948), Braithwaite (1953), and Rudner (1953) that scientists ought not to engage in acceptance and rejection of hypotheses, on much the same grounds as Fisher. Since then, the debate has continued, with Levi (1960, 1967, 1970, 1984), Kyburg (1968), Hempel (1965, 1981), and Maher (1993), among others, advocating acceptance, and Carnap (1968, 1971), Rosenkrantz (1977), Jeffrey (1970, 1991), among others, on the other side.

After briefly reviewing the debate, we will propose a limited role for acceptance, along the lines advocated by Myrvold (2011), which has similarities to the role proposed by Frankish (2004) and Sargent (2009).

3.1 Background in statistical debates: Fisher and Neyman-Pearson. The modern theory of statistical hypothesis testing has its roots in the work of Galton and Karl Pearson. Fisher made seminal advances in the early decades of the twentieth century, and, building on Fisher's work, Jerzy Neyman and Karl Pearson's son Egon Pearson (1928, 1933) developed the theory into (something like) the form found in textbooks today (see Lehmann 2011 for some of the history). Fisher introduced the use of what we now call p -values as tests of statistical significance. The idea is to consider the probability distribution of some appropriate function of the data, adopted as a test statistic, on the supposition of some hypothesis H_0 . We consider the "tail probabilities," that is, for any value of the test statistic, the probability, conditional on H_0 , of obtaining a value that is at least as extreme as the one obtained. This probability, now known as a p -value, is regarded as an indication of the level of significance of the result obtained. In an influential passage, Fisher wrote,

If P is between .1 and .9 there is certainly no reason to suspect the hypothesis tested. If it is below .02 it is strongly indicated that the hypothesis fails to account for the whole of the facts. We shall not often be astray if we draw a conventional line at .05 and consider that higher values of χ^2 indicate a real discrepancy. (1925, p. ?)

Fisher's method does not provide a uniform prescription for picking out, from the space of possible values of a given test statistic, an appropriate "tail region," or a clear rationale for choosing one reason over another, though in the examples considered by Fisher there seems to be a clear natural choice. It is this lacuna that Neyman and Pearson sought to fill. In the problem setting considered by Neyman and Pearson, we have a hypothesis, H_0 , to be compared with one or more alternative hypotheses. Data is gathered, and a decision is to be made whether or not to reject the hypothesis H_0 on the basis of the data. Neyman and Pearson's innovation is to think of the choice of a statistical decision procedure as a task involving a trade-off between the risks of two potential errors; Type I errors, which consist of rejecting H_0 when it is true, and Type II errors, those of failure to reject H_0 when one of the alternatives considered is true.

Neyman and Pearson were clear that the choice of a decision procedure could not be made uniformly, without consideration of the costs associated with the two types of errors.

Let us now for a moment consider the form in which judgments are made in practical experience. We may accept or we may reject a hypothesis with varying degrees of confidence; or we may decide to remain in doubt. But whatever conclusion is reached the following position must be recognized. If we reject H_0 , we may reject it when it is true; if we accept H_0 , we may be accepting it when it is false, that is to say, when really some alternative H_i is true. These two sources of error can rarely be eliminated completely in some cases it will be more important to avoid the first, in others the

second. We are reminded of the old problem considered by LAPLACE of the number of votes in a court of judges that should be needed to convict a prisoner. Is it more serious to convict an innocent man or to acquit a guilty? That will depend on the consequences of the error; is the punishment death or fine; what is the danger to the community of released criminals; what are the current ethical views on punishment? From the point of view of mathematical theory all that we can do is to show how the risk of the errors may be controlled and minimised. The use of these statistical tools in any given case, in determining how the balance should be struck, must be left to the investigator (1933, 295–296).

This is the argument from what Heather Douglas (2000) has called “inductive risk.”

Textbooks on statistics vary in the extent to which this message is taken to heart. The best textbooks (e.g. Lehmann 1959 and subsequent editions) include a passage of this sort. Unfortunately, practice has been to adopt fixed p -values as criteria of significance, without regard to the particular risks associated with the two types of error (a practice that has garnered extensive criticism), and Fisher’s suggested convention of $p = 0.05$ as indicating a significant result has, despite Fisher’s frank acknowledgment of its arbitrariness, been widely adopted.

One thing that Neyman and Pearson don’t tell us is how to balance the risk of the two types of error. If we are to use an expected utility analysis, then more is needed than the two error probabilities and the costs associated with each type of error; we also need prior probabilities for the two hypotheses. An alternative was suggested by Abraham Wald (1949, 1950), who introduced a minimax procedure.

Although Neyman and Pearson thought of themselves as elaborating on Fisher’s work, the key move, that of adopting a decision procedure for acceptance and rejection based on balancing of the risks of the two types of errors, was one that Fisher was not willing to follow them on. In an essay, “Statistical Methods and Scientific Induction,” published in 1955, and also in his 1956 book, *Statistical Methods and Scientific Inference*, Fisher sharply criticized “[t]he attempt to reinterpret the common tests of significance used in scientific research as though they constituted some kind of acceptance procedure and led to ‘decisions,’ in Wald’s sense” (1955, 69); he also added a brief section, 12.1, to the seventh edition (1960) of *The Design of Experiments*. His chief criticism is that, though acceptance procedures are needed and are useful in industrial settings, in which there is a single decision to be made on the basis of the statistical information obtained, such as whether or not to accept a consignment of manufactured goods as free of defects. This is all well and good in such contexts, Fisher says, and “I am thankful, whenever I travel by air, that the high level of precision and reliability required can really be achieved by such means.” However, he continues, things are different in the natural sciences; “the logical differences between such an operation and the work of scientific discovery by physical or biological experimentation seem to me so wide that the analogy between them is not helpful, and the identification of the two sorts of operation is decidedly misleading” (p. 70).

The reason for the difference is the open-ended nature of the use to which information will be put in the natural sciences; in contrast to the single decision to be made in the industrial context, a multitude of decisions may be made on the basis of the information obtained, and, precisely

because of the sensitivity of the choice of an appropriate acceptance criterion on what is at stake in the case of error, no one decision procedure will suffice for all these uses.

Finally, in inductive inference we introduce no cost functions for faulty judgements, for it is recognized in scientific research that the attainment of, or failure to attain to, a particular scientific advance this year rather than later, has consequences, both to the research programme, and to advantageous applications of scientific knowledge, which cannot be foreseen. In fact, scientific research is not geared to maximize the profits of any particular organization, but is rather an attempt to improve *public* knowledge undertaken as an act of faith to the effect that, as more becomes known, or more surely known, the intelligent pursuit of a great variety of aims, by a great variety of men, and groups of men, will be facilitated. We make no attempt to evaluate these consequences, and do not assume that they are capable of evaluation in any sort of currency.

When decision is needed it is the business of inductive inference to evaluate the *nature* and *extent* of the uncertainty with which the decision is encumbered. Decision itself must properly be referred to a set of motives, the strength or weakness of which should have had no influence whatever on any estimate of probability. We aim, in fact, at methods of inference which should be equally convincing to all rational minds, irrespective of any intentions they may have in utilizing the knowledge inferred.

We have the duty of formulating, of summarising, and of communicating our conclusions, in intelligible form, in recognition of the right of other free minds to utilize them in making their own decisions. (1955, 77).

Though, for Fisher, it is not the place of the scientific researcher to introduce cost functions, he is *not* advocating a picture in which scientists are absolved from considerations of the impact of their work on society. Rather, the idea is that it will be most beneficial to the public if scientists refrain from introducing criteria for acceptance and rejection of hypotheses, precisely because such criteria must involve consideration of the cost of error, and the application of the scientists' work may well extend well beyond circumstances those scientists can foresee, and involve costs that they are not in a position to estimate. Far from being a picture on which scientists are absolved from bringing ethical considerations to bear on their work; on this view, it is their *duty* to communicate as clearly as they can the results of their work, including communication of the extent of uncertainty involved.

3.2 The acceptance debate in philosophy of science. Following Wald, Braithwaite (1953) advocated a minimax decision procedure. [Say something about Churchman.] Richard Rudner, citing Neyman, used the dependence of statistical decision procedures on the risks associated with error to argue that "The Scientist *qua* Scientist Makes Value Judgments" (1953).

In an early paper (Jeffrey 1956), written while he was still a graduate student, Richard C. Jeffrey weighed in, taking the Fisherian line against Braithwaite, Churchman, and Rudner. Jeffrey's

argument is couched as a *reductio* of the claim that it is the task of a scientist to accept or reject hypotheses. If the scientist accepts or rejects hypotheses, then Rudner is correct that the criteria for acceptance and rejection will depend on what is at stake. However, it is rarely the case that a single decision will be based on the scientific evidence; the same study might be used to decide whether a given drug is safe for use on humans, or whether it may be used for animals, and one might demand a higher standard of assurance of safety for accepting its use on humans than on animals. Merely reporting acceptance or rejection will, in such a case, lead to some decisions being made in a less-than-optimal way.

Jeffrey summarizes the argument in the concluding section,

On the Churchman-Braithwaite-Rudner view it is the task of the scientist as such to accept and reject hypotheses in such a way as to maximize the expectation of good for, say, a community for which he is acting. On the other hand, our conclusion is that if the scientist is to maximize good he should refrain from accepting or rejecting hypotheses, since he cannot possibly do so in such away as to optimize every decision which may be made on the basis of those hypotheses.

In this early paper (and in his dissertation, completed in the following year), we find an adumbration of the position that Jeffrey would later call “radical probabilism,” which was to be a guiding theme of his career. This is a view that rejects the centrality of epistemic attitudes such as acceptance or rejection of propositions and replaces them, instead, by probabilities reflecting an agent’s judgment of the credibility of a hypothesis. It is *radical* probabilism because Jeffrey does not require such judgments to be grounded in certainty about anything. Rudner’s argument afforded Jeffrey an opportunity to argue against the notion that scientists accept and reject hypotheses, because, according to Jeffrey, the very considerations of value that Rudner invokes should lead Rudner to reject his premise that scientists accept and reject hypotheses.

Because Jeffrey was replying, in part, to a paper entitled “The Scientist *qua* Scientist Makes Value Judgments,” one might expect him to be arguing for the negation of this, and indeed, he has been read in this way.¹ But this claim forms no part of his argument, which can go through without commitment on this matter one way in the other, and nowhere in his paper does deny Rudner’s titular thesis, but, instead, remains neutral on it; note that in the above-quoted passage we have “*if* the scientist is to maximize good” But if we do accept the premise that is permissible for a scientist to perform her job in the way that is most useful to society,² then considerations of value lead to the conclusion that, since any other procedure would lead to suboptimal decisions, the scientist *should* refrain from accepting or rejecting hypotheses.

If the argument were intended as an argument for the thesis that a scientist made be absolved from the responsibility of making value judgements, it is, as Douglas points out, ill-suited for doing so. The scientist *qua* scientist makes decisions, and the process of decision-making doesn’t get off the

¹ The first to do so seems to have been Levi (1960). Heather Douglas (2009), in an influential discussion of Jeffrey, counts Jeffrey among those who argued for a vision of science isolated from broad social concerns, and this reading has become a prevalent one.

² And, seriously, who doesn’t?

ground without considerations of the good or bad of potential consequences of decisions. The process of experimentation involves a multitude of decisions, as does the choice of procedures used to analyze the data, and, at the end of all that, there is the choice of what to publish and how.

There is a problem for radical probabilism, raised by Rudner, which Jeffrey at this early stage does not know how to answer. That is that the scientist must nevertheless accept a determinate probability judgment. Is the would-be radical probabilist not sure of something after all, namely, her own probability judgments?

The problem shifts in Jeffrey's later work. In *The Logic of Decision* (1965) is presented a system that uses an agent's preferences between acts to attribute to her both judgments of desirability (value judgments) and judgments of probability. The agent's preferences are permitted to be a partial order; there may be pair of acts between which she neither has a preference nor is committed to regarding as equally good; she just doesn't know what to say when asked which is better. As a consequence, the representation theorem used by Jeffrey does not yield unique probability judgments. The belief state of an agent is, rather, a *set* of probability judgments.

This can be thought of as representing a sort of second-order uncertainty, uncertainty about what probability to assign to a given proposition. This eliminates the need for acceptance of judgments about particular probability values, but, obviously, only pushes the problem up one level; the agent, it seems, accepts a certain definite *set* of probability functions.

The correct response, I think, is that to have a belief state is not the same as accepting a proposition to the effect that this is my belief state, any more than to have a preference between two options is equivalent to accepting a proposition to the effect that one is more desirable than the other. A radical probabilist need not accept any statements other than logical truths with certainty, even statements about her own belief state.

But a related problem comes up in the context of communication. Suppose we agree that in cases of non-negligible uncertainty the epistemic situation is one in which communicating probability assignment is more appropriate than communicating an assertion of acceptance. As we have seen in our first example, there might be second-order uncertainty, about which probability assignment is most appropriate. We will return to this matter in section 4.

3.3 A limited role for acceptance. There are statements that I regard as so well established that any reasonable person's credence will be, though strictly less than 1, so close to 1 as to make no difference in any decision that any of us is likely to encounter. For such decisions, there is no harm in labelling them as simply *accepted*, and there is considerable gain, in relieving ourselves of the cognitive burden of estimating just how high our credence in such propositions is, and to avoid cluttering up our communications with assessments of the credibility.

For example: Standard equilibrium statistical mechanics, applied to a gas, attributes nonzero probabilities to all sorts of pressure fluctuations. Small fluctuations are rare, and can be observed; larger fluctuations, though rarer, still have a nonzero probability. Whether or not I have full confidence in standard equilibrium statistical mechanics, if I have nonzero credence that these

calculations are even approximately correct, I have a miniscule, though nonzero, credence that in the next three minutes all the air in this room will squeeze itself in a cubic centimetre in the upper left-hand corner.

How small? I don't know. I could work it out, if need be, but there's no need to bother; it's small enough to be completely, utterly, negligible for all practical purposes. No decision that I expect to make will be made any differently than it would if my credence in such an eventuality were actually zero.

If someone asks me what will happen to an inflated bicycle tire that has a hole in it, I will reply, simply, that it will deflate. There is no need to hedge or qualify this assertion with an estimate of the statistical mechanical probability that the tire will remain inflated, as this probability is so small as to be completely negligible.

4. The representation of an epistemic state: Good's hierarchy. For some propositions, namely, those for which there is no harm in neglecting uncertainty, we may represent an agent's state of belief as including that proposition as provisionally accepted (bearing in mind that, in the light of new evidence, the uncertainty might become relevant again).

Should we, then, represent a state of uncertainty by a probability assignment, representing the degree to which belief in various propositions is warranted by the evidence?

It would be absurd to suppose that we will ever be in a position to settle on a completely precise probability assignment that we can reasonably regard as uniquely the best in a given situation. As I. J. Good puts it, "it would only be a joke if you were to say that the probability of rain tomorrow (however sharply defined) is 0.3057876289" ([1979] 1983, 95).

Now, weather forecasters do not, of course, issue probability estimates to 10 decimal places; typically we get integral percentages. For the sorts of decisions to be made on the basis of those forecasts, that is all that is needed; would the difference between 37% probability of rain and a 38% probability of rain make any difference to your behaviour?

Where the uncertainty matters, it will be useful to provide an interval range of probability estimates, or, even better, a second-order probability function, indicating, for any interval, the degree to which one can be confident that the optimal probability estimate is in that interval. This could be useful in conveying, for example, the fact that, as one goes further into the future, one becomes less confident that one has gotten the probability right.

But of, course, one might be uncertain about what the best choice of a second-order distribution is, and so consider a third-order distribution, and so on, as recommended by Good (1952, 1979). We should not expect the regress to be terminated at some level with a probability function that is entirely determinate; rather, one would expect that, the higher one goes up the hierarchy, the harder it will be to uniquely specify a probability function.

It might seem that this inevitably lands us in a hopeless infinite regress. This might make things seem hopeless. It need not, though; as Good remarked in a discussion of decision making with a hierarchical representation of belief states,

the higher the type the woollier the probabilities. It will be found, however, that the higher the type the less the woolliness matters, provided that the complications do not become too complicated ([1952] 1983, 14).

Considerations of the cost (in complication of the analysis) of continuing further up the hierarchy, compared with benefit (in terms of improved decisions) can indicate an end to the analysis. In connection with this, Good remarked,

Isaac Levi (1973, p. 23) says, “Good is prepared to define second order probability distributions ... and third order probability distributions over these, etc. until he gets tired.” This was funny, but it would be more accurate to say that I stop when the guessed expected utility of going further becomes negative if the cost is taken into account. ([1979] 1983, 99).

5. Values and context of reporting. The decision of whether to publish the results of a scientific investigation, and, if so, what to publish, is a context in which ethical considerations clearly have a role to play, as well as considerations such as impact on others and on the scientist’s own career. Typically, the decision to publish involves a decision that publication would be beneficial to someone—society at large, the scientist’s employer, or the scientist herself (and in the best case, all three). There needs to be sufficient motivation for publication to compensate for the cost in time and effort associated with doing so.

That ethical considerations regarding whether to publish have a legitimate role to play is most easily seen in extraordinary situations; for example, cases in which knowledge could be used to do harm.

Consider, then, the sort of decision facing a researcher who has done some research and is deliberating about what to publish. Alice has done some research on a certain topic, and has arrived at what she regards as well-informed credences on that topic, expressed either as a probability function, or a second order probability, function, or whatever is most appropriate. She is tasked with reporting her credences. Other agents will be making some decisions on the basis of Alice’s report.

In the simplest case, all the agents have the same value-judgments, and the decisions that they will make, if they adopt Alice’s reported credences, are exactly the decisions that Alice will endorse. Alice has no incentive to do anything other than report the credences she has arrived at, in the form she deems most valuable.

This sort of case is rarely, if ever, realized. In the more interesting case, among those who will be utilizing Alice’s report will be some who have different goals than Alice.

One thing that Alice could do might be to try to manipulate these others into making decisions that she would prefer, rather than the decisions that they themselves would prefer on an honest assessment of the evidence. Under some circumstances this might be morally permissible; for example, in a time of war it may be morally acceptable to attempt to mislead or misinform the enemy.

But in a situation of democratic deliberation, in which all agents are expected to respect each other's right to make an informed choice, whether or not they themselves endorse this choice, this would be regarded as an impermissible betrayal of trust, if Alice has accepted a societal role as a source of expert judgment. We would ask Alice to set aside her own personal preferences and report her honest assessment of the evidence. That is, Alice would be expected to respect the following.

Principle of trust. A person who has accepted a societal role as a source of expert opinion has a duty to honestly report her judgments.

Typically, we think of independent scientists as having a role in society that subjects them to the principle of trust, and hence to judge those who violate it to be guilty of a violation of professional ethics. Journalists have traditionally been thought to occupy such a role (hence the slogan, "without fear or favour"). A salesperson or corporate spokesperson is not thought of as occupying such a role, and their opinions taken with a grain of salt. This is the reason that PR firms find it useful to adopt the "Third Man" strategy, of representing their client's point of view as if it came from an independent source (see Rampton & Stauber 2001), and it is the reason that scientists who play that role are rightly regarded as betraying a role responsibility.

Suppose a scientist accepts the principle of trust. What does this entail, in terms of what is to be reported?

It entails, in the context of communication, careful consideration of what will be most valuable to the intended recipients, in forming their own judgments (which may be different from the scientist's) about what the best course of action is, in terms of what is known, and the degree and kind of uncertainty involved.

In some cases a proposition may be so well established on the basis of the evidence that every well-informed, reasonable person will make decisions as if it were true.

In other cases, this will not be the case, but (as in the case of weather forecasting), there may be well-established procedures that yield estimates of probability that are sufficiently determinate that any second-order uncertainty about what the best estimate can safely be neglected by all those who might be making decisions on their basis. In such cases, statements of probability estimates are appropriate.

In other cases, there may be genuine uncertainty about what the best probability estimate is, but well-defined procedures for estimating second-order uncertainty.

But what about cases in which it is known that there is uncertainty, but no well-defined procedure for estimating the extent of this uncertainty? This, for example, is the situation that obtains in assessing degree of uncertainty regarding climate projections yielded by global climate models (See IPCC AR5 WGI Ch. 12 . A case can be made that existing tools are inadequate (see Section 7, below). This is a case in which philosophers could contribute to the development of improved techniques for assessing and reporting uncertainty.

6. Conclusion. Where to go from here? The hierarchical model sketched in section 4 has, I claim, the advantage that it is friendly to a principled pluralism of techniques for reporting an epistemic state. When a statement is sufficiently well-established, or the costs of error sufficiently slight, one may make a categorical statement. Where there is more uncertainty, or in situations in which the information may be applied to a variety of situations, with varying cost functions, a statement of probability may be in order, if there are grounds for assessing one. In other situations, a range of probabilities is appropriate... and so on.

Some readers will find this machinery artificial and cumbersome. Very good! I hope that they will be moved to construct something better.

References

- Braithwaite, R. B. (1953). *Scientific Explanation : A Study of the Function of Theory, Probability, and Law in Science*. Cambridge: Cambridge University Press.
- Carnap, Rudolf (1968). "On rules of acceptance," in Lakatos, ed., 146–150.
- (1971). "Inductive Logic and Rational Decisions," in R. Carnap and R. C. Jeffrey, eds., *Studies in Inductive Logic and Probability*, Vol. I (Los Angeles: University of California Press), 5–31.
- Churchman, C. West (1948). *Theory of Experimental Inference*. New York: MacMillan & Co.
- Douglas, Heather E. (2000). "Inductive risk and values in science." *Philosophy of Science*, 67, 559–579.
- (2009). *Science, Policy, and the Value-Free Ideal*. Pittsburgh: The University of Pittsburgh Press.
- Fisher, Ronald A. (1955). "Statistical Methods and Scientific Induction." *Journal of the Royal Statistical Society, Series B* 17, 69–78.
- (1956). *Statistical Methods and Scientific Inference*. Oxford: Hafner Publishing Co.
- Frankish, K. (2004). "Partial belief and flat-out belief," in F. Huber & C. Schmidt-Petri, eds., *Degrees of belief: An anthology* (Oxford: Oxford University Press), 75–93.
- Good, I.J. (1952). "Rational Decisions." *Journal of the Royal Statistical Society Series B* 14, 107–114. Reprinted in Good (1983), 3–14.
- (1979). "Some History of the Hierarchical Bayesian Methodology," in J. M. Bernardo, *et al.*, eds., *Bayesian Statistics* (Valencia: University of Valencia Press), 489–515. Reprinted in Good (1983), 95–105.
- (1983). *Good Thinking: The Foundations of Probability and its Applications*. Minneapolis: The University of Minnesota Press.
- Hempel, Carl G. (1965). "Coherence and Morality." *The Journal of Philosophy* 62, 539–542,
- . (1981). "Turns in the Evolution of the Problem of Induction." *Synthese* 46, 389–404. Reprinted in Hempel (2001), 344–356.
- (1983). "Valuation and Objectivity in Science," in R. S. Cohen and L. Laudan, eds., *Physics, Philosophy, and Psychoanalysis: Essays in Honor of Adolf Grünbaum* (Dordrecht: D. Reidel Publishing Company), 73–100. Reprinted in Hempel (2001), 372–396.

- (2001). *The Philosophy of Carl G. Hempel.*, ed. James H. Fetzer. Oxford: Oxford University Press.
- Ioannidis, John P.A. (2005). “Why Most Published Research Findings are False.” *PLoS Medicine* **2**, 696–701.
- Jeffrey, Richard C. (1956). “Valuation and Acceptance of Scientific Hypotheses.” *Philosophy of Science* **23**, 237–246. Reprinted in Jeffrey (1992), 14–29.
- (1965). *The Logic of Decision*. New York: McGraw-Hill Book Company.
- (1970). “Dracula Meets Wolfman: Acceptance vs. Partial Belief,” in Swain, ed., 157–184.
- (1992). *Probability and the Art of Judgment*. Cambridge: Cambridge University Press.
- Knutti, Reto, David Masson, and Andrew Gettleman (2013). “Climate Model Genealogy: Generation CMIP5 and How We Got There.” *Geophysical Research Letters* **40**, 1184–1199.
- Kyburg, Henry (1968). “The Rule of Detachment in Inductive Logic,” in Lakatos, ed., 98–119.
- Lakatos, Imre, ed. (1968). *The Problem of Inductive Logic: Proceedings of the International Colloquium in the Philosophy of Science, London, 1965, volume 2*. Amsterdam: North-Holland Company.
- Lehmann, Erich L. (1959). *Testing Statistical Hypotheses*. New York: John Wiley & Sons.
- (2011). *Fisher, Neyman, and the Creation of Classical Statistics*. Berlin: Springer.
- Levi, Isaac (1960). “Must the Scientist Make Value Judgments?” *The Journal of Philosophy* **57**, 345–357.
- (1967). *Gambling with Truth: An Essay on Induction and the Aims of Science*. Cambridge: The MIT Press.
- (1970). “Probability and Evidence,” in Swain, ed., 134–156.
- (1976). “Acceptance Revisited,” in Radu J. Bogdan, ed. *Local Induction*. Dordrecht: D. Reidel Publishing Company.
- (1984). *Decisions and Revisions: Philosophical essays on knowledge and value*. Cambridge: Cambridge University Press.
- Maher, Patrick (1993). *Betting on Theories*. Cambridge: Cambridge University Press.

- Met Office, “2014 one of the warmest years on record globally.” Press release, 26 January 2015. <http://www.metoffice.gov.uk/news/release/archive/2015/2014-global-temperature>
- Myrvold, Wayne C. (2012). Epistemic Values and the Value of Learning.” *Synthese* **87** 547–568.
- NASA (2015). “NASA, NOAA Find 2014 Warmest Year in Modern Record.” Press Release, January 16, 2015. <http://www.nasa.gov/press/2015/january/nasa-determines-2014-warmest-year-in-modern-record/>
- NASA (2016). “NASA, NOAA Analyses Reveal Record-Shattering Global Warm Temperatures in 2015.” <http://www.nasa.gov/press-release/nasa-noaa-analyses-reveal-record-shattering-global-warm-temperatures-in-2015>
- Neyman, J., and E.S. Pearson (1928). “On the Use and Interpretation of Certain Test Criteria.” *Biometrika* **20A**, 175–240, 263–295.
- (1933). “On the Problem of the Most Efficient Tests of Statistical Hypotheses.” *Philosophical Transactions of the Royal Society of London, Series A* **231**, 289–337.
- Rampton, Sheldon, and John Stauber (2001), *Trust Us, We’re Experts*. Penguin.
- Rosencrantz, Roger (1977). *Inference, Method, and Decision: Towards a Bayesian Philosophy of Science*. Dordrecht: D. Reidel Publishing Company.
- Sargent, M. (2009). “Answering the Bayesian challenge.” *Erkenntnis* **70**, 237–252.
- Schmidt, Gavin (2015). “Thoughts on 2014 and ongoing temperature trends.” Blog post, *Real Climate*, 22 January 2015. <http://www.realclimate.org/index.php/archives/2015/01/thoughts-on-2014-and-ongoing-temperature-trends/>
- Schmidt, Gavin A., and Thomas R, Karl (2015). “NOAA/NASA Global Analysis for 2014.” <http://www.ncdc.noaa.gov/sotc/briefings/201501.pdf>
- Swain, Marshall, ed., (1970). *Induction, Acceptance, and Partial Belief*. Dordrecht: Reidel.
- Tokyo Climate Center, Japan Meteorological Agency (2015). “State of Global Warming in 2014.” *TCC News* **39**, Winter 2015. <http://ds.data.jma.go.jp/tcc/tcc/news/tccnews39.pdf>
- (2016). “State of the Global Warming in 2015.” *TCC News* **43**, Winter 2016. <http://ds.data.jma.go.jp/tcc/tcc/news/tccnews43.pdf>
- Wald, Abraham (1949). “Statistical Decision Functions.” *Annals of Mathematical Statistics* **20**, 165–205.
- (1950). *Statistical Decision Functions*. New York: John Wiley & Sons.

World Meteorological Organization (2015). “Warming Trend Continues in 2014.” <http://public.wmo.int/en/media/press-release/warming-trend-continues-2014>. Published 2 February 2015.

——— (2016). “2015 is hottest year on record.” <http://public.wmo.int/en/media/press-release/2015-hottest-year-record>. Published 25 January 2016.