Inductive Risk and Values in Science *

Heather Douglas† ‡

Department of Philosophy, University of Puget Sound

---

abstract

Although epistemic values have become widely accepted as part of scientific reasoning, non-epistemic values have been largely relegated to the "external" parts of science (the selection of hypotheses, restrictions on methodologies, and the use of scientific technologies). I argue that because of inductive risk, or the risk of error, non-epistemic values are required in science wherever non-epistemic consequences of error should be considered. Illustrated with examples from dioxin studies, I show how non-epistemic consequences of error can and should be considered in the internal stages of science: choice of methodology, characterization of data, and interpretation of results.

1. Introduction. Despite its central importance to the philosophy of science, the issue of whether, which, and how values have a role to play in science has been discussed little in the past 50 years. With some exceptions (see below), the common wisdom of philosophers of science has been that only epistemic values have a legitimate role to play in science. While many have claimed that non-epistemic (i.e. social, ethical, political) values in fact do play a role in science, the normative standard of "value-free" (read "non-epistemic value free") science remains.

In this paper, I will challenge that normative standard for large areas of science. I will argue that non-epistemic values are a required part of the internal aspects of scientific reasoning for cases where inductive risk includes risk of non-epistemic consequences. In these cases, value-free science is inadequate science; the reasoning is flawed and incomplete. Thus the normative standard needs to be reconsidered. For science which has clear non-epistemic impacts, being "value-free" is not a laudable goal. As I will note at the end of my paper, this does not mean any argument whatsoever is a good argument in science. Accepting the role of values in science does not eliminate the requirement for good arguments. It only modifies the understanding of what can count as a good argument.

It must be noted that this challenge to the normative standard is obviously not the only one set forth recently. In the past ten years, feminist philosophers of science have challenged the normative standard, primarily by challenging the epistemic/non-epistemic distinction on which it rests. (Rooney 1992; Longino 1996) I find their arguments largely persuasive and I have written elsewhere on the porousness of this distinction. (Machamer and Douglas 1999) While the distinction cannot do the boundary work many philosophers would like it to (i.e.,

maintaining the boundary between legitimate uses of values and illegitimate uses that would threaten the objectivity of science), the distinction can serve to remind us which goals the values primarily serve within a particular context. This is how I will use it for the remainder of this paper.

My focus for the role of values in science centers on Hempel's concept of "inductive risk." Hempel's articulation of inductive risk encapsulates the main arguments over values in science from the debates on that issue from 1945-1965. After using Hempel's work to provide the necessary background on inductive risk, I will discuss how inductive risk fits in with other work on the legitimate use of non-epistemic values in science. To illustrate how consideration of inductive risk can require the use of non-epistemic values, I will discuss examples from recent laboratory animal studies of dioxin's ability to induce cancer.[1,2] It is precisely such contentious areas as this that have caused public questioning of science and that generate much heat but little light on the values in science question. With these examples, I hope to convince the reader that a decision by the scientists in the examples would not be complete without a consideration of non-epistemic values through inductive risk.

2. Inductive Risk. From 1948-1965, a series of papers[3] raised the issue of whether values could be a legitimate part of scientific reasoning. All those that argued for the legitimate use of values in science did so on the basis of the concept of inductive risk. Inductive risk, a term first used by Hempel (1965)[4], is the chance that one will be wrong in accepting (or rejecting) a scientific hypothesis. While papers such as Churchman 1948 and Rudner 1953 argued that the risk of inductive error meant that values must play a role in science generally,

other papers, such as Jeffrey 1956 and Levi 1962, sought to limit the influence of non-epistemic values in science. Because Hempel's views encapsulate considerations from both sides of this debate, I will focus on Hempel to introduce the concept of inductive risk. A more in-depth historical examination of this material must await a future paper.

In his 1965 essay, "Science and Human Values," Hempel articulates the traditional view of philosophers regarding the possibility that values could act as presuppositions for scientific arguments. (Hempel 1965) According to Hempel, value statements have no logical role to play when one is trying to support a scientific statement. Judgments of value lack "all logical relevance to the proposed hypothesis since they can contribute neither to its support nor to its disconfirmation." (*ibid*., 91) This traditional view does not encapsulate the entirety of Hempel's thinking on science and values, however. Hempel holds that values can serve as presuppositions to what he calls scientific *method*. Because no evidence can establish a hypothesis with certainty, "acceptance (of a hypothesis) carries with it the 'inductive risk'" that the hypothesis may turn out to be incorrect. (*ibid*., 92) Inductive risk is the risk of error in accepting or rejecting hypotheses.

Hempel then considers what rules should be used by a scientist when accepting or rejecting hypotheses, arguing that values do have an important role to play in the rules of acceptance. (*ibid*., 92) Hempel considers rules of acceptance to be "special instances of decision rules" (such as maximizing expected utility), which must consider both the possibility that the decision to accept a hypothesis (or reject a hypothesis) proves right and the possibility that it proves wrong. As Hempel states:

When a scientific rule of acceptance is applied to a specified hypothesis on the basis of a given body of evidence, the possible 'outcomes' of the resulting decision may be divided into four major types: (1) the hypothesis is accepted (as presumably true) in accordance with the rule and is in fact true; (2) the hypothesis is rejected (as presumably false) in accordance with the rule and is in fact false; (3) the hypothesis is accepted in accordance with the rule, but is in fact false; (4) the hypothesis is rejected in accordance with the rule but is in fact true. The former two cases are what science aims to achieve; the possibility of the latter two represents the inductive risk that any acceptance rule must involve. (*ibid*., 92)

In order to formulate the acceptance rules properly, Hempel suggests, one must decide how one values the various outcomes: "The problem of formulating adequate rules of acceptance and rejection has no clear meaning unless standards of adequacy have been provided by assigning definite values or disvalues to those different possible 'outcomes' of acceptance or rejection." It is in this way that value statements act as legitimate premises in whether or not to accept or reject a scientific hypothesis. Values are needed to weigh the consequences of the possible errors one makes in accepting or rejecting a hypothesis, i.e. the consequences that follow from the inductive risk.

Depending on the outcomes, different kinds of values will be required for the justification of an acceptance rule. For some cases, acceptance (or rejection) of a hypothesis will lead to a particular course of action and outcomes with non-epistemic effects. In these cases, the outcomes of the potential actions need to be evaluated using non-epistemic values in order to

formulate rules of acceptance. In other cases, where the acceptance of a hypothesis will not

lead clearly to any particular course of action, Hempel believes the question of how to assign

values to the outcomes to be considerably more difficult.[5] Instead of valuing the practical

outcomes, one must instead consider the outcomes in terms of the goals of science, which

Hempel describes as "the attainments of an increasingly reliable, extensive, and theoretically

systematized body of knowledge." (*ibid*., 93) In current terms, Hempel is providing a

potential set of epistemic values with which to determine what our rules of acceptance ought

to be: reliability, extensiveness, and systematization.[6]

Although inductive risk is present in scientists' decisions to accept a theory, it is not obvious

that scientists should consider all the consequences entailed by inductive risk. One might

argue, as both Richard Jeffrey and Ernan McMullin have, that we should not expect or

demand that scientists consider the consequences of accepting a theory erroneously. (Jeffrey

1956; McMullin 1983) Such considerations, McMullin argued in his 1982 Presidential

Address to the Philosophy of Science Association, should be made by those using or

applying the science. Under this view, the value judgments attached to various outcomes, or

"utilities," are not the concern of the scientists. As McMullin stated: "Such utilities are

irrelevant to theoretical science proper and the scientist is not called upon to make value-

judgments in their regard as part of his scientific work." (McMullin 1983, 8)

This argument, however, overlooks the authority science and scientists have in our culture

and the important role scientists play in practical decision-making. Where science is "useful"

it will have effects beyond the development of a body of knowledge. In many contexts, if a

scientist affirms something as true, or accepts a certain theory, that statement is taken as

authoritative and will have effects, potentially damaging ones, if the scientist is wrong. And, as I will argue below, scientists also take inductive risks in stages of science before acceptance or rejection of theories, thus considering risks never brought to the light of public decision-making. Public decision-makers thus have no ability to consider McMullin's "utilities," leaving the job to the scientists. To claim that scientists ought not consider the predictable consequences of error (or inductive risk) is to argue that scientists are somehow not morally responsible in for their actions as scientists. To defend a completely "value-free" science would require such a move, one which seems to be far more dangerous than openly grappling with the role of values in science. Arguing that scientists have the same moral responsibilities as the rest of us is beyond the scope of this paper.

3. The Structure of Values in Science. Inductive risk provides just one way for values to play a role in science. It is a critical way, I believe, both for the dismantling of a value-free normative standard for science and for a clearer understanding of how and why scientific disputes occur. In this section, I will place inductive risk into the context of other views on science and values, showing the general structure of how and where values play a role in science. In the process, I will expand on Hempel's limited view of inductive risk for acceptance of hypotheses, arguing that inductive risk is relevant throughout the scientific process.[7]

To place the idea of inductive risk in context, it should be noted that there are three decision points in the scientific process where non-epistemic values are widely recognized as having a legitimate role. (Here, I follow Longino 1990, 83-85) First, values (both epistemic and non-epistemic) play important roles in the selection of problems to pursue. Second, the direct use

to which scientific knowledge is put in society requires the consideration of non-epistemic values. For example, if science enables the development of a new technology, values are (or should be) consulted to determine whether such a technology is desirable. Third, non-epistemic values place limitations on methodological options, such as limitations on how we can use humans in experimentation.

In all of these cases where non-epistemic values are recognized as legitimate, those values play a direct role in the decision-making. One is not considering the consequences of error, as one is with inductive risk, but considering the direct consequences of a particular course of action. Performing experiments on humans without their consent is unethical not because of the chance that the methodological choice will lead to inaccurate and misused results, but because of the direct consequences of the methodology on the human subjects. Similarly the use of science to develop an unwanted technology is unethical not because of the potential unintended consequences of the technology (although those might also be a problem), but because of the *intended* consequences. In these cases, moral concerns over the direct consequences of actions override any potential epistemic benefits, thus limiting choices. Longino has pointed out that these three roles for values in science are compatible with an "externality" picture of values in science. (Longino 1990, 85-86) Under this model, the "internal" process of scientific reasoning can go forward without the necessary inclusion of non-epistemic values. Non-epistemic values serve as constraints for some scientific choices, but do not interfere with internal scientific reasoning.

Consideration of values through inductive risk does not fit with the "externality" model, however. First, the role of values is indirect, instead of direct. As Hempel rightly pointed

out, value judgments have no direct place in the argument for what should be taken to be true. However, because error is always a possibility, we are required to consider the consequences of error alongside the arguments concerning evidence. And the consideration of the consequences of error *require* the consideration of values, both epistemic and non-epistemic. The role for values is there, even if it is not direct.

Second, with inductive risk, the places in the scientific process where the values play a role are not limited to the outskirts of science.[8] In the externality model, the internal stages of science remain free of the values. Consideration of inductive risk, however, occurs throughout the scientific process. Although Hempel focused entirely on inductive risk at the point of theory acceptance, there are other places in the scientific process where inductive risk is relevant. If one follows the general schema of the methodology from a scientific research paper, significant inductive risk is present at each of the three "internal" stages of science: choice of methodology, gathering and characterization of the data, and interpretation of the data. At each point, one can make a wrong choice (i.e. epistemically incorrect), with consequences following from that choice. A chosen methodology assumed to be reliable may not be. A piece of data accepted as sound may be the product of error. An interpretation may rely on a selected background assumption that is erroneous. Thus, just as there is inductive risk for accepting theories, there is inductive risk for accepting methodologies, data, and interpretations.

By expanding where we see relevant inductive risk, the potential role for non-epistemic values has also expanded. Hempel was right in asserting that whether or not a piece of evidence is confirmatory of a hypothesis (given a set of background assumptions) is a

relationship in which value judgments have no role. What evidence is available to support a theory or which background assumptions we chose to hold, however, does involve value judgments through the consideration of inductive risk. In cases where the consequences of making a choice and being wrong are clear, the inductive risk of the choice should be considered by the scientists making the choice. In the cases I discuss below, the consequences of the choices include clear non-epistemic consequences, *requiring* non-epistemic values in the decision-making. Thus, where the weighing of inductive risk requires the consideration of non-epistemic consequences, non-epistemic values have a legitimate role to play in the internal stages of science. The externality model is overthrown by a normative requirement for the consideration of non-epistemic values, i.e., non-epistemic values are required for good reasoning.

In these cases where inductive risk is involved, non-epistemic values are not the sole determinant of whether to accept a given option. The scientist will need to consider both the quantity of evidence or degree of confirmation to estimate the magnitude of inductive risk and the valuation of the consequences that would result from error to estimate the seriousness or desirability of the consequences. The weighing of these consequences, in combination with the perceived magnitude of the inductive risk (i.e. how likely one is to be wrong), determines which choice is more acceptable. Where non-epistemic consequences follow from error, non-epistemic values are essential for deciding which inductive risks we should accept, or which choice we should make.

4. Inductive Risk in Methodological Choice: Statistical Significance. As noted in the previous section, that non-epistemic values have a role to play in the making of

methodological choices is little disputed. When a methodological option has direct consequences that are ethically unacceptable, that methodology is not considered a viable choice. Such is the case particularly for studies with human subjects, which must be performed with a constant eye towards the appropriate treatment of the subjects. Valuing direct consequences are not the only way in which non-epistemic values play a role in methodological choice, however. In this section, I describe how inductive risks are taken when making the methodological choice of the appropriate level for statistical significance, and how these risks entail non-epistemic consequences. Under most circumstances, the choice of a level of statistical significance is not made through the explicit consideration of arguments for different statistical choices, but by the tradition of an area of research or the choice of a computer statistical package. I will not argue that such conventions have not worked reasonably well.[9] Instead I will examine the reasoning required to make a deliberate and explicit choice of a level of statistical significance for conducting toxicological studies, showing how non-epistemic values would play a role in such a choice.

The deliberate choice of a level of statistical significance requires that one consider which kind of errors one is willing to tolerate. For any given test, one must find an appropriate balance between two types of error: false positives and false negatives. False positives occur when one accepts an experimental hypothesis as true and it is not. False negatives occur when one rejects an experimental hypothesis as false and it is not. Changing the level of statistical significance changes the balance between false positives and false negatives. If one wishes to avoid more false negatives and one is willing to accept more false positives, one should lower the standard for statistical significance. If one wishes, on the other hand, to avoid false positives more, one should raise the standard for statistical significance. For any

given experimental test, one cannot lower both types of error; one can only make trade-offs from one to the other. In order to reduce both types of error, one must devise a new, more accurate experimental test (such as increasing the population size examined or developing a new technique for collecting data).

In laboratory animal studies, tests for statistical significance are used to determine when the response of the dosed or exposed animals is significantly different from the non-dosed or control animals. Because of the control exerted in the lab over the conditions of the animals, if there is a response that is significantly different between the exposed and the control animals, that difference can generally be attributed to the dose given to the animals. The statistical comparison between exposed and control animals is particularly important for cancer studies, where both control animals and exposed animals will usually exhibit some cancer. What must be determined is whether the exposed animals exhibit significantly more cancer than the control animals. Only when a cancer rate is significantly different from the control cancer rate is it considered a genuine result of the dosing. Thus, setting the standard for statistical significance will impact what is considered a response caused by the dosing. The stricter the standards for statistical significance, the greater the difference between dosed animals and control animals will need to be for the response to be considered significant. Stricter standards lead to a reduction in the rate of false positives, and an increase in the rate of false negatives. On the other hand, if one has a laxer standard for statistical significance, a smaller difference in cancer rates between exposed and control populations will be considered significant and attributed to the dosing regimen. This increases the likelihood of false positives, but lowers the likelihood of false negatives.

In setting the standard for statistical significance, one must decide what balance between false positives and false negatives is optimal. In making this decision, one ought to consider the consequences of the false positives and false negatives, both epistemic and non-epistemic. I will focus here on the non-epistemic consequences. In laboratory animal studies testing the potential harms of environmentally pervasive chemicals (such as dioxins), the results are used to determine both whether the chemical has a particular effect and what the dose-response relationship is for the chemical and the effect. The results are then extrapolated to humans (a controversial subject I will not address here) and used to set regulatory standards for the chemical.

In testing whether dioxins have a particular effect or not, an excess of false positives in such studies will mean that dioxins will appear to cause more harm to the animals than they actually do, leading to overregulation of the chemicals. An excess of false negatives will have the opposite result, causing dioxins to appear less harmful than they actually are, leading to underregulation of the chemicals. Thus, in general, false positives are likely to lead to stronger regulation than is warranted (or overregulation); false negatives are likely to lead to weaker regulation than is warranted (or underregulation). Overregulation presents excess costs to the industries that would bear the costs of regulations. Underregulation presents costs to public health and to other areas affected by damage to public health. Depending on how one values these effects, an evaluation that requires the consultation of non-epistemic values, different balances between false positives and false negatives will be preferable.

In addition to whether a substance has a particular effect, laboratory animal studies also are used to determine the dose-response relationship for the effect. Two different models are used when interpreting dose-response data: the threshold model and the linear extrapolation model. (The choice of a model is discussed further below.) The threshold model assumes that there is no response or effect caused by the chemical under study below a certain dose. The dose below which there is no effect is called the threshold. The linear extrapolation model, on the other hand, assumes that the chemical under study is capable of producing biological effects with decreasing rates at ever decreasing doses. Thus one must extrapolate a curve back from the tested doses, usually a linear extrapolation through the origin.

Deciding whether an exposed group's response differs significantly from the control group's response is essential for determining the details of either dose-response model. The statistical significance test tells us whether or not the difference in response is significant and attributable to the dose. Thus, the standard for statistical significance will affect both what is considered a response and the shape of the dose-response curve. For the threshold model, the no-response level is determined by where there is no *statistically significant* response, i.e., the threshold is defined in terms of observable response and observability is defined in terms of statistical significance. Thus a false negative generally means that the "safe" dose is set higher than it should be, which will be less protective of public health. Because dose levels in animal studies are usually set one order of magnitude apart, the "safe dose" resulting from a false negative will be at least one order of magnitude less protective than it should be. For a dose-response extrapolation curve, a false negative will have varied results, depending on the shape of the curve, but it will in general produce a less dangerous looking curve, leading to laxer regulations. False positives, on the other hand, will produce excessively protective

safe doses (in the threshold model) or more dangerous looking dose-response curves (in the extrapolation model), generating stricter than necessary regulation.

In finding the appropriate balance between false positive and false negative errors, we must decide what the appropriate balance is in the consequences of those errors: overregulation and underregulation. Selecting an appropriate balance will depend on how we value the effects of those two consequences, whether we are more concerned about protecting public health from dioxin pollution or whether we are more concerned about protecting industries that produce dioxins from increased regulation. To value one objective and not to value the other at all does not seem to be a plausible position; we would not want to choose to have only false positives or only false negatives. Finding the balance requires, among other things, weighing the non-epistemic valuations of the potential consequences.

Reducing the possibility of any error by increasing the power of the study would help mitigate the dilemma here, but doing so is extremely difficult. For example, one way to reduce the risk of both false positives and false negatives is to increase the animal populations under study. Currently, most studies use 50-100 animals in each dose group. Increasing those numbers would decrease the chance of false positives and false negatives. However, it is extremely expensive and difficult to do larger studies. A single two-year cancer study using 200 rats (50 at three dose levels and 50 controls) costs roughly $3 million. (Graham and Rhomberg 1996, 18) The cost and logistics of increased dose groups can be overwhelming. Perhaps some other solutions to strengthen studies will present itself, but in the meantime, some balance must be struck. Where the balance should lie for dioxin studies

is currently unclear. Regardless, determining the balance clearly requires an ethical value judgment in the internal stages of a scientific study.

5. Inductive Risk in Evidence Characterization: Rat Liver Tumors. Once one has implemented the method chosen for the study, the data must be gathered and characterized. Evidence characterization occurs late in the studies I examine here, after the laboratory animals have been dosed for many months. One can encounter difficulties in evidence characterization that are unforeseen when one is selecting a methodological approach. Deciding how to grapple with ambiguities in data sources that are not expected is my concern in this section. I will show how, as with the other stages in science, there is inductive risk in making choices, and one must decide what amounts (or levels) and kinds of inductive risk are acceptable. Some of the consequences of the risks are non-epistemic, and thus non-epistemic values are needed to weigh the consequences and to make the choice.

In dioxin cancer studies, rodents (the animal group of choice because of their relatively short lifespan and rapid breeding cycles) are dosed for two years, close to a natural life-span. At the end of those two years, full body autopsies are performed on the animals to gather the endpoint data.   Because dioxins appear to affect more than one organ site, all potential areas for cancerous growths are checked. In the studies relied upon by regulators for making decisions about dioxins, tissue and organ samples have been mounted on slides to be evaluated by toxicologists. One particular study, published in 1978 by Richard Kociba and other toxicologists at Dow Chemical, has been central to regulators in setting acceptable levels for dioxins in the environment. (Greenlee *et al*. 1991, 567; Huff *et al*. 1991, 72) The first long term cancer study performed for dioxins, the Kociba study focused attention on

cancers of the liver in female rats. (Kociba *et al*. 1978) The female rat liver slides have undergone at least three evaluations by pathologists, with different results.

In table 1, three different evaluations of the rat liver slides from the Kociba studies are given. The first evaluation, from 1978, was originally reported in the Kociba study. The second evaluation, from 1980, was performed at the behest of the EPA by Dr. Robert Squire. The third evaluation, from 1990, was performed by a team of seven pathologists assembled by a private contracting company PATHCO, Inc. at the request of the paper industry. (EPA 1994, 6-5)   The table reports the numbers of benign (B), malignant (M) and total (T) tumorous livers as a fraction of the total livers examined from female rats at each of the three dose levels plus the control group. Note the changing numbers of rats at each dose perceived to have liver tumors under the different evaluations. These evaluations are of precisely the same slides; the total experiment was not being re-performed. Only the classification of the rat liver slides was reevaluated.

The third slide evaluation in 1990 was precipitated by a reconsideration of the understanding and evaluation of rat liver changes. In the mid-1980s, the basic categories for rat liver anomalies came into question and a new system was put into place that was supposed to add clarity to rat liver abnormality evaluations. (Maronpot *et al*. 1986) Yet when these new standards were used by the seven pathologists in the 1990 re-evaluation, agreement was not readily achieved on how the various abnormalities should be classified. The seven experts resorted to majority voting to reach an opinion about the slides. (The article describing the voting claimed that "consensus was reached when at least four out of seven pathologists agreed." (Goodman and Sauer 1992) This is not a consensus as normally understood, but a

simple majority.) That the pathologists resorted to voting indicates that there is still a significant degree of judgment required in the evaluation of the rat liver slides. Even with the new standards, how the standards are to be implemented is viewed differently among experts.

As evidenced by the lack of agreement among expert pathologists, the judgment of whether a tissue sample has a cancerous lesion or not has proven to be more subtle than one might initially suppose. Thus, for a slide evaluation, there is significant uncertainty in whether a judgment is correct. With this uncertainty comes significant inductive risk and the need to evaluate consequences of potential errors. Although not as formal as setting a level for statistical significance, the pathologists must be similarly concerned with false positives and false negatives. Suppose a pathologist chooses to take all borderline cases and judge them to be non-cancerous lesions. Such an approach will insure few false positives, but will likely lead to several false negatives. The consequences for such an approach will be an underestimation of malignancies and thus an underestimation of risk. Because the Kociba studies have been so important in regulation, a lowered estimate of risk arising from a reinterpretation of the studies will likely lead to a relaxed regulation (as the 1990 reevaluation did in the state of Maine). (Brown 1991, 17) If the regulation has been lowered on the basis of erroneous judgments (false negatives), then the lowered regulation may cause increased harm to the public. In choosing to judge all borderline cases as non-cancerous, the pathologists must judge this an acceptable risk.

Another pathologist may take a different approach, judging all borderline cases to be malignancies. This approach, in contrast, will reduce false negatives, but will likely produce

several false positives. Such false positives will increase the apparent rate of malignancies in the rats, thus increasing the appearance of risk. For a data set as important as the Kociba studies, such a (false) appearance would lead to a more stringent (than necessary) regulation. Although choosing to judge borderline cases as malignancies will more amply protect public health, the approach does so at the economic costs of potentially unnecessary regulation. When pathologists view the slides, borderline cases will occur (as evidenced by the lack of agreement among pathologists). Some judgments must be made by the pathologists regarding how to classify the liver slides. Depending on how one values these consequences of false positives and false negatives, one would want to make questionable judgments in favor of one direction or another.

One might argue that blinding the pathologists to which dose groups the slides belong could alleviate the need to consider the consequences of the errors in judgment. If pathologists have a tendency towards avoiding false positives and thus produce more false negatives (or vice versa) but distribute that tendency evenly among control and dosed slides, then the *difference* between control and dosed slides will be roughly the same (assuming the number of borderline slides in each group is the same). In addition, the consequences of errors on particular slides cannot be determined if one doesn't know to which dose group the slide belongs, thus shielding individual judgments from such consequences. Such blinding techniques are particularly useful to prevent the unacceptable *direct* use of non-epistemic values in the characterization of data, i.e. the seeing of something because one *wants* to see it. The argument that one ought to consider the consequences of error in making judgments should not be construed as an argument against blinding techniques used to *prevent* blatant errors. However, before such blinding techniques are used, one should be sure to consider

how successful they will be in a particular case (whether there are other clues that will offset the blinding) and whether potential errors will be distributed evenly among study groups. If the chance of errors are not distributed evenly among study groups, one is still required to consider inductive risk because the consequences of error are foreseeable. For example, if borderline cases are likely to be found only in dosed groups, then false positives or false negatives will have predictable consequences.

In the dioxin rat liver case, complete blinding is not possible with these slides. At the two higher dose levels, signs of acute liver toxicity were also visible in the slides, signs with which the pathologists are familiar. Thus, the two higher dose levels are identifiable through the microscope which pathologists use to examine the slides for cancerous lesions. And it is in these higher dose levels (10 ng/kg/day and 100 ng/kg/day) that the most cancerous growths were found and the most borderline cases were observed and evaluated. Thus, the pathologists were aware of the non-epistemic consequences of their decisions in categorizing the slides. They are thus required to consider those consequences using non-epistemic values when making the judgment call on how to categorize the slides.

This case demonstrates that there is inductive risk in how one applies categories used in data characterization and that such inductive risk can be linked to non-epistemic consequences. A tendency to either overestimate or underestimate the number of cancerous growths in these slides will have an impact on how dangerous dioxins appear. The consequences of the errors are identifiable and need to be weighed in order to determine which errors are more acceptable. In other cases, inductive risk may be present in the selection of the categories to be used as well as the application of the categories in the characterization of the data. In

addition, judgments are made in science concerning whether to keep data or whether to discard the data as unreliable. At all these decision points, there is the risk of error, and with that risk, the need to consider both the epistemic and non-epistemic consequences of error.

6. Values in the Interpretation of Results: Are there Thresholds? In the preceding sections, I have described how non-epistemic values can serve a legitimate function in making choices for methodology and evidence characterization. Once the methodology has been implemented, the data gathered and characterized, what the data means must be deciphered. Much debate and controversy among scientists has centered on the correct interpretation of the dioxin studies. In particular, prolonged debate has centered on the issue of whether the lab animal studies show that there is a threshold for dioxins' carcinogenic effects. Arguments can be made either for or against the idea of a threshold. Depending on which aspects of the evidence one chooses to emphasize or, more generally, which background assumptions one adopts, different interpretations are plausible. I will present the background assumptions and reasoning for the basic positions taken by scientists. Both interpretations I present are plausible, and there is significant uncertainty in adopting either interpretation. The uncertainty and the different consequences of adopting either interpretation create a decision context with considerable inductive risk. Because some of the relevant consequences are non-epistemic, non-epistemic values are needed to evaluate the risks of adopting either position and their accompanying background assumptions.

As noted above, animal studies are particularly important in risk regulation for helping determine the dose-response curves for chemicals. Human epidemiological studies, unable to be performed under controlled conditions for ethical reasons, rarely exhibit clear dose-

response relationships for environmental or occupational exposures. Thus, controlled animal studies have been used in place of human studies to estimate the shape of the curve for regulatory purposes. Particular debate has centered on whether certain effects, especially cancer, have a threshold. The cancer endpoint has historical significance for regulatory purposes because cancer was thought to be the most sensitive endpoint for any chemical throughout the 1970s and most of the 1980s.[10] If dioxin's carcinogenic potential has a threshold and cancer is the most sensitive endpoint, then there is an absolutely safe dose to which one can regulate, below which there will be no detrimental effects. Unfortunately, little agreement has been reached on whether such a threshold exists.

Two opposing general background assumptions lay the groundwork for the arguments over whether there is a threshold for dioxins' carcinogenic effects. For the endpoint of cancer, two competing basic assumptions in toxicology are plausible: that there is always a threshold for toxic effects and that there is no safe dose or threshold for carcinogenic effects. On one side, one can argue that a threshold for toxic effects should be the basic assumption. For toxicologists in general, one of the first maxims is "the dose determines the poison." Or as Paracelsus said more eloquently: "All substances are poisons; there is none which is not a poison. The right dose differentiates a poison and a remedy." (as quoted in Timbrell 1989, 9) Under this view, it is generally assumed that every poison has some threshold for its toxic effects; nothing is biologically potent at any dose whatsoever.

The opposing basic assumption, that no threshold exists, arises from work on cancer formation, particularly cancer due to radiation. In the 1960s, scientists came to understand that a tumor could arise from just one mutant cell, and a mutant cell could arise from just one

stray hit of radiation (e.g., one beta or alpha particle). For cancer caused by radiation, therefore, the dose does not determine the poison. Any amount of radiation, no matter how small, has the ability to cause cancer. It just isn't likely. Instead, a cell damaged by radiation is more likely to self-destruct, or to be destroyed by the immune system, if the cell threatens uncontrolled cancerous growth. But there is no threshold for radiation's ability to produce cancer. One then needs to have a curve of the likelihood of cancer plotted against the dose. An acceptable dose is determined not by a threshold but by what risk one is willing to take. This no-threshold model for radiation's carcinogenic effects has been adopted by U.S. regulatory agencies for chemicals that are mutagenic, as evidenced by the use of no threshold linear extrapolation model by agencies when performing risk assessments for mutagenic chemicals. A mutagenic chemical is one that can damage the DNA such that cancerous cells result. Mutagenic chemicals are thought to act in roughly the same way as radiation; and thus it is thought that no dose of a mutagen is "safe," just as with radiation. One is either above or below an acceptable risk level, currently set at 1 in 1,000,000 lifetime cancer risk.

These two opposing intuitions on whether there is a threshold in the dose-response curves of cancer-causing chemicals are complicated by the fact that not all cancer causing chemicals are mutagens. Some chemicals do not damage or alter the DNA, but instead promote the growth of cancerous cells once the mutations are already present. In lab tests, these chemicals will not promote cancer on their own. One must first place a mutagen into the cell culture. Then, the addition of the promoter greatly increases the incidence of cancer compared with the mutagen alone. Dioxin appears to be a very potent promoter of cancer in this way, while not being mutagenic. For the remainder of this discussion, I will assume that dioxin is a promoter only and not a mutagen. It is unclear whether we should assume that

such promoters have thresholds similar to other toxic effects or that they are more like mutagens, with no threshold for their potential damage, only decreasing likelihood with decreasing dose. Which general background assumption, that toxins always have thresholds or that carcinogens do not, should be adopted in the dioxin case is open for debate.

Against this backdrop of competing general assumptions, interpretations more specific to the dioxin case must be made. The Kociba female rat liver data, discussed in the previous section, has been central to debates over a threshold in dioxin's cancer promoting ability. Recall from the discussion above that three separate evaluations have been made of the female rat liver slides from the Kociba study. Regardless of which evaluation one chooses to consider reliable, there appears to be a significant increase in tumor rates among the rats given 10-100 ng/kg/day, whereas the animals given 1 ng/kg/day do not have a response significantly different from the control animals. (See Table 1.) Is this data evidence that there is a threshold in the ability of dioxins to produce cancer in rats? Depending on which aspects of the data one chooses to emphasize and which assumptions one adopts, different interpretations arise.

For those critical of the threshold interpretation, the issue of the statistical power of the studies is crucial. The sample size for the Kociba rat liver slides is 50 female rats at each dose level. If the dose level of 1 ng/kg/day produces cancer in 1% more of the dosed rats than the control rats, this study will not be able to detect that effect. A one in one hundred effect is not statistically detectable in the sample size of 50 rats. And whether an effect is "observed" is determined by whether an effect is statistically significant.[11] Thus, a 1% increase in cancer rate is not observable in the Kociba study. Suppose dosing the rats at 1

ng/kg/day causes an elevation in the cancer rate to this small degree. Then the "threshold" is only in our ability to observe the effects given our limited resources (only 50 female rats); it represents a limit of detection, not a limit in the activity of dioxins. It is not a threshold that can be relied upon by the regulators, who must be concerned with the 1% and even lower chance effects, when applied across a human population of millions. In sum, if there is a threshold, and that threshold is at a dose which produces fairly high rates of responses, then the threshold should be easily detectable. For example, if there is a threshold for effects and at that threshold, 10% or more of the animals will be effected, we should be able to detect this a Kociba-type study. If, on the other hand, the threshold is a low probability threshold, we won't be able to detect this with such a study. If there is no threshold, we won't be able to detect this with such a study either.

While the relatively weak statistical power of the animal studies lends support to the no-threshold interpretation, other aspects of the study bolster the threshold model. One crucial argument arises from the possibility that the rat liver tumors were caused not by dioxin tumor promotion directly, but indirectly through dioxin's acute toxic effects to the liver. As argued recently in the rat liver slide evaluation debate, the increase in liver cancer rates among the female rats could be due to the acute toxicity dioxin causes in the liver, which results in cell death in the liver, forcing the liver to attempt a rapid regeneration of cells. (Brown 1991; Goodman and Sauer 1992) At the highest dose level, all livers with cancer showed signs of acute toxicity. At the intermediate dose level, 8 of the 9 livers with confirmed cancer showed signs of acute toxicity. At the lowest dose and in the control groups, which had 1 and 2 cancers respectively, the livers showed no sign of toxicity. (See Table 1, far right column.)

Liver toxicity is generally thought to be a threshold effect, following the classic toxicology intuition that a dose determines the poison for such acute toxic effects. If one considers the liver toxicity to be a threshold effect and if the cancer found in rat livers is merely a by-product of that toxicity, it would be reasonable to presume that there is a threshold for dioxin's cancer promoting ability in the rat livers. Under this line of reasoning, one could argue that the threshold for cancers should be placed just above the level where liver toxicity ceases, i.e. that no increased risk of cancer occurs at the dose of 1 ng/kg/day. If liver toxicity causes the cancer, and is not simply concurrent with it, the cancer promotion in the liver is plausibly a threshold effect. On the other hand, one might argue that the acute liver toxicity, while perhaps assisting dioxin's tumor promoting abilities, is not the cause of the cancer. Instead, one could argue that dioxin's tumor promoting abilities are separate from the acute liver damage, pointing to evidence that dioxin promotes tumors in other animals organs as well.

In sum, one can present two opposing and plausible interpretations of the Kociba rat liver data, depending on one's choice of background assumptions and emphasis. The background assumptions that lead to the threshold position include that the toxicity seen in the livers is a likely cause of the cancers and that such cancer promotion is likely to be a threshold phenomenon, making it more plausible that an apparent threshold in the data is an actual threshold. The background assumptions that lead to the opposing viewpoint include that the statistical sensitivity of the studies is not sufficient to detect a threshold and that the link between toxicity and cancer promotion is correlation but not necessarily causation.

In making a choice between these positions, scientists must consider the consequences of their choice, particularly if they are wrong (the inductive risk). Different selections of background assumptions and interpretations of dioxin data has had a significant impact on the stringency of dioxin regulation. Countries that have assumed a threshold model have set acceptable levels of dioxins in the environment about 100X higher than countries such as the U.S. that have relied on non-threshold extrapolation models. (Finkel 1988; Greenlee *et al.* 1991) If one adopts a threshold model and one is wrong, the regulations will likely be inadequately protective of public health. If one adopts a non-threshold model and one is wrong, the regulations will likely be overly stringent. How one values these two possible errors should play an important role in determining which inductive risk one is more willing to take, and thus which assumptions one is more willing to adopt.

7. Conclusion. The three examples in the previous sections serve to illustrate how non-epistemic values play a role in the internal stages of dioxin science through consideration of inductive risk. The legitimate and required role of non-epistemic values is not a direct role, as with the constraints ethical and societal values place on the external aspects of science. Instead, it is through the consideration of the consequences of error. This means that not all scientific decisions in the internal stages of science will require the consultation of non-epistemic values. When there is very low uncertainty, such that a scientist believes there is virtually no chance of being wrong, there is little gained by considering the consequences of being wrong-- the chance of error is so small that consequences of being wrong become insignificant. This parallels everyday thinking, where, for example, the chance of being struck by a meteorite is so small, it is not worth considering the consequences of such an

event. Nor does understanding the importance of inductive risk mean that scientists should cease to attempt to reduce the chance of error. If anything, it should increase that motivation.

In addition, there are some areas of science where making a wrong choice has no impact on anything outside of that area of research. One may think, for example, of research into the coherence properties of atom beams.[12] It is very difficult to fathom how errors in such research could have non-epistemic consequences. Hence, scientists doing such research need not consider non-epistemic values. However, this type of research is rapidly becoming a minority in modern science, as most funding goes toward "applied" research and funding that goes toward "basic" research increasingly needs to justify itself in terms of some eventual applicability or use.

Finally, there are cases where the science will likely be useful but the potential consequences of error may be difficult to foresee. This gray area would have to be debated case by case, but the fact that such a gray area exists does not negate the basic argument: that when non-epistemic consequences of error can be foreseen, non-epistemic values are a necessary part of scientific reasoning. What this means for the "objectivity" of science must await future work. As a closing comment, it must be noted however, that, as Hempel pointed out, the argument "I want X to be true, therefore X is true" remains a bad argument, both within and without science.

References

Brown, W. Ray (1991), "Implication of the Reexamination of the Liver Sections from the TCDD Chronic Rat Bioassay", in Michael Gallo, Robert J. Scheuplein and Kees A. Van der Heijden, (ed.), *Biological Basis for Risk Assessment of Dioxins and Related Compounds*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 13-26.

Churchman, C. West (1948), "Statistics, Pragmatics, and Induction", *Philosophy of Science* 15: 249-268.

———(1956), "Science and Decision-Making", *Philosophy of Science* 22: 247-249.

Cranor, Carl F. (1993), *Regulating Toxic Substances: A Philosophy of Science and the Law*. New York: Oxford University Press.

Environmental Protection Agency (1994), *Health Assessment Document for 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) and Related Compounds*. Washington, D.C.: U.S. Environmental Protection Agency.

Finkel, Adam M. (1988), "Dioxin: Are We Safer Now Than Before?", *Risk Analysis* 8: 161-165.

Goodman, Dawn G., and Robert M. Sauer (1992), "Hepatoxicity and Carcinogenicity in Female Sprague-Dawley Rats Treated with 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin (TCDD): A Pathology Working Group Reevaluation", *Regulating Toxicology and Pharmacology* 15: 245-252.

Graham, John D., and Lorenz Rhomberg (1996), "How Risks are Identified and Assessed", *The Annals of the American Academy of Political and Social Science* 545: 15-24.

Greenlee, William F., Melvin E. Anderson, and George W. Lucier (1991), "A Perspective on Biologically-Based Approaches to Dioxin Risk Assessment", *Risk Analysis* 11: 565-568.

Hempel, Carl G. (1965), "Science and Human Values", in *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*. New York: The Free Press, 81-96.

Huff, J. E., A. G. Salmon, N. K. Hooper, and L. Zeise (1991), "Long-Term Carcinogenesis Studies on 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin and Hexachlorodibenzo-*p*-dioxins", *Cell Biology and Toxicology* 7: 67-94.

Jeffrey, Richard C. (1956), "Valuation and Acceptance of Scientific Hypotheses", *Philosophy of Science* 22: 237-246.

Kociba, Richard (1991), "Rodent Bioassays for Assessing Chronic Toxicity and Carcinogenic Potential of TCDD", in Michael Gallo, Robert J. Scheuplein and Kees A. Van der Heijden, (ed.), *Biological Basis for Risk Assessment of Dioxins and Related Compounds*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 3-12.

Kociba, R. J., D. G. Keyes, J. E. Beyer, R. M. Carreon, C. E. wade, D. A. Dittenber, R. P. Kalnins, L. E. Frauson, C. N. Park, S. D. Barnard, R. A. Hummel, and C. G. Humiston (1978), "Results of a Two-Year Chronic Toxicity and Oncogenicity Study of 2,3,7,8-Tetrachlorodibenzo-*p*-Dioxin in Rats", *Toxicology and Applied Pharmacology* 46: 279-303.

Kuhn, Thomas (1977), "Objectivity, Value, and Theory Choice", in *Metahistorical Studies*, 320-339.

Levi, Isaac (1962), "On the Seriousness of Mistakes", *Philosophy of Science* 29: 47-65.

Longino, Helen E. (1990), *Science as Social Knowledge; Values and Objectivity in Scientific Inquiry*. Princeton: Princeton University Press

——— (1996), "Cognitive and Non-Cognitive Values in Science: Rethinking the Dichotomy", in Lynn Hankinson Nelson and Jack Nelson, (ed.), *Feminism, Science, and the Philosophy of Science*. Dordrecht: Kluwer, 39-58.

Machamer, Peter and Heather Douglas (1999), "Cognitive and Social Values," *Science and Education* 8: 45-54.

Maronpot, Robert R., Charles A. Mongomery Jr., Gary A Boorman, and Ernest E. McConnell (1986), "National Toxicology Program Nomenclature for Hepatoproliferative Lesions of Rats", *Toxicologic Pathology* 14: 263-273.

McMullin, Ernan (1983), "Values in Science", in Peter D. Asquith and Thomas Nickles, (ed.), *Proceedings of the 1982 Biennial Meeting of the Philosophy of Science Association, Volume 1*. East Lansing: Philosophy of Science Association, 3-28.

Rooney, Phyllis (1992), "On Values in Science: Is the Epistemic/Non-Epistemic Distinction Useful?", in David Hull, Micky Forbes, and Kathleen Okruhlik, (ed.), *Proceedings of the 1992 Biennial Meeting of the Philosophy of Science Association, Volume 2*. East Lansing: Philosophy of Science Association, 13-22.

Rudner, Richard (1953), "The Scientist Qua Scientist Makes Value Judgments", *Philosophy of Science* 20: 1-6.

Timbrell, John A. (1989), *Introduction to Toxicology*. New York: Taylor and Francis.

FOOTNOTES

[1] By "dioxin" I will be referring to the most toxic congener of the class of chemicals known as dioxins, 2,3,7,8-tetrochlorodibenzo-*p*- dioxin (or 2,3,7,8-TCDD), which is also the best studied.

[2] The case studies on dioxin are drawn from my dissertation, *The Use of Science in Policy-making: a Study of Values in Dioxin Science*, The University of Pittsburgh, 1998. I thank Donald Mattison for assisting me with understanding dioxin studies.

[3] Thanks to John Beatty for making aware of this work.

[4] My thanks to Eric Angner for bringing this article to my attention.

[5] A point made by Jeffrey (1956, 242).

[6] Levi (1962) emphasized the use of epistemic values to the exclusion of all others.

[7] Churchman (1956, 247) recognizes the many places where scientists make decisions. My work can be seen as a clarification and further development of his views.

[8] Longino (1990, 86 and 128-132) also grapples with the role of values in the internal processes of science, but she does not argue that non-epistemic values are *required* of internal science. Instead, she argues that values can affect science through background assumptions, a non-normative argument. Nor does her account make clear how background assumptions, ostensibly epistemic statements, carry ethical or societal values with them. This gap spurred me to deeper consideration of the problem.

[9] In *Regulating Toxic Substances*, Carl Cranor argues that the standard assumptions have not been serving us well, leaving us with too many false negatives compared to false positives. He estimates that false negatives are in fact more costly to society than false positives. (see Cranor 1993, 71-78, 122-129, 135-137, and 153-157)

[10] "Most sensitive endpoint" means that if you regulate to prevent or acceptably reduce risk from that endpoint (e.g. cancer), all other toxic effects will also be adequately prevented. Whether cancer is the most sensitive endpoint for dioxins is a matter of debate.

[11] Choice of a level for statistical significance is discussed above. In these studies, "significant" results must be significant to the 95% level, i.e. the false positive rate will be 5% or less.

[12] I thank my colleague at the University of Puget Sound, Greg Elliott, for this example.