Scientific Consensus and Expert Testimony in Courts
Lessons from the Bendectin Litigation

Boaz Miller

Forthcoming in *Foundations of Science*

Abstract A consensus in a scientific community is often used as a resource for making informed public-policy decisions and deciding between rival expert testimonies in legal trials. This paper contains a social-epistemic analysis of the high-profile Bendectin drug controversy, which was decided in the courtroom inter alia by deference to a scientific consensus about the safety of Bendectin. Drawing on my previously developed account of knowledge-based consensus, I argue that the consensus in this case was not knowledge based, hence courts' deference to it was not epistemically justified. I draw sceptical lessons from this analysis regarding the value of scientific consensus as a desirable and reliable means of resolving scientific controversies in public life.

Keywords Expert Testimony · Consensus · Science and Policy · Science and Technology Studies (STS) · Social Epistemology

1. Introduction

A consensus in a community of researchers is commonly perceived and treated as a mark of knowledge shared by its members. It is argued that because scientists as individuals and science as socially organized and institutionalized inquiry strive at the truth, a scientific consensual view, when such exists, is the best available source to which a non-specialist can turn to gain knowledge on a disputed question (Jones 2002; Anderson 2011). Scientific consensus is therefore widely deferred to when adjudicating expert disagreement in legal trials and public policy arenas. By deference, I mean inferring that a view is correct from the mere fact that there is a scientific consensus over it, namely, regarding the existence of a scientific consensus over a certain view as a good and independent reason to believe that this view is correct over and above the reasons the parties to the consensus have to believe it is correct.

There are, however, good reasons to doubt the epistemic value of practices of blind or uncritical deference to consensus. At times, experts reach an artificial consensus to advance collective non-epistemic goals, such as the effective dissemination of policy by means of creating the mere appearance of objectivity (Solomon 2007). Likewise, experts occasionally decide to mask existing disagreements within them and speak in one voice. This may be because they paternalistically decide that it is better for the public that they show a unified front; they might worry that their social status might be undermined if disagreements among them became public; or they may wish to gain material support. To promote the group's collective interests, dissenters may refrain from voicing their own views or even express views that they do not personally hold (Beatty 2006). There are also cases in which a consensus is reached when one group within the relevant scientific community that supports a certain position overpowers other groups, unjustifiably silencing opposition to its view.

In this paper, I focus on the case study of the controversy in U.S. courts in the 1980's and early 1990's over the safety of the drug Bendectin in order to give a normative critique of the practice of deferring to scientific consensus for arbitrating between rival expert testimonies in courts. The Bendectin controversy was a high-profile debate, which eventually led to the U.S. Supreme Court's Daubert (1993) decision which established new rules for the admissibility of scientific expert testimony. The controversy was settled partly by courts' deferring to the scientific consensus that emerged in the late 1980's. I argue that this deference to consensus was, in fact, epistemically unjustified. I draw general lessons from this example with respect to limited value of scientific consensus in resolving legal and policy-related disputes.

My analysis of the case study heavily draws on my theoretical account of knowledge-based consensus (Miller 2013). The term "knowledge-based consensus" denotes a consensus that is grounded in knowledge shared by its members, as opposed to mere agreement, which is not necessarily of epistemic value. It identifies three conditions a consensus needs to meet to qualify as knowledge-based, which pertain both to the evidential support the consensual view enjoys, as they manifest themselves in the epistemic community, and the characters of the social agreement around it. The application of my theory to the Bendectin case study reveals that the consensus does not satisfy the conditions for being knowledge based. Hence, I argue that it could not legitimately serve as a resource in arbitrating the dispute in court. I therefore advise general caution in deferring to a scientific consensus in order to decide legal and policy-related disputes.

This paper consists of seven sections. In the following Section 2, I review the state of literature in social epistemology about expert testimony, particularly the problem of adjudicating expert disagreement. In Section 3, I give a brief overview of my account of knowledge-based consensus, which constitutes a novel approach for dealing with some cases of expert disagreement. In Section 4, I provide the essential historical and legal background needed for understanding the Bendectin controversy. In Section 5, I discuss the Bendectin case study and argue that the consensus in this case was not knowledge based. In Section 6, I discuss the meaning and of this example and its negative implications to the practice of deferring to a consensus in order to resolve legal and policy-related disputes.

2. The Social Epistemology of Expert Testimony

Over the past several decades, courts have been increasingly deferring to scientific expert testimony and scientific evidence. For example, in torts trials, courts defer to experts to determine whether a specific toxic substance caused the plaintiff's illness, and in criminal trials, courts rely on forensic evidence such as DNA profiling to determine the defendant's guilt.[1] Often, courts are faced with expert disagreement. In such cases, how can a scientifically lay judge or jury evaluate scientific expert testimony or evidence and reach epistemically warranted decisions based on it? How can they assess its reliability? After all, the court defers to the experts precisely because it lacks the competence to decide the

[1] For accounts of expert testimony in different legal contexts see Golan (2004); Adler (2007); Cole (2001); Aronson (2007); and Lynch et al. (2009).

matter on its own.

Social epistemologists have addressed the problem of adjudicating between rival experts. Some have resorted to scepticism. For example, Frances (2005) argues that in cases of disagreement between experts, a lay person has no sufficient grounds for preferring the view of one expert over another, and therefore cannot have knowledge of the disputed matter. He further argues that the traditional solutions in philosophy to scepticism, which aim at dismissing the sceptic, do not apply to this new scepticism, which is live and cannot be easily dismissed. Brewer (1998, pp. 1617-1671) makes a similar argument in the context of scientific expert testimony in courts. He argues that in many cases, a scientifically lay judge or juror does not have sufficient epistemic competence to evaluate the content of an expert's testimony. Thus, when she faces rival experts who have seemingly equivalent credentials, she does not have a non-arbitrary basis to prefer the testimony of one expert over another. Because factual legal judgments should not be arbitrary, Brewer argues that expert testimony may pose a major problem to due process, which a core value of the legal system.

Other philosophers are more optimistic. Goldman (2001) describes several heuristics that can help a layperson decide between competing experts, such as examining how well the experts react to criticism, examining their credentials, evaluating their past "track-records", and evaluating evidence about the experts' interests and biases. Goldman's approach is further developed by Matheson (2005) and Gelfert (2011a), who argue that laypeople may sometimes be able to assess expert testimony by evaluating the experts' dialectical performance. With respect to expert disagreement in courts, Goldman (1999, pp. 304-311) recommends that courts in the adversarial legal system, in which each side brings its own experts, rely more on court-appointed independent experts and expert panels, who are less biased. Similarly, Haack (2004, pp. 24-25) proposes that courts in the U.S., which adhere to the adversarial structure, should use more often their statutory power to appoint independent scientific experts, similar to the common practice in inquisitorial legal systems, like Germany and the Netherlands.

Such strategies and recommendations may prove useful, especially in controversies that are artificially blown-up for the purpose of the trial, but they have their downsides as well. As Haack (2004, pp. 24-25) herself points out, her suggestion to rely more on court-appointed experts implies relinquishment of discretion on part of the judge, and thus raises normative worries as to whether it is democratic to hand judicial authority to experts. In addition, as Jasanoff (1995, pp. 55-57) argues, this suggestion overlooks the positive critical role of cross examination of experts in the adversarial system. In the U.S., for instance, experts on behalf of defendants in criminal trials exposed serious defects in the DNA profiling techniques used by the state, and it is doubtful whether court-appointed experts would have done the same. Furthermore, independent advisers to the court are rarely free of bias. This is because of what is known as the "expert-paradox": "precisely what qualifies certain individuals to serve as advisers can also prevent them from objectively assessing the literature" (Wolf 2007, p. 609). Experts have vested interests in their theories. Their professional development, funding, and reputation often depend on the success and acceptance of a specific view or theory (Pickering 1982). It is hard to find experts who are truly objective, especially in cases of genuine controversies. Last, for the most part, the above suggestions do not break new ground in the *practical* problem of expert disagreement, but rather explain and clarify the epistemic rationales underlying heuristics that are already being used to adjudicate between rival.

In sum, existing attempts by philosophers to deal with the problem of adjudicating between rival experts have shown limited success. They offer good reasons to reject radical scepticism about laypeople's ability to justifiably adjudicate expert testimony, but they do not significantly enhance the layperson's epistemic arsenal. In the next section I review my account of knowledge-based consensus, which introduces a new tool to this epistemic arsenal. It specifies determinable conditions under which a consensus in an epistemic community is based on knowledge shared by its members, rather than mere agreement. When a consensus is knowledge-based, the court can legitimately defer to it to adjudicate scientific expert testimony.

3. Knowledge-Based Consensus

My account of knowledge-based consensus aims at finding conditions a consensus needs to meet to qualify as knowledge based. Elsewhere (Miller 2013) I give a fuller explication and defence of it. Within this paper, I will only review its main features. Let us assume a situation in which a person, e.g., a policy maker, is faced with a scientific consensus over a certain matter, and wants to determine if she may legitimately defer to this consensus. That is, this person wants to know if she may legitimately infer that a certain position is correct from the mere fact that there is a consensus over it. By assumption, this person has no other way to establish the truth or falsehood of the consensual view.

Note that the judgment of whether a consensus is knowledge-based is a *social* judgment, which judges and juries are epistemically qualified to make. It is not a scientific judgment. The judgment of whether a consensus is knowledge-based is about whether a certain state of affairs exists in a scientific community. It is a second-order judgment about the state of a science, rather than a first-order judgment about the subject matter of the science. Interpreting social reality is part of the routine work of judges and juries and their capacity to do so is assumed by the legislature. Common law contains vague or abstract legal terms such as "the reasonable person", "good faith", or "reasonable care". These expressions reflect standards of conduct or appropriate behaviour. These terms are not explicitly defined in the legislative corpus. Even if they have a specific legal meaning, it is still parasitic on common meanings. Their meaning is contextual and depends on general cultural norms and the specific circumstances of the case at hand. The meaning of such terms may also change over time (Pound 1965, pp. 57-58). Such terms assume that judges are competent and qualified to interpret social reality and make social judgments.

My account starts by investigating the meaning of "knowledge-based consensus". It draws on the definition of knowledge in standard philosophical theories of knowledge. Most common definitions of knowledge are based on some variant of a conception of knowledge as justified true belief. How are we to understand this conception in the context of consensus, particularly scientific consensus? Let us start with truth. The notion of truth simpliciter does not adequately characterize *scientific* knowledge. Rather, it is better to adopt the notion of partial or approximate truth, because even our best scientific knowledge is never perfect or complete. It is constantly improving, evolving, and being corrected.

What about belief? How is consensus to be characterized in terms of belief? In what sense does a group of people share a belief as a group? According to Gilbert's influential account of group belief, which my account adopts, members of a population share a collective belief if and only if they are jointly committed to believing it *as a body* (2002, p. 42). This account does not require that members of the population individually have the relevant belief. Rather, it means, roughly, that they have agreed to let the content of the shared belief stand as the position of the group. The joint commitment entails that they endorse this position when participating in group activities, and publicly defend it when acting as group representatives. Gilbert's account of collective belief can accommodate a wide spectrum of agreements, varying in strength and explicitness. It can accommodate a joint commitment arrived at by an explicit process, such as a written consensus statement or verbal agreement (1987, p. 124-125), or an implicit agreement which is perceived as binding by the members of the group (Gilbert 1994).

A related question is the scope of agreement that is required to say that a consensus exists. Tucker (2003, pp. 509-510) argues that agreement amounts to consensus only when it is complete and there is no dissent whatsoever. I reject this view. In real-life cases, complete agreement is hardly ever present, and it is doubtful whether any wide agreement in the history of science has ever been complete. The mere existence of dissent is not enough for discounting certain wide agreements as genuine consensuses. For example, it is widely believed today that the HIV virus causes AIDS. There are few dissenting scientists who disagree, and their view is dismissed as groundless by the vast majority in the scientific community. The belief that HIV causes AIDS constitutes a basis for research, treatment, and prevention. It is taught and presented as a fact in university curricula and to the public. The dissenters are suppressed and marginalized by the mainstream scientific community, and face extreme difficulties challenging the majority view. Though a pocket of dissent exists, it is fair to say that there is a consensus that HIV causes AIDS.

What about justification? A common way to understand the justification condition in modern theories of knowledge is as ensuring that when the belief is true, the relation between the belief its truth is not accidental or merely lucky.[2] Pritchard (2005, p. 146) calls the kind of luck that a belief needs to exclude to count as knowledge "veritic epistemic luck". For example, if I form a belief by looking at a broken watch that happens to show the right time, although my belief is true and arguably justified, it is not knowledge. It is just a fluke. A knowledge-based consensus cannot be vertically lucky. That is, a consensus is not knowledge-based if its members just happed to reach the truth.

There is, however, another type of epistemic luck, the elimination of which is central to any account of knowledge. I call it *epistemic misfortune*. It refers to circumstances in which agents are justified or seemingly justified in their views but there are factors that

[2] This anti-luck condition is also meant to exclude so-called Gettier cases, which have been a major concern in epistemology in recent decades.

systematically and/or deliberately mislead them or inhibit their gaining knowledge. What distinguishes a mere bias from epistemic misfortune is that in epistemic misfortune, left to their own devices, agents' chances of correcting their misguided views are slim to null. Sometimes careful reflection may lead agents to realize that their views are biased, for example by sexist or racial prejudice, and lead them to revise them accordingly. However, if the epistemic misfortune is genuine, they will not be motivated or disposed to perform such reflection, and even when they are, they may fail. To emphasize the difference between veritic epistemic luck and epistemic misfortune, veritic epistemic luck is meant to exclude situations in which the belief is true, but this is only due to mere luck, while epistemic misfortune is meant to exclude cases in which the belief is false, but the subjects who hold it are stuck in a situation in which it seems true to them, and their chances of getting themselves out of this situation are slim.

While in epistemic misfortune, the parties to the consensus are victims of the circumstances, sometimes they are less innocent – they deliberately form a consensus despite their awareness that the consensual view falls short of knowledge. These are cases of *non-cognitive consensus,* which knowledge-based consensus also aims to exclude. They are not knowledge based because people are likely to form the consensus whether the consensual view is true or not. People form a non-cognitive consensus for many reasons. Consensus is a powerful political tool for advancing policy and promoting social aims. A group of experts may reach a consensus and mask existing disagreements within it to advance a goal. They might worry that their social status might be undermined if disagreements among them became public; they may wish to gain material support; etc. To promote the group's collective interests, dissenters may refrain from voicing their own views or even express views that they do not hold outside the expert community and even inside it (Beatty 2006, pp. 53-56).

To assess whether a situation of knowledge obtains, we need to eliminate veritic luck, epistemic misfortune and non-cognitive reasons as possible good explanations of the consensus. That is, a consensus is knowledge based when knowledge is the best explanation thereof. I suggest three identifiable conditions that a consensus must jointly meet for us being able to infer knowledge as its best explanation.

The *first* condition is that the agreement must be genuine. Roughly, scientists must give the same meaning to the same terms, share the same fundamental background assumptions, etc. Let us call this condition *social calibration*. Sometimes an alleged agreement exists, but a closer look reveals that it is loose, superficial, set around vague terms that are susceptible to multiple interpretations, and hence not genuine (Halfon 2006). Such an alleged agreement is not a putative knowledge-based consensus. To avoid such a scenario, agents must agree on some minimal content. The minimal requirement is that scientists share agreements of three types, which are the essential elements of a Kuhnian paradigm (Kuhn 1970, pp. 182-191). The first is *shared formalism* (or in Kuhn's terms, "symbolic generalizations"), for example, $f=ma$ in Newtonian physics and $2H_2O \rightarrow 2H_2 + O_2$ in analytic chemistry. The second type is *ontological schemes* ("metaphysical models") which are descriptions or models of the building blocks and furniture of the world, such as that matter is composed of particles. The third is evidential standards ("exemplars") which are model solutions that show how to apply the formalism to solve specific problems and define what solutions are acceptable.

Such a minimal agreement is sufficient because mature formalisms constitute the *syntax and minimal semantics* required to calibrate collective research. A commitment to a formalism is *ipso facto* a commitment to accepting derivations and inferences done with it. It constitutes a rigid set of rules of manipulation of symbols, derivation, and inference (Gelfert 2011b). At the same time, members to the consensus may explicitly diverge in their metaphysical views about the reality of objects populating the ontological schemes they use.

According to the *second* condition, a knowledge-based consensus needs to be epistemically robust. That is, it needs to exhibit apparent consilience of different lines of evidence,[3] such as both animal studies and human studies. Producing evidence using multiple techniques and under different background assumptions aims at eliminating influences that are *accidental to the particular way* a hypothesis is tested. For example, if the same pattern is observed when the same sample is placed in different types of microscopes, it is likely that the observed pattern is accurate, rather than a by product of the particular way a certain microscope operates (Stegenga 2009). The more a consensus exhibits such apparent consilience, the more it is likely to be knowledge-based, and less likely to be a mere accidental by product of a particular technique.

The *third* condition requires that the consensus be socially diverse. This is another aspect of the robustness condition discussed in the previous paragraph. In social epistemology, the robustness condition is often cashed out in terms of social diversity, as diverse people tend to have different background assumptions (Longino 2002, p. 131). The rationale behind the diversity condition is eliminating veritic epistemic luck. If for a putatively successful theory *T*, it very easily could have been that had we thought of an alternative *T**, we would not have accepted *T*, then if *T* is true, we are lucky to have accepted it. Diversity increases the number of alternatives, hence if *T* is true, it is not merely veritically lucky. Diversity controls for two types of influences – expected and unexpected, just as randomization in clinical trials controls for both known confounders (sex, age, etc.)

[3] One may wonder why only apparent consilience of evidence is required, rather than actual consilience. This is because actual consilience is too demanding a condition. The difficulty with determining whether the evidence is actually consilient, as opposed to seemingly consilient is threefold. First*, de facto* multiple methods for combining and weighing different lines of evidence are available and may generate different outcomes for the same body of evidence (Douglas 2012; Stegenga 2011). Second, even if there were only one widely agreed upon method for combining and weighing evidence, appealing to it for answering the question of when consensus is knowledge based merely because it enjoys a wide consensus would be question-begging, but all existing methods have their pros and cons. Third, all our current best methods for determining evidential support leave room for subjectivity (Douven 2009, 347-350; Douglas 2012).

and unknown confounders.

The third condition of diversity is an independent condition from the second condition of apparent consilience of evidence. Think about the following case. You find out that a consensus exists that passive smoking does not raise the chances of lung cancer, and this consensus exhibits consilience of different lines of evidence – studies of different types support this conclusion. Then you discover that all parties to the consensus have some financial ties to the tobacco industry. Will you still think that knowledge is the best explanation of that consensus? Probably not. Hence, we should separately address the question of whether all lines of evidence seem to support the consensual view, and whether the consensus is socially diverse.

To sum up this section, my account addresses the question of when a consensus is knowledge based. It identifies three conditions under which this is likely to be the case: social calibration, apparent consilience of evidence, and social diversity. It is important to stress that the fact that a consensus is not knowledge based does not necessarily mean that the consensual view is false or unwarranted. Rather, it means that the fact that an agreement exists does not carry additional credence for the view that is being agreed upon. It does not give us a reason to infer that it is true from the mere fact that a consensus exists over it. Only when there is evidence that a consensus is knowledge based, the fact that the consensus exists gives us an independent reason to justifiably believe that it is correct. Only then, is deference to a view because it is a consensual view justified. In the next two sections, I argue that the consensus over the safety of Bendectin was not knowledge based, hence the court's deference to it was not justified.

4. <u>The Bendectin Trials – General Background</u>

In the 1980's and early 1990's, there was a series of mass tort trials in U.S. Federal Courts involving birth defects allegedly caused by a drug called Bendectin, which was manufactured by Merrell. At that time, the scientific community was divided on the question of whether Bendectin could cause birth defects in human babies. From the mid 1980's, however, a scientific consensus emerged that it could not. This consensus was used *inter alia* as a resource for deciding the case in courts, under the assumption that a scientific consensus amounts to knowledge.

Several sources give thorough and detailed accounts of the history of the Bendectin litigation (Green 1996; Sanders 1998; Edmond & Mercer 2000). My aim in this paper is not to provide an alternative historical account, but rather to provide an alternative analysis of the role and epistemic significance of the scientific consensus that emerged in this case. In this section, I therefore provide only the essential background that is relevant to the issue at hand, rather than a comprehensive historical account.

The Bendectin trials were mass tort trials, namely civil actions involving numerous plaintiffs against one or a few corporate defendants. In the U.S. federal legal system, where the Bendectin trials were conducted, there are three instances. The lower instance consists of 94 district courts, which are trial courts. A higher instance consists of 13 circuit courts, which are appeal courts. Their decisions set binding precedents for district courts in their circuit, but not for other circuits. The highest instance is the U.S. Supreme Court, which is an appeal court, the decisions of which set binding precedents to all lower courts.

In tort cases, in establishing a legal case for compensation, a plaintiff must show that the defendant had a legal duty to prevent harm, the defendant violated it, the plaintiff suffered a legally compensable injury, and the defendant's action was the factual and legal cause of the injury (Cranor 2005, pp. 143-144). Toxic torts, i.e., torts cases that deal with injuries from allegedly harmful substances, pose unique epistemic problems in proving causation. The causal mechanism behind the injury is often highly complex and poorly understood. There is often long latency period between the exposure and the injury. Most injuries are the result of long-term exposure to an agent at low doses (Parascandola 1997, p. 147). The subject is often exposed to more that one chemical agent, sometimes produced by different parties, and it is difficult to determine which agent or combination of agents is responsible for the injury. Toxic agents do not always cause harm. They cause harm only at a certain proportion of instances. Some conditions are not uniquely associated with one chemical agent, and sometime occur without exposure to known agents (Simon 1992, p. 35).

Generally speaking, prior to 1993, the legal test for admissibility of scientific evidence or expert testimony was that it was "generally accepted in the relevant scientific community", – a criterion established in the *Frye* (1923) decision – and helpful to juries in their fact-finding mission. Judges tended to permit experts to testify and let cross-examination during trial determine whose experts the jury believed. Cross-examination in front of a jury was the main screen for expert testimony. In the Bendectin litigation, starting from the mid 1980's, however, judges started making admissibility tests stricter – and in the *Daubert* (1993) decision, which was one of the Bendectin cases, the Supreme Court set new standards for admitting scientific evidence. In *Daubert*, the Supreme Court determined that judges should act as "gatekeepers" and admit only "reliable evidence", which it took to be "good science" that was produced by the application of "the scientific method". They should also admit only evidence that is "relevant", i.e., fits the particular case. According to *Daubert*, judges should keep "bad science" out, as it is unreliable. Daubert recognized four indicators of good science: falsifiability, publication in peer-reviewed journals, known chance of error, and general acceptance within the scientific community.[4]

De facto, *Daubert* institutionalized the stricter admissibility tests and procedures that had developed in the course of the Bendectin litigation. Today, a "Daubert" review of

[4] For a critical discussion of the *Daubert* decision and its flawed understanding of what it took to be the relevant issues in epistemology and philosophy of science see Haack (2003; 2004; 2005).

experts is a regular procedure in trials of this sort, and it occurs early in the timeline of events leading to a trial. After the initial complaints and answers have been filed, and after discovery, during pretrial hearings, a judge reviews whether the experts will be permitted to testify. Experts for both sides would typically submit reports that provided the basis of their testimony, which would be subject to review by the opponents and by the court. Often one side or the other will file a *Daubert* motion to have the others' litigants excluded from testifying. If an expert critical to a litigant's case is not admitted, the litigant (typically the plaintiff) may be unable to establish needed factual premises, in which case the judge can dismiss the attempted legal action by means of a summary judgment because there is no material issue of fact for the jury to decide. Thus, a preliminary review of whether one's experts may testify can result in dismissal of the case without a trial (Cranor 2005, pp. 143-145).

The Bendectin trials took place from the early 1980's until the mid 1990's. While in the early 1980's district courts tended to rule for the plaintiffs, from the mid 1980's they started to rule against the plaintiffs on account of not showing causation. Summary judgements became more common, and circuit courts tended to reverse previous district courts' jury decision in favour of the plaintiffs. While this was the main trend, because courts in each circuit can rule independently of one another, there were also exceptions. The last Bendectin case was the *Oxendine* (1996) decision, in which the D.C. Superior Court (an appeal court) reversed a previous decision and ruled in favour of Merrell.[5]

The change in courts' views from accepting the plaintiffs' claims in the early cases to rejecting them later on was justified *inter alia* by an emerging scientific consensus that epidemiological studies in humans are the most reliable indicator for detecting causation between Bendectin and birth defects, and that such studies do not indicate such causation.

[5] See a table of the Bendectin cases and their outcomes in Edmond & Mercer (2000) at 304-5.

In this case, the consensus that emerged in the scientific community was used to settle a scientific dispute in court. As I have argued, however, the mere existence of consensus, even among scientists, is not necessarily indicative of shared knowledge. Thus, a question arises regarding whether the consensus in this case was in fact knowledge based. Assuming that deference to a scientific consensus to settle a scientific dispute in court is justified only when the consensus is knowledge based, another way to put this question is to ask whether the deference to the consensus was justified in this case. In the next section I answer this question in the negative

5. Was the Consensus on Bendectin Knowledge Based?

According to Huber's polemical account (1991, Ch. 7), before the scientific consensus emerged, scientists studying Bendectin and testifying to their findings in courts were doing legitimate science. However, after the consensus emerged, scientists still researching Bendectin or testifying in court about its alleged harmful effects were motivated by greed and no longer doing legitimate science, but junk science. The problem with this account is that the most reasonable way to interpret the term "junk science" for Huber is "science no longer deemed acceptable by the majority of the scientific community". Unless we have independent reasons to think the majority must always be right, however, Huber's claim becomes circular: The scientists' testimony should not have been accepted because it was junk science, and it was junk science because it was no longer acceptable. With respect to greed, as I will show, it cannot be categorically stated that scientists who had remaining concerns about the safety of Bendectin were all motivated by greed. We should also keep in mind that while some scientists may have been motivated by greed, Merrell had its own financial interests and resources to promote research that pointed to the opposite conclusion and suppress research that did not.

Green (1996) and Sanders (1998) both offer non-polemical and more nuanced accounts. Similarly, however, they both argue that the deference to the emerging scientific consensus by the courts in order to settle the legal controversy was justified, without addressing the question of whether the consensus was knowledge based or not. Green praises the legal system for eventually ruling correctly relying on the accumulating reliable epidemiological evidence that indicated lack of causation (1996, pp. 21-22; 314-316). He does not explain, however, why courts were entitled to determine that the epidemiological evidence takes precedence over other kind of evidence that pointed to the other direction, where this was one of the main question on which the experts were divided.

Edmond and Mercer (2000) dispute these claims, and provide a symmetrical analysis of the events; that is, an analysis that does not favour one of the sides to the debate, and does not assume that the party to the debate that eventually won had been correct all along. They argue that the question of whether epidemiological evidence is the most reliable way of indicating causation and how it should be amalgamated with other types of evidence was one of the questions courts were required to decide. Thus, they argue, Green retroactively legitimates the courts' decisions by the standards they eventually reached. Furthermore, Edmond and Mercer argue that courts played a significant instrumental role in bringing about the scientific consensus about the supremacy of epidemiological standards, on which they later relied as a normative epistemic source (Edmond & Mercer 2000, pp. 272-273).

According to Edmond and Mercer, Sanders similarly offers a Whiggish reconstruction of the events, assuming that the final conclusion lay dormant in the evidence, waiting to be discovered by the courts. They criticize both the courts for resting on an inadequate ideal of science as a value-free enterprise that produces knowledge that is independent of social influences and interests, in particular by the courts themselves, and Sanders' account of the trials as a "sociology of error" that assumes the correctness of the final outcome reached and works backwards to explain decisions leading to it in terms of rationality and decisions that stood in the way in terms of biases and social interests. As an alternative, they offer a symmetrical reconstruction of the events:

> …care should be made to attempt to document and articulate some of the key processes through which the primacy of epidemiology as an integral part of the closure of the Bendectin litigation emerged. Rather than assume that this conclusion somehow lay dormant ready to be discovered once the dust of legal distortions had cleared, we will document how the context of litigation helped constitute the outcome (Edmond & Mercer 2000, p. 282).

Though they offer a more balanced account, in my view, that accords better with the events, Edmond and Mercer's analysis leaves us dissatisfied. While they are right, as I will argue, to analyze the events in terms of negotiations and mutual social construction of the outcome by the scientists and the courts, they refrain from passing a normative judgement on this construction. In the social context of mass torts in the U.S., *any* final outcome reached would have been a social construction negotiated by courts and scientists. However, not any outcome reached is equally epistemically justified. In particular, the mere fact that courts played an active role in bringing about the scientific consensus *in itself* does not necessarily mean that it was not knowledge based.

My analysis aims at taking Edmond and Mercer's account one step further, namely not only pointing out the contingent circumstances and events that brought about a particular consensus rather than another, but also determine whether it was knowledge based. In what follows, I will present Sanders' depiction of the events leading to the consensus. As I will show, Sanders' depiction of the events actually supports Edmond and Mercer's analysis rather than his own, as he seems to downplay the role of non-epistemic factors in bringing about the consensus. I will also show why the consensus that eventually emerged cannot be said to be knowledge based.

Four main types of evidence were used in the Bendectin trials: structure-activity (chemical studies), in vitro (experiments on single cells, organs, tissue samples, etc.), in vivo (animal studies), and epidemiology (human studies). Each of these types of studies corresponds to a different line of evidence. Each type of evidence belongs to what Hacking (2002, pp. 159-199) calls "a style of scientific reasoning" – one of seven fundamental ways of reasoning he identifies as constitutive of Western science. This information is summarized in Table 1.[6]

Evidence Type	Scientific Discipline	Style(s) of Reasoning	Benefits	Drawbacks
Structural Activity	Medicinal Chemistry	Analogical Model	• Cost • Facilitate explanation and understanding	• Prone to error – very similar molecules sometimes have very different effects.
In Vitro	Toxicology	Laboratory Style	• Cost • Possibility of cross species extrapolation by comparing human and animal samples	• Chemicals may behave differently in the living body than in the test tube due to their accumulation only in certain organs or being rapidly metabolized into different substances.
In Vivo	Toxicology	Laboratory & Statistical Styles	• More accurate knowledge and control of the dose than human studies • Better control of confounding factors than human studies	• To reduce costs, trials are done on a small population of animals and so the dose is increased, but at very large doses almost every substance becomes harmful. • Different physiology in humans and animals – some chemicals are harmful to humans and not animals and vice versa. • No single agreed-upon method of extrapolation to humans.
Epidemiology	Epidemiology	Statistical Style	• Direct knowledge of risk factors in humans – fewer problems for extrapolation	• Data are obtained from what people report and depends on their memory. Less reliable knowledge of what doses were taken and when. • Less control and knowledge of possible confounding factors. • Existence of an effect at very small rates is hard to identify due to insufficient sample size.

Table 1: Evidence Used in the Bendectin Trials

When talking about the validity of studies, a distinction is drawn between internal validity, which concerns inference with respect to the population or samples actually tested, and external validity, which concerns the extrapolation of the results of the study to the target population. All types of studies are prone to problems of internal and external

[6] This table relies on Sanders (1998, pp. 46-61). See there for full bibliographical references.

validity. Generally speaking, evidence belonging to the laboratory style tends to be more prone to problems of external validity, as the subjects or samples differ in relevant ways from the target population, and the laboratory settings try to create an idealized environment that shields the experiment from influences that may exist in the target population (Sanders 1998, pp. 53-60).

Another possible problem for in vivo and epidemiological studies that belong to the statistical style is the validity of the statistical conclusions. One problem is the conditions under which it is legitimate to infer causation from correlation. These conditions vary with circumstances and give rise to disputes, especially in the legal context of tort law that has its own understanding of causation.[7] Another problem is the trade off that exists in statistical studies between increasing the chances of false positives and decreasing the chance of false negatives and vice versa (Sanders 1998, pp. 50-53).

The latter issue is particularly important in the Bendectin controversy. In vivo studies of a substance on several types of animals are very effective in preventing birth defects. There have been no known new drug-induced epidemics of birth defects since the implementation of laws requiring animal testing (Sanders 1998, p. 57). While in vivo studies are very effective in preventing birth defects, they do so at the cost of a high rate of false positives, namely, they predict more birth defects in humans than there are. Sanders claims that this feature makes in vivo studies valuable in the regulatory context, which aims at prevention of birth defects, but not so much in the legal context, which aims at causally explaining existing ones (1998, p. 58). However, he seems to overlook Edmond and Mercer's point that the exact level of tolerance of the legal system to false positives was not pre-determined but constructed and negotiated during the Bendectin trials.

From the mid 1980's epidemiological studies conducted failed to replicate the

[7] See Parascandola (1997) and Simon (1992) for discussions of the relations between correlation and causation in tort law.

positive results of the previous studies. Sanders argues that within the field of epidemiology, the question whether the studies showed causation was "purely factual". He argues that

> None of the scientists involved in the Bendectin studies had tied the outcome of the research to an important theoretical issue. In this environment, the scientists were not prepared to invest the intellectual energy necessary to provide nonsubstantive explanations for replication failure. Even central players […] were prepared to interpret failure to replicate as a factual finding, and to revise their opinion accordingly (Sanders 1998, p. 86).

However, Sanders' analysis conflates two different issues. The results of the studies were not purely factual, but theory laden. For example, as he mentions, an important question in epidemiological studies was in which stage of the pregnancy the mothers took Bendectin. It was theorized that if it were a teratogen, Bendectin would work at the organogenesis stages of their pregnancy – the early stages in which internal organs develop, so women who took it later should be excluded from the population study (Sanders 1998, p. 53). This example shows that theory is involved in the design and interpretation of the study. The second claim is that epidemiologists, even those who initially claimed that Bendectin was harmful, did not have a strong vested social-cognitive interest (Pickering 1982) in a particular theory of causation, which might have motivated them to offer an alternative theoretical interpretation of the failed replications.

Sanders agrees that the controversy over the supremacy of epidemiological studies over other types of evidence was theoretical and not purely factual, but argues:

> If the in vivo studies had clearly pointed in a different direction from the epidemiology, the controversy would not have been so easily resolved. However, controversy concerning the primacy of epidemiological evidence is largely irrelevant to the Bendectin controversy. The in vivo and aggregate time trend data support the epidemiological evidence that either Bendectin is not a teratogen or is such a weak teratogen that its effects are undetectable. In vitro studies are more ambiguous, but it is generally agreed that this evidence is less probative than in vivo and epidemiological studies (Sanders 1998, p. 86).

This assessment, however, does not accord with the events as Sanders himself depicts them. The majority of epidemiological studies conducted from 1981 to 1994 concluded with a negative result (no effect observed), and a few were inconclusive (Sanders 1998, 70). By contrast, nine in vivo studies were conducted from 1981 to 1988; three concluded with a negative result, four with a positive result, and two with inconclusive results (Sanders 1998, pp. 66-7). In 1985, nine toxicologists wrote a letter to the journal *Teratology* expressing scepticism about whether the safely of Bendectin had in fact been established (Brown et al., 1985; Sanders 1998, p. 88). Two additional letters in this issue discuss the safety of Bendectin and there is clear dissent within the community of toxicologists (Brent 1985, Holmes 1985). From a social-epistemic point of view, the existence of this dissent is significant.

Moreover, as Sanders himself mentions, the in vivo studies practically stopped from the mid 1980's, not because the scientific community was satisfied that the epidemiological evidence was better, but because *courts* were:

> The direct evidence on human effects provided by the epidemiological evidence diminished the value of in vivo studies insofar as they are designed to answer the question of whether a drug causes harm to humans. Increasingly, courts simply refused to rely on or even admit in vivo evidence. *Considering the declining demand for in vivo research, it is not surprising that the supply has been limited* (Sanders 1998, p. 68; emphasis added).

In other words, when dissenting views within the toxicologist community called for more in vivo research, it was not done because courts stopped relying on it or admitting it.

It is important to bear in mind that the practice of not admitting in vivo research did not emerge in a vacuum, but in a specific social context. The U.S. legal system is a jury system, where the jury – twelve citizens with no legal background – assume the role of the trial fact finder. This practice was established in a period of growing public concerns, especially by conservatives, which were expressed *inter alia* by Foster et al. (1993) and Huber (1993), about jurors' abilities to carry out their role as fact finders in trials involving complex scientific evidence. Specifically, the problem arose in mass tort trials, such as the

Bendectin trials, where the plaintiffs claimed that their exposure to a harmful substance produced by the defendant caused their illness. In such cases, it was argued that jurors tend to sympathize with the plaintiffs, who are typically ordinary working-class citizens, and be hostile to the defendant, typically a big corporation. Thus, they tend to decide in the plaintiffs' favour based on "junk science" or irrelevant evidence presented by the plaintiffs' experts and purporting to show a causal link between the suspected substance and the plaintiffs' illness (cf. Sanders 1998, Ch. 1 & 2). Such practice by the jury, it was argued, not only averts justice, but also inhibits economic development, as it compels big corporations to spend major funds on defending themselves in legal trials and paying compensations.[8] A good explanation of the emergence of the practice of not admitting in vivo research – namely preventing it from reaching a jury, is that it was a response to such worries.

Sanders claims that "in the case of the Bendectin controversy, it seems certain that nonepistemic factors did play a role in the general consensus that formed around Bendectin" (1998, 87). He expresses great confidence that the internal mechanisms of group rationality of the scientific community can resolve controversies correctly, and regrets that the legal system lacks such mechanisms (1998, p. 116). It seems that because of this confidence, though he is aware of them, Sanders seriously downplays the importance of external influences on the scientific community and the mutual relations between it and the legal system. In particular, he downplays the role the court played in blocking further in vivo research, which Edmond and Mercer (2000, p. 276) find pivotal in constructing the decision about Bendectin. Thus, his appeal to the consensus that was formed as a justification for the court decisions is illegitimate.

My account of knowledge-based consensus supports this conclusion. The scientific consensus that Bendectin does not cause birth defects exhibits neither apparent consilience

[8] For a discussion and critique of these claims see Cranor (2005).

of different lines of evidence. Rather, it was achieved by one line of evidence associated with one dominant group within the scientific community, namely epidemiological studies conducted by epidemiologists, taking over the rest, in particular over in vivo studies conducted by toxicologists, which were still inconclusive. Since epidemiologists and toxicologists constitute different social circles, and the consensus was *de facto* only among epidemiologists, it did not exhibit social diversity as well. In other words, in this case, various social factors led the legal and scientific communities to favour the negative results obtained through epidemiological studies carried by epidemiologists over the inconclusive results that were obtained through in vivo studies, and to abandon further in vivo research in spite of dissenting calls among toxicologists to the contrary. The consensus that was formed in the mid 1980's and early 1990's was not knowledge-based. Hence, the mere existence of consensus in this case could not serve as a justification or a normative resource for courts on which to base their decisions.

6. The Wider Picture – General Lessons from the Bendectin Case Study

What are we to make of the claim that the consensus in the Bendectin case study was not knowledge based? It is important to distinguish between two questions. The *first* question is whether Bendectin causes birth defects. This is a scientific question about human physiology. The *second* question is whether the scientific consensus that Bendectin does not cause birth defects was knowledge based. This is a meta-scientific question, about the state of a science, and it belongs to the realm of social epistemology. In this paper, I addressed the second question, not the first.

Generally speaking, it is possible that a proposition *p* is true, and that there is a scientific consensus that *p*, but the consensus is not knowledge based. Suppose, for example, that *p* is true, but all the parties to the consensus are biased toward the truth of *p*, and would probably all believe that *p* even if *p* were false. The opposite situation is also possible.

It is *possible* though *implausible* according to my account of knowledge-based consensus that there will be a knowledge-based consensus over *p*, when *p* is in fact false.

The statement that a consensus over *p* is *not* knowledge based does not mean that *p* is false. Rather, it means that the mere fact that an agreement exists in an epistemic community does not carry epistemic weight. It does not give us any reason to believe that *p* *over and above* the reasons the members of the community have to believe that *p*. It may still be the case that these reasons alone are compelling and sufficient for justifiably believing that *p*, or it may not. When we can determine that a consensus is knowledge based, this means that we have an independent reason to justifiably believe that *p*.

Applying this general discussion to the Bendectin case, my claim that the Bendectin consensus was not knowledge based should be confused neither with the claim that Bendectin does not cause birth-defects, nor with the claim that the courts ruled incorrectly, nor with the claim that the evidence for the theory that Bendectin does not cause birth defects was insufficient for justifiably believing so. Rather, it means that the courts could not defer to the consensus and justify their decisions by appealing to the fact that a scientific consensus on this matter existed. It also means that the scholars who justify the court decisions by appealing to the emergence of the consensus on this matter are wrong in their epistemic analysis of this case study.

Another question that arises is whether animal research on Bendectin was terminated prematurely. Solomon (2001) would seem to answer this question in the positive. Solomon argues that as long as a theory enjoys some empirical success, some research efforts should be allocated to it. In the Bendectin case, so she would seem to argue, since some animal studies showed positive results, animal research was terminated prematurely.

I am not committed to this position, and my analysis does not necessarily imply that it was wrong to terminate further Bendectin research in order to conclusively determine whether it caused birth defects. There might be overriding social considerations that would justify terminating further research in this case. Recall that in the mid 1980's, when animal research was *de facto* terminated because courts no longer accepted animal studies as stand-alone evidence, Bendectin had already been taken out of the market. It no longer posed a potential threat to public health. It might thus be argued that further public resources in order to conclusively establish whether it caused birth defects or not just for the sake of the Bendectin trials should not have been allocated. Such an argument, though, cannot be supported by the existence of consensus, since the consensus in this case was not knowledge-based. It may be supported by non-epistemic considerations that concern the just distribution of research resources in society. Addressing such considerations exceeds the scope of this paper. It follows from my analysis, however, that if such further research had been done, it would not have been "junk science".

In my view, the Bendectin case study constitutes a warning sign against misusing scientific consensus in public debates, and against a demand by the public or by decision makers for a scientific unified front on a given matter as a condition for legitimizing policy or warranting action. Requiring a scientific consensus assumes a distorted image of science as a body that always speaks in one voice. A demand for consensus undermines scientific knowledge. It sets the wrong message that any dissent is illegitimate, and thus incentivizes those who object to a certain policy to artificially manufacture doubt. A demand for consensus may also lead to silencing legitimate dissenting voices within the scientific community, inhibiting progress, and may simply set too high a bar for science to meet (de Melo-Martín & Intemann 2013, p. 233).

Such a demand also ignores the normality of scientific pluralism and dissent, and their positive epistemic role in justifying our theories and discovering the truth, which has been often highlighted in the philosophy of science. As Dascal (1998, p. 147) argues, rather than being abnormal and detrimental to science, "controversies are indispensable for the formation, evolution and evaluation of theories, because it is through them that the essential role of criticism in engendering, improving, and controlling the 'wellformedness' and the 'empirical content' of scientific theories is performed". Effective critical deliberation requires *real* dissent by members of an epistemic community who are convicted in their beliefs, rather than some members' merely playing the Devil's advocates. But if this is the case, the deliberation may still produce a mixed outcome rather than consensus (Beatty & Moore 2010. p. 203).

It might be objected that although a scientific consensus may be wrong or biased, and despite the epistemic benefits of dissent, policy makers have no better choice than deferring to a scientific consensus, i.e., while imperfect, deference to a scientific consensus is the most epistemically reliable way to inform policy makers about matters of fact. This seems to be Anderson's (2011) and Oreskes' (2007) position, who, despite their awareness of various cases in the history of science, in which a consensus was false or biased, advocate deference to a consensus as the best means of guiding policy. This position has been also recently advocated by sociologist of science Harry Collins:

> If there's a consensus among experts, and you think you can trust these people, and they're working with integrity and trying to argue that opposition are wrong using the normal ways of arguing in science—rather than political suppression—then you should base policies on the consensus even if you can't be sure that consensus is the truth. [...] Sometimes experts will be wrong, but what else are you going to do? Will you just ask your mum, or flip a coin? (quoted in Koerth-Baker 2011).[9]

I reject this view. There are better ways to go about than deferring to a consensus, asking mum, or flipping a coin. For example, as de Melo-Martín & Intemann (2013, p. 233) suggest, rather than demanding a scientific consensus as a condition for public policy, the

[9] In a co-authored paper by Collins and Evans, they acknowledge that the use of science in court complicates things, and makes this advice problematic: "In courtrooms and the like, even the most routinized procedures with the longest historical entrenchment can be the subject of heated and detailed analysis" (Collins and Evans 2002, pp. 267-268).

public may be educated to understand the normality of scientific dissent and its positive epistemic role, taking the sting out of the attempts to artificially manufacture uncertainty. Or as Beatty & Moore (2010, p. 209) suggest, the scientific community can aim not at achieving a consensus on a theory, but rather on achieving a collective agreement by both the majority and minority that the deliberation process of the theory was adequate and in this respect the theory can be said to stand for the view of the group.

But even in the current state of affairs, in which such social reforms have not yet been implemented, decision makers, judges and jurors included, are able to do better than blindly or uncritically deferring to the scientific consensus. Decision makers can assess the consensus and examine to what extent it is knowledge based. Only if it is knowledge based, should they defer to it – or so, at least, my account of knowledge-based consensus suggests, and the Bendectin case study reviewed in paper illustrates.

7. <u>Conclusion</u>

I reviewed the controversy that arose in the 1980's and early 1990's in a series of legal tort trials in the U.S. over the question of whether the drug Bendectin caused birth defects. I argued that while courts deferred to the consensus that emerged in the scientific community that the drug did not cause birth defects, the consensus was not knowledge based. As I demonstrated, the consensus in this case did not exhibit an apparent consilience of evidence or social diversity. Rather, due to certain contingent social settings, epidemiological evidence triumphed over all other types of evidence. This entails that it could not be legitimately referred to, and generally suggests that courts and policy makers can use consensus as an epistemic resource less frequently than they do.

I distinguished between the question of whether the consensus in this case was knowledge based from the question of whether Bendectin was harmful as claimed. I argued that a negative answer to the first question does not necessarily imply a negative answer to the second question, but rather disqualifies the deference to the consensus as a legitimate way of resolving the second question in this particular case.

References

Alder, Ken. 2007. *The Lie Detectors: The History of an American Obsession*. New York: Free Press.

Anderson, Elizabeth. 2011. Democracy, Public Policy, and Lay Assessments of Scientific Testimony. *Episteme* 8(2): 144-164.

Aronson, Jay D. 2007. *Genetic Witness: Science, Law, and Controversy in the Making of DNA Profiling*. New Brunswick, NJ: Rutgers University Press.

Beatty, John, & Alfred Moore. 2010. Should We Aim for Consensus? *Episteme* 7(3): 198-214.

Beatty, John. 2006. Masking Disagreement among Experts. *Episteme* 3(1): 52-67

Brent, Robert. 1985. Editorial Comment on Comments on "Teratogen Update: Bendectin". *Teratology* 31(3): 429-430.

Brewer, Scott. 1998. Scientific Expert Testimony and Intellectual Due Process. *The Yale Law Journal* 107(6): 1535-1681.

Brown, Kenneth S., John M. Desesso, John Hassell, Norman W. Klein, Jon M. Rowland, A. J. Steffek, Betsy D. Carlton, Cas. Grabowski, William Slikker Jr., & David Walsh. 1985. Comments on "Teratogen Update: Bendectin". *Teratology* 31(3): 431.

Cole, Simon. 2001. *Suspect Identities: A History of Fingerprinting and Criminal Identification*. Cambridge, MA: Harvard University Press.

Collins, Harry M. & Robert Evans. 2002. The Third Wave of Science Studies: Studies of Expertise and Experience. *Social Studies of Science* 32(2): 235-96.

Cranor, Carl F. 2005. The Science Veil over Tort Law Policy: How Should Scientific Evidence be Utilized in Toxic Tort Law. *Law and Philosophy* 24(2): 139-210.

Dascal, Marcelo. 1998. The Study of Controversies and the Theory and History of Science. *Science in Context* 11(2): 147-154.

Daubert v. Merrell Dow Pharmaceuticals, Inc. 1993. 509 U.S. 579.

de Melo-Martín, Inmaculada & Kristen Intemann. 2013. Scientific Dissent and Public Policy: Is Targeting Dissent a Reasonable Way to Protect Sound Policy Decisions? *EMBO Reports* 14(3): 231-235.

Douglas, Heather. 2012. Weighing Complex Evidence in a Democratic Society. *The Kennedy Institute of Ethics Journal* 22(2): 139-162.

Douven Igor. 2009. Uniqueness Revisited. *American Philosophical Quarterly* 46(4): 347-360.

Edmond, Gary & David Mercer. 2000. Litigation Life: Law-Science Knowledge Construction in (Bendectin) Mass Toxic Tort Litigation. *Social Studies of Science* 30(2): 265-316.

Feyerabend, Paul. 1970. Consolations for the Specialist. In *Criticism and the Growth of Knowledge*, eds. Imre Lakatos and Alan Musgrave. Cambridge: Cambridge University Press.

Foster, Kenneth R., David E. Bernstein & Peter W. Huber. 1993. *Phantom Risk: Scientific Inference and the Law*. Cambridge, MA: MIT Press.

Frances, Bryan. 2005. *Scepticism Comes Alive*. New York: Oxford University Press.

Frye v. United States. 1923. 293 F.2d 1013 (D.C. Cir.)

Gelfert, Axel. 2011a. Expertise, Argumentation, and the End of Inquiry. *Argumentation* 25(3): 297-312.

Gelfert, Axel. 2011b. Mathematical Formalisms in Scientific Practice: From Denotation to Model-Based Representation. *Studies in History and Philosophy of Science* 42(2): 272-286.

Gilbert, Margaret. 1987. Modeling Collective Belief. *Synthese* 73(1): 185–204.

Gilbert, Margaret. 1994. Remarks on Collective Belief. In F. F. Schmitt (Ed.), *Socializing Epistemology: The Social Dimensions of Knowledge*, 235-256. Lanham, MD: Rowman & Littlefield.

Gilbert, Margaret. 2002. Belief and Acceptance as Features of Groups. *Protosociology* 16: 35-69.

Golan, Tal. 2004. *Laws of Men and Laws of Nature: The History of Scientific Expert Testimony in England and America*. Cambridge, MA: Harvard University Press.

Goldman, Alvin I. 2001. Experts: Which Ones Should You Trust? *Philosophy and Phenomenological Research* 63(1): 85-111.

Green, Michael D. 1996. *Bendectin and Birth Defects: The Challenge of Mass Toxic Substances Litigation*. Philadelphia: University of Pennsylvania Press.

Haack, Susan. 2003. Entangled in the Bramble-Bush: Science and the Law. In *Defending Science – within Reason: Between Scientism and Cynicism*, 233-264. Amherst, NY: Prometheus Books.

Haack, Susan. 2004. Truth and Justice, Inquiry and Advocacy, Science and Law. *Ratio Juris: An International Journal of Jurisprudence and Philosophy of Law* 17(1):15-26.

Haack, Susan. 2005. Trial and Error: The Supreme Court's Philosophy of Science. *American Journal of Public Health* 95(S1): S66-S73.

Hacking, Ian. 2002. *Historical Ontology*. Cambridge, MA: Harvard University Press.

Halfon, Saul. 2006. The Disunity of Consensus: International Policy Coordination as Socio-Technical Practice. *Social Studies of Science* 36(5): 783-807.

Holmes, Lewis B. 1985. Response to Comments on "Teratogen Update: Bendectin". *Teratology* 31(3): 432-432.

Huber, Peter W. 1991. *Galileo's Revenge: Junk Science in the Courtroom.* New York: Basic Books.

Jasanoff, Sheila. 1995. *Science at the Bar: Law, Science, and Technology in America*. Cambridge, MA: Harvard University Press.

Jones, Ward E. 2002. Dissident Versus Loyalist: Which Scientists Should We Trust? *The Journal of Value Inquiry* 36(4): 511-520.

Koerth-Baker, Maggie. 2011. The Scientist who Studies Scientists—An Interview with Harry Collins. *Boing Boing* (Apr 14). http://boingboing.net/2011/04/14/the-scientist-who-st.html

Kuhn, Thomas S. 1970. *The Structure of Scientific Revolutions*, 2nd ed. Chicago: The University of Chicago Press.

Longino, Helen. 2002. *The Fate of Knowledge*. Princeton: Princeton University Press.

Lynch, Michael, Simon A. Cole, Ruth McNally & Kathleen Jordan. 2009. *Truth Machine: The Contentious History of DNA Fingerprinting*. Chicago: University of Chicago Press.

Mill, John S. 1859/1993. On Liberty. In *Utilitarianism, On Liberty, Considerations on Representative Government*, ed. J. M. Dent, 69-187. London: Everyman.

Miller, Boaz. 2013. When is Consensus Knowledge Based? Distinguishing Shared Knowledge from Mere Agreement. *Synthese* 190(7): 1293-1316.

Oreskes, Naomi. 2007. The Scientific Consensus on Climate Change: How Do We Know We're Not Wrong? In *Climate Change: What It Means for You, your Children, and your Grandchildren*, eds. Joseph F. C. DiMento & Pamela Doughman, 65-99. Cambridge, MA: MIT Press.

Oxendine v. Merrell Dow Pharmaceuticals, Inc. 1996. WL 680992 (D.C. Super. 1996).

Parascandola, Mark. 1997. Chances, Individuals and Toxic Torts. *Journal of Applied Philosophy* 14(2): 147-157.

Pickering, Andrew. 1982. Interests and Analogies. In *Science in Context: Reading in the Sociology of Science*, eds. Barry Barnes and David Edges, 125-146. Cambridge, MA: MIT press.

Pound, Roscoe. 1965. *An Introduction to the Philosophy of Law,* rev. ed. New Haven: Yale University Press.

Pritchard, Duncan. 2005. *Epistemic Luck*. New York: Oxford University Press.

Rescher, Nicholas. 1993. *Pluralism: Against the Demand for Consensus*. New York: Oxford University Press.

Sanders, Joseph. 1998. *Benediction on Trial: A Study of Mass Tort Litigation*. Ann Harbor: University of Michigan Press.

Simon, Michael A. 1992. Causation, Liability and Toxic Risk Exposure. *Journal of Applied Philosophy* 9(1): 35-44.

Solomon, Miriam. 2001. *Social Empiricism.* Cambridge, MA: MIT Press.

Solomon, Miriam. 2007. The Social Epistemology of NIH Consensus Conferences. In *Establishing Medical Reality: Essays in the Metaphysics and Epistemology of Biomedical Science*, eds. Harold Kincaid & Jennifer McKitrick, 167-177. Dordrecht: Springer.

Stegenga, Jacob. 2009. Robustness, Discordance, and Relevance. *Philosophy of Science* 76(5): 650-661.

Stegenga, Jacob. 2011. Is Meta-Analysis the Platinum Standard of Evidence? *Studies in History and Philosophy of Biological and Biomedical Science* 42(4), 497-507.

Wolf, Joan B. 2007. Is Breast Really Best? Risk and Total Motherhood in the National Breastfeeding Awareness Campaign. *Journal of Health Politics, Policy and Law* 32(4): 649-654.