UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

-- x

IN RE METHYL TERTIARY BUTYL ETHER
PRODUCTS LIABILITY LITIGATION

This document pertains to:

City of New York v. Amerada Hess Corp., et al., Case No. 04-CIV-3417.

-- x

Master File No. 1:00-1898
MDL 1358 (SAS), M21-88

DECLARATION OF HARRY T. LAWLESS, PH.D., IN OPPOSITION TO DEFENDANTS' JOINT MOTION IN LIMINE TO EXCLUDE THE OPINION OF PLAINTIFF'S EXPERT HARRY T. LAWLESS, WITH EXHIBITS

I, HARRY T. LAWLESS, declare:

1. I have been retained on behalf of plaintiff in the matter of City of New York v. Amerada Hess et al. (MDL 1358, Southern District of New York). I make this Declaration from my personal knowledge, and if called to do so, could and would testify competently to the matters set forth below.

2. I am currently, and have been since 1989, a Professor in the Food Science Department of Cornell University, where I specialize in teaching sensory evaluation methods and performing research on taste and smell. My full qualifications are given in my February 5, 2009, Expert Report in this case and my curriculum vitae, which is appended to my Expert Report, which, in turn is attached to Defendants' motion as Defendants' "Exhibit B." My March 20, 2009, Expert Rebuttal Report is attached to Defendants' motion as Defendants' "Exhibit E."

3. I have served as a testifying expert in litigation only five times prior to my work in this case. I have not testified or consulted in any litigation involving MTBE prior to my engagement in this case.

4. In my opinion, the levels at which detection of MTBE occurs are, or

should be, directly relevant to a drinking water supplier interested in maintaining the quality of the product it supplies. My reasoning is given in detail below.

5. My methodology is neither flawed nor unprecedented. There is ample precedent in the peer-reviewed literature. Three analyses of the Stocking et al. data, including one by the authors of the study, one by EPA and my own analysis all come to remarkably close estimates of the percentage of people detecting MTBE at various concentrations. My reasoning and support are given below.

My Opinions and methodology

6. As I state in my Expert Report, my primary opinion in this case is:

> Based on the best study of odor detection of MTBE (methyl tertiary butyl ether), it is my opinion that MTBE can be detected by a significant portion of the population by smell in drinking water at levels of 1-2 parts per billion (ppb). Based on the best study of odor detection of MTBE to date, this proportion is approximately 10%.

(Exp. Rept. ¶ 6.) I also conclude, based on the same study, that 25% of the population can detect MTBE in water at concentrations of 3 ppb and 50% of the population can detect MTBE in water at concentrations of 14 ppb. (Exp. Rept. ¶ 16.)

7. The study on which I relied is the so-called "Stocking Study," an MTBE odor study involving 57 untrained consumer panelists carried out by the National Food Laboratory in 1998, and later described in a published report, *Implications of an MTBE Odor Study in Setting Drinking Water Standards* by Andrew J. Stocking, Irwin H. Suffet, Michael J. McGuire and Michael C. Kavanaugh, 93 J. Am. Water Works Ass'n 95 (March 2001) (a copy of which is attached as Exhibit 1 to my Declaration.) After a detailed review of other MTBE odor studies in the literature, I concluded that the Stocking study is the best and most trustworthy based on several factors detailed in my Expert Report. (Exp. Rept. ¶ 18.)

8. For my opinions, I accepted both the test protocol and the resulting dataset from the Stocking Study as given. However, rather than calculating an overall "threshold" value that represents the concentration at which half the population can detect MTBE in water, I wanted to find out what the Stocking data could tell us about the concentrations at which other significant, but smaller proportions of the population can detect MTBE.

9. To answer that question, I performed an interpolation on a dose-response curve for percent correct as a function of concentration. This statistical methodology is also known as the "chance-corrected percentile method." My application of the chance corrected percentile method is described in notes 4-6 of my Expert Report, but is further explained below:

A. Using the the matrix of correct and incorrect responses of the 57 panelists (*see* Stocking (2001) (Ex. 1) at 102, Fig. 1), I added the number of correct choices (x's) down each column to get a sum for each concentration level, and then divided by the number of panelists and multiplied by 100 to convert to a percent correct at each concentration.

B. I then fit a line to the data using two alternative statistical tools:

   1) A linear best fit (least squares) line, with the form $y = bx + a$, where y is the percent correct, x is the log of the concentration, b is the slope and a is the intercept; and

   2) A logistic regression equation, $\log[P/(1-P)] = a+b*\log(C)$, where P is the percent correct, C is the concentration of MTBE, a is the intercept and b is the slope.

C. Once the equation under each of the above alternatives was derived, using the standard Abbot's formula, where the corrected percentile = (observed percentile – chance probability)/(1 – chance probability), I corrected for guessing at the 66.6% correct, 50% correct and 40% correct concentration levels to get the true detection levels at 50%, 25% and 10%, respectively.

D. Finally, I converted the log concentration values to ppb.

10. I found that the linear best fit (least squares) and logistic regression approaches yielded very similar values, but ultimately relied on the linear best fit values for his

opinions, based on an R-squared analysis which showed that it produced a slightly better fit to the data (.93 v. .90). (Exp. Rept. n.6.) I also calculated 95% confidence intervals around my 10% detection value (0.5 to 4 ppb), 25% detection value (1.1 to 9.6 ppb) and 50% detection value (5 to 43 ppb). (Exp. Rept. ¶ 17, n.7) These confidence intervals indicate a relatively narrow margin of error, especially at the lower percentiles.

Issue: Allegation that rejection and not detection is the only critical issue in this case

11. In my deposition testimony, I referred to the levels at which consumers would reject drinking water containing MTBE as "important in this litigation." I am not a lawyer and do not see myself as qualified to offer an opinion regarding what evidence is and is not relevant in a lawsuit. No data exists regarding the concentrations at which consumers might "reject" or "object to" MTBE. While I believe that it would be helpful and informative to all parties in this case to have data on this issue to examine, the absence of data on MTBE rejection threshold levels has no bearing on my opinions in this case, which related to detection.

12. The levels at which detection of MTBE occurs are directly relevant to a drinking water supplier interested in maintaining the quality of the product it supplies. The level at which consumers actually reject their drinking water is a level to be avoided by a water provider. In other words, it informs us of a danger zone, a level to be avoided, or a ceiling. This does not mean that water providers would not want to consider a lower level as a standard for purposes of insuring consumer satisfaction and avoiding complaints.

13. Existing studies of MTBE's taste and odor almost exclusively consider detection. Because no data are available concerning rejection or objection thresholds, the detection data are that much more important. They are the only measures to inform the jury.

14. Defendants assert that the objection level is necessarily higher than the

detection level. The consumer objection level for a substance may be higher than the detection level, but the consumer objection level has not been determined for MTBE. In the absence of any data, the assertion that the objection level is higher than the detection level is speculation; it is equally valid to presume that consumers will object to any foreign taste or odor in their water.

Issue: Assessment of levels other than 50% detection; there are ample precedents.

15. The assessment of detection levels other than 50% is has ample precedent.

16. In my Expert Rebuttal, I cited several sources showing that levels other than 50% detection are commonly used in food and beverage testing for practical purposes, especially when the goal is to gain evidence that a change in flavor is *not* detected, as is the case here with drinking water standards. I cited one further example of a threshold method which used percentages other than 50%. (Rebuttal at 2-3, n.9-11.)

17. Regarding the choice of other percentages than 50%, this has precedent from reports using the Stocking data itself. In an extensive statistical analysis of the Stocking data, the USEPA in 2001 examined various methods to estimate of levels other than 50% detection for MTBE (from the Stocking data), including 5%, 10%, and 25%. (*See* Declaration of Andrew E. Schulman, Ex. 2.) This extensive report was peer-reviewed by a distinguished body of advising scientists. This report performed an extensive and complex statistical analysis of the dose-response curve for MTBE from the Stocking data. Their conclusion, using several complex statistical models, was that the 10% detection level of MTBE was at 2.2 ppb, which corresponds within a reasonable error to my estimate of 1 – 2 ppb, and most importantly is lower than the group threshold estimate (50% detection) of 15 ppb.

18. Even the authors of the Stocking study performed a similar analysis. In a draft report of the Stocking study, those authors had themselves calculated levels of detection

other than 50%, including 5%, 10% and 25%. (*See* Schulman Decl. Ex. 3.) The value from the draft Stocking report for 10% detection was also 2.2 ppb. So the very authors of the Stocking report have also taken this approach. Defendants' own expert is one author.

19. The California Department of Health Services (DHS), in a report adopting a secondary maximum contaminant level for MTBE, also affirmed that examining proportions below half the population would be advisable in setting standards (*See* Declaration of Todd E. Robins, Ex. 2 at 9-10.)

20. Furthermore, in Dale, M.S., Moylan, M.S., Koch, Bart, and Davis, M.K., 1997, Taste and odor threshold determinations using the flavor profile method: Proceedings of the American Water Works Association Water-QualityTechnology Conference, Denver, Colorado.Dale et al. (1997), researchers with the Metropolitan Water District of Southern California – which Defendants' expert refers to as one of "the largest water utilities in the U.S." – also alluded to the importance of looking at more sensitive groups in water contamination standards, and not just the population average when they stated, "to ensure maximum consumer satisfaction, water utilities would benefit from targeting concentrations that would be perceived by the most sensitive portion of the general population (e.g. less than 10 percent)."

Issue: Allegations That My analysis was flawed and unprecedented

21. Defendants assert that my method was "unprecedented" and that it has no support. This is incorrect.

22. The use of the correction for guessing is well-established. Abbott's formula, the standard formula for correction for guessing, has been with us since at least 1924. In my Expert Report I cite multiple literature references that demonstrate use of Abbot's formula as an established method for adjusting for baseline response. (Exp. Rept. n. 4.) Hundreds more

could be found. There is nothing mysterious about a correction for guessing; it is widely used in educational testing, sensory evaluation and other fields.

23. More specifically, within the realm of choice tasks in sensory research, the use of correction for guessing, specifically in forced-choice tasks (commonly known as "multiple choice" tests in everyday experience), such as the method of ASTM E- 679, was proposed as early as 1978 by Morrison, D. G. 1978, A probability model for forced binary choices. American Statistician, 32, 23 – 25, and is a common treatment for such data. The use of chance corrected values for determining thresholds was employed in the peer-reviewed paper for diacetyl thresholds by Antinone et al. 1994, *The importance of diacetyl as a flavor component in full fat cottage cheese*, 59 Journal of Food Science 38 at 39, a study I co-authored and cited in my original Expert Report. (Exp. Rept n. 5.) As my co-authors and I wrote in Antinone et al. (1994), "[i]n order to correct for chance performance, Abbot's formula has been widely used in psychology for decades." A copy of Antinone et al. (1994) is attached as Exhibit 2 to my Declaration. Abbot's formula is also cited in my textbook chapter on detection threshold methods. A an excerpted copy of Chapter 6 from my textbook, Lawless H.T and Heymann H, 1998, *Sensory Evaluation of Food: Principles and Practices*, is attached as Exhibit 3 to my Declaration.

24. Even more pertinent to the issue of pollutants, a group of chemical engineers working in the air pollution area examined this correction-for-guessing approach, specifically in reference to the methods of ASTM committee E-18, the authors of method E-679, in Viswanathan, S., Mathur, G. P., Gnyp, A. W. and St. Peirre, C. C. 1983, Application of probability models to threshold determination, Atmospheric Environment, 17, 139 – 143. This study found similar thresholds from the correction-for-guessing analysis and the geometric mean.

25. My analysis of the E-679 data using chance-corrected percentiles has stood the test of time and presentation to students and colleagues. As stated in my Rebuttal, I have used this analysis of chance corrected percentiles for over 10 years in my classroom and laboratory exercises, specifically dealing with the method of ASTM E-679, in an upper level course for Cornell seniors and graduate students. (Rebuttal at 3.) We routinely compare the chance-corrected percentile analysis to the geometric mean approach. To the best of my recollection, no Cornell student has ever objected to the chance corrected percentile analysis as nonsensical, flawed or inappropriate. It is that simple and obvious. I have incorporated this analysis into the revision of the threshold chapter of my textbook using the Stocking MTBE data as a case study. This chapter was read, edited and approved by my co-author, Dr. Hildegarde Heymann of the University of California, Davis. She raised no objections to my analysis of the MTBE data using chance-corrected percentiles. Defendants' allegation that I "never made any such suggestion [of alternate analyses] to his colleagues" (Motion, pg. 18) is, therefore, incorrect. An excerpted copy of my revised textbook chapter is attached as Exhibit 4 to my Declaration.

26. Neither the logistic regression nor linear best fit (least squares) equations I used are new or untried. They are both well-established, alternative statistical methods for doing just what I did with the Stocking data: fitting a linear mathematical model to measurements obtained from an experiment. There is nothing unusual about fitting a line to a set of points. Thousands, if not millions, of scientists and students do it every day.

27. Specifically, logistic regression is simply a well-recognized method for fitting an S-shaped curve (the cumulative form of the familiar bell curve) to a set of data. Logistic regression is a general procedure commonly used in threshold methods. *See, e.g.*, Walker, J. C., Hall, S. B., Walker, D. B., Kendall-Reed, M. S., Hood, A. F. and Nio, X.-F.

2003. Human odor detectability: New methodology used to determine threshold and variation. Chemical Senses, 28, 817 – 826.

28. Linear best fit (least squares) is also commonly used in this context. In Campden & Chorleywood Food Research Association Group, 2003, *Odour Threshold of Methyl Tertiary Butyl Ether (MTBE) in Water*, Report S/REP/74638 (copy of which is attached as Exhibit 5 to my Declaration), the authors used the same approach to calculate the geometric mean from their data set. There is no scientific reason a best-fit line could not be used to determine other corresponding percentages and concentrations along the dose-response curve. Defendants' expert, Dr. Suffet, reprints the linear best fit graph from Campden (2003) in his own review of the study, I.H. Suffet, 2007, *A re-evaluation of the taste and odor methyl tertiary butyl ether (MTBE) in drinking water*, Water Science & Technology, 55, No. 5, 265-273, a copy of which is attached as Exhibit 6 to my Declaration.

Issue: Chance-corrected percentile analysis is not specifically prescribed by ASTM e-679

29. The assertion that because ASTM E-679 does not specifically prescribe the chance-corrected analysis of percentiles, that therefore they are somehow inappropriate or flawed, is unjustified. That ASTM prescribes the use of the geometric mean for the purpose of estimating a group threshold does not address what other analyses might be equally legitimate. ASTM standards do not rule out alternative analyses. They do not state, by default, what must not be done. They describe one simple reproducible method for making a measurement that can be duplicated in other laboratories. The goals of a standard method are simplicity and reproducibility. They are concise descriptions of a procedure and do not go into any discussion of alternative analyses in the manner that a scientific paper would do so. A copy of ASTM E679-91, the version of the standard practice used in the Stocking Study, is attached as Exhibit 7.

A copy of ASTM E1432-04, a related ASTM standard practice that discusses the limitations of E679 and endorses use of Abbot's formula to correct for chance, is attached as Exhibit 8.

30. I have not modified or proposed to modify the data collection method. My analysis is merely an alternative and legitimate way of viewing the data set and drawing scientifically sound conclusions.

31. One reason why ASTM might not have considered the chance-corrected analysis is that they were dealing with very small panels, and the level of uncertainty around a percent correct is inversely proportional to the square root of the panel size. The Stocking study itself is a departure from method E-679, as it involved 57 panelists and 456 forced-choice presentations. Thus, the Stocking Study presents far less uncertainty in terms of deriving percent-correct values along a dose-response curve than would the typical ASTM E679 data set.

Issue: Allegation that I "manufactured" an opinion to "amplify" the City's damages claims

32. Defendants' assert in the motion that the values I derived were "manufactured" lower threshold estimates, and that I was motivated to "amplify" the City's damages claims in this case. These assertions about my motivation and causal chain are speculative and unfounded. I was not motivated in any way to find a number because it would support amplified damage claims. I was retained to examine the literature and come to reasoned conclusions, and I did so using sound scientific practices.

I, Harry T. Lawless, hereby declare under penalty of perjury under the laws of the United States, that the foregoing is true and correct.

Executed this 14th day of May, 2009 at Santa Fe, New Mexico.

HARRY T. LAWLESS, PH.D.