

UNITED STATES DISTRICT COURT  
SOUTHERN DISTRICT OF FLORIDA

CASE NO. 11-20427-WILLIAMS/TURNOFF

DISNEY ENTERPRISES, INC.,  
TWENTIETH CENTURY FOX FILM CORPORATION,  
UNIVERSAL CITY STUDIOS PRODUCTIONS LLLP,  
COLUMBIA PICTURES INDUSTRIES, INC., and  
WARNER BROS. ENTERTAINMENT INC.,

*Plaintiffs,*

v.

HOTFILE CORP., ANTON TITOV, and  
DOES 1-10.

*Defendants.*

\_\_\_\_\_ /

HOTFILE CORP.,

*Counterclaimant,*

v.

WARNER BROS. ENTERTAINMENT INC.,

*Counterdefendant.*

\_\_\_\_\_ /

**REPLY DECLARATION OF DR. RICHARD WATERMAN IN SUPPORT OF  
PLAINTIFFS' MOTION FOR SUMMARY JUDGMENT AGAINST DEFENDANTS  
HOTFILE CORP. AND ANTON TITOV**

PUBLIC VERSION

I, Richard Waterman, hereby declare as follows:

1. I am an Adjunct Professor of Statistics at The Wharton School at the University of Pennsylvania, and the President and Co-Founder of Analytic Business Services, Inc., a consultancy focused on providing expert advice and opinions in the field of statistical analysis. I have been retained by the plaintiffs Disney Enterprises, Inc., Twentieth Century Fox Film Corporation, Universal City Studios Productions LLLP, Columbia Pictures Industries, Inc., and Warner Bros. Entertainment Inc. (“plaintiffs”) to provide analysis related to the above-captioned case. The statements made in this declaration are based on my personal knowledge or application of my specialized knowledge to facts or data of which I am aware. If called to testify, I would testify based on the best of my knowledge, information, and belief, as follows:

2. I received my Ph.D. in Statistics from the Pennsylvania State University in 1993. I have substantial experience designing and reviewing sampling protocols for various large organizations, such as the United States Postal Service, for whom I designed and analyzed a national multi-stage sample for the estimation of operational characteristics. Further explanation of my experience, qualifications, publications, previous testimony, and compensation is included in the declaration I previously submitted to the Court in support of plaintiffs’ motion for summary judgment, filed February 17, 2012.

3. I have reviewed the Declaration of Professor James Boyle in Support of Defendants’ Opposition to Plaintiffs’ Motion for Summary Judgment and its attached Exhibit 2, which is a copy of his Rebuttal Report dated January 6, 2012. In paragraph 53 of his rebuttal report, he calculates certain statistics based on the results of the statistical analysis that I designed and about which I testified in my previous declaration. According to his report, he claims to have added up the total number of “paidfor” counts and total number of downloads (from

“dailydownload”) for each file within certain of Mr. Zebrak’s categories – “Confirmed Infringing,” “Highly Likely Infringing,” “Noninfringing,” and “Unknowable.” He then calculated a purported “conversion rate” for each category, consisting of the sum of the “paidfor” counts (which reflect users “converting” to Premium status) over the sum of downloads.

4. The resulting statistics – the “conversion rates” for each category – do not statistically support the conclusions that Prof. Boyle makes about non-infringing files on Hotfile being more likely to convert users to Premium than infringing files. First, Prof. Boyle calculated these quantities without consideration of the underlying sampling design and they are therefore incorrect. Properly calculated, the data does not show any statistically significant difference between the conversion rate for infringing and non-infringing files. Second, Prof. Boyle did not calculate the margin of error on the conversion rate for the non-infringing files. However, because the sub-sample of non-infringing files in the study is so small, and the characteristics of the files in that sub-sample are skewed, the margin of error on the conversion rate for non-infringing files is relatively high.

5. First, Prof. Boyle attempts to reach conclusions about the characteristics of certain *files* on Hotfile rather than the characteristics of certain *downloads*. In this case, I developed a protocol in order to select a sample to draw conclusions about overall daily downloads on Hotfile (whether the downloads were infringing). However, Prof. Boyle re-purposes that data and draws conclusions about categories of different files (the “conversion rate”). The protocol that was implemented in this case was not designed, without a suitable weighting adjustment, for that purpose. In order to draw general conclusion about files on Hotfile from the sample he selected, he must perform additional calculations and analyses. It is

not statistically valid to extrapolate the “conversion rates” he calculated to apply to files on Hotfile in general.

6. If Prof. Boyle had attempted to weight the sample to adjust for the probability that a file was selected into the sample, and had properly taken into account the statistical uncertainty in the calculations, he would have found that there was no statistically significant difference between the conversion rates on the infringing and non-infringing categories. In order to weight the sample correctly, he would have needed to take into account the number of downloads of a file from the date in which the file was selected, so that the calculations reflected the probability of the file being selected on that day. Performing that calculation produces conversion rates for the infringing and non-infringing categories that are different from those reported by Prof. Boyle but not statistically significantly different from each other.<sup>1</sup> In other words, the observed differences in conversion rates are accounted for by random variation in the data, and from a statistical point of view the rates are indistinguishable.

7. Second, independent of whether he used the right calculations, Prof. Boyle did not calculate the margin of error on the conversion rate for the non-infringing files. However, when Prof. Boyle seeks to examine the characteristics of only the non-infringing files, he is examining a sample size of 87 files. The small size of this sub-sample contributes to a relatively large margin of error on the conversion rate for non-infringing files. Further, even a cursory look at the sub-sample of non-infringing files shows that three specific files are extreme outliers in terms of their popularity in downloads and conversions – iREB, sn0wbreeze, and Jdownloader. As a result, the characteristics of the files in the sub-sample are highly skewed. This also contributes to the relatively large margin of error. Overall, the margin of error tells us the precision of the

---

<sup>1</sup> This is determined using a 0.05 level of significance, which is equivalent to 95% confidence.

estimate that is calculated. The true conversion rate could be (with 95% confidence)<sup>2</sup> within the range of the estimated rate, plus or minus the margin of error.

8. If Prof. Boyle had properly calculated the conversion rate as I explain above in paragraph 6, the margin of error would be *more than 57%* of the value of the “conversion rate” estimate itself.<sup>3</sup> This indicates that the calculated “conversion rate” lacks precision. Most sampling protocols in the real world would be designed to produce a substantially lower coefficient of variation (a statistic calculated from the margin of error) than the very large coefficient obtained on the conversion rate for non-infringing files.<sup>4</sup> Further, even looking at the flawed methodology that Prof. Boyle used to calculate the conversion rate for non-infringing files in his rebuttal report, the margin of error on his conversion rate is approximately 33% of the estimated value. Properly calculated, the margin of error shows that there is no statistically significant difference between the conversion rate on infringing and non-infringing files.

9. Finally, I also note that if one looks at the number of premium conversions per category of infringing versus non-infringing content in Prof. Boyle’s sample described in paragraph 53 of his rebuttal report, there are many more conversions on infringing rather than non-infringing files as a category. Moreover, as I noted above, three specific files are extreme outliers in terms of their popularity in downloads and conversions. Even if the calculated conversion rate had any statistical significance (and it does not), a reasonable conclusion would be that these files, not non-infringing files generally, drive a higher rate of conversions.

---

<sup>2</sup> The 95% refers to the repeated application of the sampling design, in the sense that if the entire sampling design were repeated many times, then one would expect that 95% of the confidence intervals to cover the true but unknown population level conversion rate.

<sup>3</sup> I used the widely accepted and reliable “bootstrap” technique to calculate the margin of error. See Bradley Efron, (1987), *The Jackknife, the Bootstrap, and Other Resampling Plans*.

<sup>4</sup> A typical goal for coefficient of variation would be of the order of 5%, whereas the coefficient of variation here was approximately 28.5%, which is substantially larger.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on March 19<sup>th</sup> 2012, at Philadelphia, PA.



---

Dr. Richard Waterman