

EXHIBIT A

PUBLIC VERSION

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF FLORIDA

CASE NO.: 11-CIV-20427-WILLIAMS/TURNOFF

DISNEY ENTERPRISES, INC.,
TWENTIETH CENTURY FOX FILM CORPORATION, a
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.,

_____ /
Counter-Defendant.

DECLARATION OF DR. DANIEL S. LEVY IN
OPPOSITION TO PLAINTIFFS' MOTION TO
STRIKE PORTION OF TESTIMONY OF
PROFESSOR JAMES BOYLE

I, Daniel S. Levy, declare as follows:

1. I am National Managing Director for Advanced Analytical Consulting Group. This declaration is based on personal knowledge and all statements contained in this declaration are true and correct to the best of my knowledge. If called as a witness, I could and would testify to the facts set forth in this declaration.

2. I have a Ph.D. in Economics from The University of Chicago. I have designed and implemented statistical sampling protocols for business analysis and litigations over the course of more than 25 years. I have provided testimony involving surveys, sampling, statistics, econometrics, economics and business, among other topics, before state and Federal courts. I have served as an expert for the US Department of Justice, the US Securities and Exchange Commission, the New York State Attorney General and served as an Expert Arbitrator for the Internal Revenue Service. I have testified in a range of matters over a number of years. My curriculum vitae is attached to my original report in this matter dated January 6, 2011 in support of Hotfile's opposition to Plaintiffs' motion for summary judgment.

3. It is my understanding that Plaintiffs have moved to exclude Prof. Boyle's analysis of conversion rates, presented in his Rebuttal Report, on the grounds that it is based on statistical and sampling expertise which Plaintiffs claim Professor Boyle does not have. Plaintiff motion is baseless because Professor Boyle's analysis stands as a matter of straight forward calculation of percentages and does not require the support of statistics or sampling science. So while Professor Boyle's analysis is based on precisely the same set of download records that Dr. Waterman used in his analysis, which left Dr. Waterman's analysis invalid and outside the bounds of statistical science, Professor Boyle's analysis, even though it rests on precisely the same set of records, is valid for the purposes it is put to in Professor Boyle's report, which is a calculation of the characteristics of the population of 1750 downloads selected by Dr. Waterman.

4. The reason that Professor Boyle's use of this set of downloads is valid, when Dr. Waterman's analysis is incurably errant, is that unlike Dr. Waterman, Professor Boyle's report does not attempt to extrapolate his findings about the 1750 downloads to a broader population. Dr. Waterman, in significant contrast, repeatedly stated in his report and at his deposition, that his purpose in selecting the 1750 download records was to extrapolate his findings to the broader population of downloads in January 2011. Although Dr. Waterman's sampling process was invalid and unscientific for the use of extrapolating to the full population of downloads even in

January 2011, let alone to the broader population of downloads from Hotfile during the course of its operations, it did produce a set of 1750 download records that Professor Boyle can validly comment about. Professor Boyle's report is simply reporting what one finds by looking within these 1750 downloads.

5. In contrast to Dr. Waterman's report, Professor Boyle's report does not make any statements about the broader population of downloads in his analysis. Since Professor Boyle's analysis is about the population of the 1750 records and the proportion of conversions within that population and does not attempt to extrapolate to some broader population, his statements stand on the basis of straightforward calculations of percentages within the population of 1750 records selected by Dr. Waterman. Professor Boyle is not appealing to the properties of sampling and statistics, and does not need to, in order to make statements are about the conversion percentages found within the 1750 download records. Dr. Waterman's, albeit invalid, attempt to extrapolate his findings to the broader population of downloads beyond the 1750 downloads *does* require the properties of sampling and statistical science because he is attempting to extrapolate from a sample to a population.

6. To clarify this point, take the example of four class rooms of 35 sixth graders each. Based on a selection of 20 taken from one sixth grade class, the average age of those 20 sixth graders can be calculated. There is no variance on that average. The average age of those 20 sixth graders is known for certainty. In fact, it is not even necessary that the 20 sixth graders be selected randomly to calculate the average age of those 20 sixth graders as long as there is no use of that average to estimate the average age of sixth graders in the broader population of sixth graders such as in the four classes. Furthermore, the fact that the 20 selected students came from just one class and that the selection process was clearly not random would not impact the calculation of the average age of those 20 sixth graders. As long as the statements about the average age of those sixth graders is meant to apply only to the 20 selected sixth graders, the selection or sampling process does not impact the calculation, and the average age of the 20 sixth graders can be calculated with complete certainty; there is no variance, and the calculation does not rest on the process of, or knowledge of, the science of statistics or sampling.

7. However, if one wants to extrapolate the findings based on the 20 selected sixth graders to make some statements about the entire sixth grade then the fact that the sample was not scientifically drawn becomes critical. In this example, since the average age of the 20

selected sixth graders was not a valid scientific probability sample of the broader population, it cannot be used as a valid scientific probability sample from which to draw an estimate of the overall sixth grade.¹ However, the average age of the 20 selected students, if sampled with a valid scientific probability sample, would estimate the average age of the overall sixth grade with some degree of variance. After all, even if the 20 students were selected based on a valid scientific probability sample, it would still only be a sample. The actual average age of the whole sixth grade would likely differ from the estimate by some amount, and the calculated variance of the estimated mean would be used by a scientist to reflect the degree of precision with which the average age of the entire sixth grade could be estimated based on the average age of the 20 selected students. But if statements about age are limited only to the selected records, and no statements are made about the broader population, then there is no variance to be calculated; it simply does not exist. In this example, the average age of the selected students is a specific number that is calculated without variance. The variance only comes into play when the selected records are to be used as an estimated measure of the age in the general population of sixth graders. But there is no variance when simply reporting the average age of the 20 sixth graders, and there is no reliance or required knowledge of the science of sampling or statistics to make such a calculation about the average age of the 20 students selected.

8. This is analogous to what Professor Boyle did in reporting the conversion rate in the 1750 records for each of the sub-populations he discussed. He presented no variance, because there is no statistical variance in the calculated average of the conversion rates when that average is simply a description of the characteristics of the population of 1750 download record. As long as Professor Boyle's statements and conclusions in his report are about the 1750 download records and do not require some extrapolation to some broader population, he does not require a background in the science of statistics or sampling. Further, there is no variance that is required to be presented for that purpose.

9. It is important to point out that the analysis of Professor Boyle, since it is based on precisely the same 1750 downloads chosen by Dr. Waterman, would have the same sampling properties as found in Dr. Waterman's analysis when addressing questions in which the relevant unit of observation is a download. Likewise, Professor Boyle's analysis would have the same sampling properties as Dr. Waterman's analysis if Professor Boyle attempted to extrapolate his

¹ William G. Cochran, *Sampling Techniques*, third edition, John Wiley & Sons, New York, 1977, P. 5.

results to the broader population of downloads. However, Professor Boyle did not attempt to make such an extrapolation and instead chose to present his findings in his report as they apply to the population of 1750 downloads. Not some broader population. As mentioned above, in performing his calculations and making his statements only about the 1750 downloads, Professor Boyle's analysis requires only the properties of basic math, not the properties of sampling or statistical science.

10. As a further point, Dr. Waterman claims that if Professor Boyle had altered his analysis to conform to what Dr. Waterman believes would be the right analysis, Professor Boyle would have reweighted the selection of 1750 downloads, which Dr. Waterman chose, and then extrapolated those results to a broader population, which would have required Professor Boyle to provide variances of these estimated conversion rates. Further, Dr. Waterman says that if Professor Boyle had done what Dr. Waterman had wanted, Professor Boyle would have found that the conversion rates between the "Highly Likely" (Infringing in Dr. Waterman's Exhibits) category and the "Noninfringing" category would not be statistically significant. Dr. Waterman's proposed analysis answers a different question than the one Professor Boyle apparently intended to address, so there is no reason why Professor Boyle should have performed the analysis that Dr. Waterman proposes. Professor Boyle's analysis details the average historical conversion rate based on a selection of downloads which were subsequently grouped into four categories. It is a valid calculation.

11. Dr. Waterman's analysis in his Reply Declaration attempts to provide estimates of the conversion rates in some broader population based on his sample of 1750 selected downloads. His analysis is again invalid because his sample is not a valid scientific sample of the broader population even for January 2011. However, putting that aside, it is simply a different question than Professor Boyle answered. In his Reply Declaration, Dr. Waterman states that "[p]erforming that calculation produces conversion rates for the infringing and non-infringing categories that are different from those reported by Prof. Boyle but not statistically different from each other." Dr. Waterman does not list these new conversion rates in his report, but fortunately they are listed in the Exhibits to his Reply Declaration. I have copied those results into Table 1 below.

Table 1

	Estimate	Std.err	WEIGHTED			
			MoE	95%CI LL	95%CI UL	cv
Illegal	0.00012					
Infringing	0.00060	0.00007	0.00014	0.00046	0.00074	11.93%
Noninfringing	0.00102	0.00029	0.00057	0.00045	0.00160	28.51%
Unknowable	0.00028	0.00024	0.00047	-0.00020	0.00075	87.17%

12 I have added the labels of the categories, at the left, which were simply missing from Dr. Waterman’s results on Page 44 of his Exhibits. But otherwise this is a replica of Dr. Waterman’s findings presented on Page 44 of his Exhibits to his Reply Declaration. Dr. Waterman produced these results based on his selection of records and statistical calculations. The “Estimate,” which is Dr. Waterman’s estimate of the average conversion rate is listed for four categories. Of interest are the estimated average conversion rates for Infringing and Noninfringing. Dr. Waterman estimated a conversion rate for the Noninfringing in his broader population of .00102, with a Std. err. of .00029. Professor Boyle found the average conversion rate of .001074 (written .1074% in Professor Boyles Report) within the population of 1750 downloads. For the Infringing category, Dr. Waterman finds an estimated conversion rate for his broader population of .00060 with a Std. err or .00007. Professor Boyle found the average conversion rate of .000586 (written .0586% in Professor Boyles Report, combined .000551 for the combined Highly Likely and Confirmed) within the population of 1750 downloads. Simply to facilitate comparison of the averages calculated by Professor Boyle and Dr. Watermen, I isolate them in Table 2 below.

Table 2

	Boyle Calculation of Average Conversion Rate within Population of 1750 Downloads	Waterman Estimates of Averaged Conversion Rate for Broader Population
Infringing	0.000551	0.00060
NonInfringing	0.001074	0.00102

Sources: Rebuttal Expert Report Professor James Boyle, Jan 6, 2012, P. 23, Reply Declaration Exhibits Dr. Richard Waterman. P. 44. Boyle presented

13 Reading across Table 2, where Professor Boyle found .000551, Dr. Waterman found .00060. Where Professor Boyle found .001074, Dr. Waterman found .00102. In think it is fair to say that Dr. Waterman's finding for the estimated average conversion rates validate, not contradict, Professor Boyles calculations of the average conversion rates he found within the population of 1750 downloads. Table 2 shows that Dr. Waterman's point estimates of the conversion rates in his broader population exhibit higher point estimates for the Noninfringing category than the Infringing category as did Professor Boyle's calculations for population of 1750 downloads. Reading the text of Dr. Waterman's Rebuttal Declaration one might expect that his reweighting and reanalysis had produced results that were very different from Professor Boyle's. However, side-by-side comparison of the results in Table 2 show that Dr. Waterman's results about the point estimates of conversion rates confirm Professor Boyle's finding, even under alternative weightings of the underlying records, which Dr. Waterman advocates and presumably finds acceptable.

14 Dr. Waterman also says that in his own reweighted result, the difference in his point estimates between Noninfringing and Infringing categories are not statistically significantly different. Below I shed further light on Dr. Waterman's calculation and claims about statistical significance which helps to clarify precisely what Dr. Waterman has found.

15 Dr. Waterman does not list within his Rebuttal Declaration the standard errors or the specific test of significance he used in his Rebuttal Declaration. However, he does list the "Std.err" of his point estimates in the related Exhibits. Based on these Std.errs, the point estimates and the numbers of observations in Dr. Waterman's selection of downloads as he reweights them on P.44 of his Exhibit, a standard test of statistical significance can be performed. Because Dr. Waterman has not provided the calculations he performed to support his opinion about his test of statistical significance, it is not possible to determine the exact calculation that Dr. Waterman used as the basis of his opinion that the conversion rates of the Infringing and Noninfringing categories were not statistically significant. It is certainly consistent with Dr. Waterman's statements that Dr. Waterman used what is known as a t-test for the difference between two means to perform his test of statistical significance. We do know from footnote 1 of Dr. Waterman's Rebuttal Declaration that his statement about the lack of

statistical significance was "...determined using a 0.05 level of significance, which is equivalent to 95% confidence." The standard interpretation of this confidence interval is that if the same sampling process were used repeatedly in a population the true average conversion rate would tend to be within confidence bounds 95% of the times.² The same statement would hold for a statement about a 99% confidence interval; in repeated samplings the true population average would tend to be within the alternative 99% confidence intervals 99% of the time. Similarly for a 90% confidence interval, the true population average would tend to be within the boundaries of the 90% confidence interval in 90% of the samples. I have performed calculations of the 95% and 90% confidence intervals based on a t-test of the difference between two means. I found that the difference in the conversion rates between Infringing and Noninfringing would be statistically significant at the 90% confidence level. So while Dr. Waterman says that "from a statistical point of view the rates are indistinguishable,"³ a fuller exposition of what Dr. Waterman's own data shows is that, in common parlance, the conversion rate in the Noninfringing files is likely to be greater than the conversion rate in the Infringing files. That difference does may not reach the 95% confidence level standard set by Dr. Waterman, but it would more than reach a 90% confidence level standard.

16 Contrary to the tone and language of Dr. Waterman's Rebuttal Declaration, Dr. Waterman's actual data and results, when featured side-by-side with Professor Boyle's, are very useful in confirm and validating Professor Boyle's findings and statements, even under the alternative methods of calculation and reweighting proposed and implemented by Dr. Waterman. Specifically, like Professor Boyle's findings for the population of 1750 downloads, Dr. Waterman's finding under his chosen methods, setting aside the invalid sampling it is based on, demonstrate that the point estimates of the conversion rate for Noninfringing are greater than that for Infringing. And further, that while this difference may not be statistically significant at a 95% confidence level, it would be statistically significant if the standard were set at the 90% confidence level.

17 Therefore, even though Professor Boyle's report does not discuss any extrapolation to a broader population outside the 1750 downloads selected by Dr. Waterman, Dr. Waterman has performed such an extrapolation based on methods that he presumably finds

² William G. Cochran, *Sampling Techniques*, third edition, John Wiley & Sons, New York, 1977, P. 12.

³ Rebuttal Declaration of Dr. Richard Waterman, March 19, 2012, P. 4.

acceptable, and that analysis buttresses Professor Boyles' conclusion that conversion rates among Noninfringing files are at least as great as Infringing files, and statistically significantly greater at the 90% confidence level.

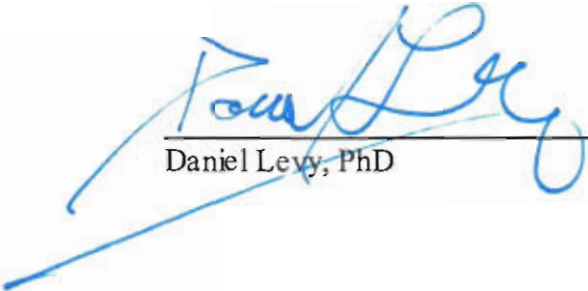
18 It is also important to point out that on P. 43 of Dr. Waterman's Exhibits to his Rebuttal Declaration, Dr. Waterman provides calculations of the average conversion rates and the Std.err of those estimated conversion rates, without the reweighting of Dr. Waterman's selection of the 1750 download; that is, based on the same weighting that Dr. Waterman used for his original report. Those unweighted average conversion rates and standard errors (Std.err) show that the conversion rates for the Noninfringing downloads are statistically significantly greater than the conversion rates for the Infringing files at the 95% confidence level chosen by Dr. Waterman. I provide these calculations in Appendix 1. Dr. Waterman does not discuss these results in his report, but the average conversion rates and Std.err needed to perform these calculations are available on P. 43 of the Exhibits to Dr. Waterman's Rebuttal Declaration. While Dr. Waterman asserts that such a calculation on the unweighted downloads, which Dr. Waterman himself selected, is not appropriate, the calculation of the conversion rate based on downloads appears to be what Professor Boyle was addressing in his original report. If the conversion rate based on each of these categories based on downloads is the point of interest for Professor Boyle, then Dr. Waterman has provided that statistical know-how and the calculations needed to extend Professor Boyle's calculation, which was limited to the 1750 download records, to the broader population of downloads, to the extent Dr. Waterman's unscientific sample can be employed for any extrapolation. And to the extent that any of Dr. Waterman's samples can be used, the results in Dr. Waterman's Exhibits to his Rebuttal Report show that Noninfringing files based on the sample of downloads have a statistically significantly greater conversion rate than Infringing files based on the sample of downloads.

19 In summary, Professor Boyle's analysis requires only the basis of standard mathematical calculation because he makes no statements about any broader population outside the population of 1750 download records chosen by Dr. Waterman. Professor Boyle's analysis therefore is unassailable on any statistical or sampling science grounds since these disciplines are not required to validate Professor Boyle's analysis. Dr. Waterman's sample itself is woefully invalid and unscientific for use in extrapolating to the broader population of January 2011 or any broader population. However, Dr. Waterman's Rebuttal Declaration analysis supports, rather

than contradicts or invalidates, the findings about conversion rates presented by Professor Boyle. Conversion rates of Noninfringing files based on a sample of files or downloads chosen by Dr. Waterman are at least as great as those of infringing files. Further, based on the type of t-test that Dr. Waterman appears to have used in his rebuttal report, the conversion rate of Noninfringing files selected in Dr. Waterman's selection of 1750 downloads is statistically significantly greater than the conversion rate for the Infringing files based on the 95% confidence test chosen by Dr. Waterman.

I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct.

Executed on this 2nd day of April 2012, at Boston, Massachusetts.



Daniel Levy, PhD

SAS Code:

```
libname hotfile "E:\Main\Clients\Hotfile\Team A\Report Analysis\Data";  
options compress=yes;
```

```
proc import datafile="E:\Main\Clients\Hotfile\Team A\Report Analysis\Data\boyle_conv_sum.csv"  
out=boyle  
dbms=csv  
replace;  
run;
```

```
proc import datafile="E:\Main\Clients\Hotfile\Team A\Report  
Analysis\Data\wm_conv_sum_weighted.csv"  
out=weighted  
dbms=csv  
replace;  
run;
```

```
proc print data=boyle;  
run;
```

```
proc print data=weighted;  
run;
```

```
title "Test for Equality of Means Conversion Rate Infringing-Conversion Rate NonInfringing:  
Boyle";
```

```
proc ttest data=boyle sides=L;  
class CATEGORY;  
var CONVERSION;  
run;
```

```
title "Test for Equality of Means Conversion Rate Infringing-Conversion Rate NonInfringing:  
Waterman-Weighted";
```

```
proc ttest data=weighted sides=L;  
class CATEGORY;  
var CONVERSION;  
run;
```

boyle_conv_sum.csv:

Category	_STAT_	Conversion	
Infringing	N	1579	
Infringing	MEAN	0.00053	
Infringing	STD	0.004768	0.00012
NonInfringing	N	87	
NonInfringing	MEAN	0.00116	
NonInfringing	STD	0.001772	0.00019

wm_conv_sum_weighted.csv

Category	_STAT_	Conversion	
Infringing	N	1579	
Infringing	MEAN	0.0006	
Infringing	STD	0.00278156	0.00007
NonInfringing	N	87	
NonInfringing	MEAN	0.00102	
NonInfringing	STD	0.00270494	0.00029

Note: Sample standard deviations were estimated using the following formula:
 $\text{sqrt}(\text{std}(\text{mean})^2 * n)$.

SAS Output

The SAS System

13:44 Monday, April 2,

2012 1

Obs	Category	_STAT_	Conversion	VAR4
1	Infringing	N	1579	.
2	Infringing	MEAN	0.00053	.
3	Infringing	STD	0.0047684	0.00012
4	NonInfringing	N	87	.
5	NonInfringing	MEAN	0.00116	.
6	NonInfringing	STD	0.0017722	0.00019

Obs	Category	_STAT_	Conversion	VAR4
1	Infringing	N	1579	.
2	Infringing	MEAN	0.0006	.
3	Infringing	STD	0.00278156	0.00007
4	NonInfringing	N	87	.
5	NonInfringing	MEAN	0.00102	.
6	NonInfringing	STD	0.00270494	0.00029

The TTEST Procedure

Variable: Conversion

Category	N	Mean	Std Dev	Std Err	Minimum	Maximum
Infringing	1579	0.000530	0.00477	0.000120	.	.
NonInfringing	87	0.00116	0.00177	0.000190	.	.
Diff (1-2)		-0.00063	0.00466	0.000513		

Category	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev
Infringing		0.000530	0.000295 0.000765	0.00477	.
NonInfringing		0.00116	0.000782 0.00154	0.00177	.
Diff (1-2)	Pooled	-0.00063	-Infty 0.000215	0.00466	0.00451
Diff (1-2)	Satterthwaite	-0.00063	-Infty -0.00026		

Method	Variances	DF	t Value	Pr < t
Pooled	Equal	1664	-1.23	0.1099
Satterthwaite	Unequal	166.85	-2.80	0.0028

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	1578	86	7.24	<.0001

Test for Equality of Means Conversion Rate Infringing-Conversion Rate NonInfringing: Waterman-Weigh 4

13:44 Monday, April 2,

2012

The TTEST Procedure

Variable: Conversion

Category	N	Mean	Std Dev	Std Err	Minimum	Maximum
Infringing	1579	0.000600	0.00278	0.000070	.	.
NonInfringing	87	0.00102	0.00270	0.000290	.	.
Diff (1-2)		-0.00042	0.00278	0.000306		

Category	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev
Infringing		0.000600	0.000463 0.000737	0.00278	.
NonInfringing		0.00102	0.000443 0.00160	0.00270	.
Diff (1-2)	Pooled	-0.00042	-Infty 0.000083	0.00278	0.0026
Diff (1-2)	Satterthwaite	-0.00042	-Infty 0.000075		

Method	Variances	DF	t Value	Pr < t
Pooled	Equal	1664	-1.37	0.0850
Satterthwaite	Unequal	96.295	-1.41	0.0812

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	1578	86	1.06	0.7591

UNWEIGHTED

	Estimate	Std.err	MoE	95%CI LL	95%CI UL	cv	var
Illegal	0.00014	0.00035	0.00068			247.73%	1.225E-07
Infringing	0.00053	0.00012	0.00023			21.72%	1.44E-08
Noninfringing	0.00116	0.00019	0.00038			16.64%	3.61E-08
Unknowable	0.00036	0.00020	0.00039			54.50%	0.00000004

Estimate_Noninfringing-Estimate_Infringing	0.00063
Var(Estimate_Noninfringing-Estimate_Infringing)	0.000000051
Std Err(Estimate_Noninfringing-Estimate_Infringing)	0.000225
Lower Bound of One-sided 90% Confidence Interval	0.0003420
Lower Bound of One-sided 95% Confidence Interval	0.0002604

Source: William G Cochran, *Sampling Techniques*, 3rd Edition, Wiley & Sons, New York, 1977, P. 180.

