Robert S. Arns, State Bar No. 65071
RSA@ARNSLAW.COM
Jonathan E. Davis, State Bar No. 191346
JED@ARNSLAW.COM
Steven R. Weinmann, State Bar No. 190956
SRW@ARNSLAW.COM
THE ARNS LAW FIRM
515 Folsom Street, 3rd Floor
San Francisco, CA 94105
Tel: (415) 495-7800
Fax: (415) 495-7888

Jonathan M. Jaffe, State Bar No. 267012
JMJ@JAFFE-LAW.COM
JONATHAN JAFFE LAW
3055 Hillegass Avenue
Berkeley, CA 94705
Tel: (510) 725-4293
Fax: (510) 868-3393

Attorneys for Plaintiffs

Redactions herein are confidential material not filed with the Court

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN JOSE DIVISION

ANGEL FRALEY; PAUL WANG; SUSAN MAINZER; JAMES H. DUVAL, a minor, by and through JAMES DUVAL, as Guardian ad Litem; and W. T., a minor, by and through RUSSELL TAIT, as Guardian ad Litem; individually and on behalf of all others similarly situated,

Plaintiffs,

v.

FACEBOOK, INC., a corporation; and DOES 1-100,

Defendants.

Case No. CV 11-01726 LHK PSG

DECLARATION OF RICHARD DROGIN IN SUPPORT OF PLAINTIFFS' MOTION FOR CLASS CERTIFICATION

Date: May 24, 2012
Time: 1:30 p.m.
Courtroom: 8
Judge: Hon. Lucy H. Koh
Trial Date: December 3, 2012

I, RICHARD DROGIN, declare and state as follows:

1. I hold a Ph.D. in statistics from the University of California at Berkeley, earned in 1970. I am currently an Emeritus Professor in the Department of Statistics at California State University, Hayward, where I have taught graduate and undergraduate courses in data analysis, non-parametric methods, regression analysis, sample surveys, probability theory, queuing theory, simulation methods and design of statistical software. I was employed at California State University, Hayward, since 1973, and became an Emeritus Professor in 1996.

2. I am a partner in the statistical consulting firm of Drogin, Kakigi & Associates. This firm provides consulting services and computerized database management. We have experience in designing and analyzing random sampling plans, organizing and managing large database systems, stochastic modeling, and performing advanced statistical analysis. Our firm has served as statistical consultants to both governmental agencies and the private sector for over twenty-five years. I have been retained as a statistical consultant in over 250 class action cases, primarily cases involving employment discrimination and wage and hour claims, and I have testified over 25 times as an expert witness (statistical analysis, computer processing) in numerous state and federal courts. I have never failed to qualify as an expert in statistics in any litigation. A copy of my current resume is attached hereto as Appendix 1.

3. Plaintiff's counsel retained me in the *Fraley v. Facebook* case pending United States District Court, Northern District of California, San Jose Division, to act as a statistical consultant. In particular, counsel has asked to me to give my opinion regarding the feasibility

of using data collected and maintained by Facebook to calculate, or estimate, various quantities related to click through rates for advertisements.

4. In connection with my work in this case, I reviewed the following documents.

a. Plaintiff's Complaint

b. The Order Denying Defendant's Motion to Dismiss

c. Plaintiff's Statement of Facts

d. Exhibit 326, spreadsheet including selected fields for named plaintiffs

e. Transcript for deposition of Chris Plambeck (PMK explaining Exh 326)

f. Spreadsheet showing Exh 326 data field explanations

g. Exhibit 693, User Level Tables (Bates 89753)

h. Exhibit 694, [redacted] description document (Bates 89754)

i. Exhibit 695, [redacted] description document (Bates 89755)

j. Exhibit 685, Ads Manager Examplar

k. Exhibit 684, Guide to the new Facebook Ads Manager

l. User Level Tables, Bates 89753

5. It is my understanding that Facebook maintains [redacted]. The database has information about [redacted]

[redacted]

6. The data described in the previous paragraph can be used to calculate the exact value of the click through rate (CTR) for a particular ad campaign, group of ad campaigns, or for all ads. The CTR is defined as the total number of times an ad is clicked upon divided by the total number of impressions. A separate CTR can be computed for Sponsored Stories ads and for Facebook Ads. I have read the declaration of Fernando Torres, and understand that Mr. Torres will testify that the difference between these two CTRs is closely related to the increase in revenue that is generated from a Sponsored Stories ad compared to a comparable Facebook Ad.

7. The formula provided by Mr. Torres gives a method for computing the additional revenue for Sponsored Stories ads. Mr. Torres' formula uses the CTR as one of its components. His formula for the additional revenue for Sponsored Stories ads can be used to calculate the additional revenue per impression, or for an entire ad campaign.

8. The random sampling techniques described below can be used to establish accurate estimates of the CTRs for Sponsored Stories, Facebook ads, and other subsets of ads. Using these estimates from a random sample in place of the exact values computed from all ads in Mr. Torres' formulas, will establish an accurate estimate of the appropriate additional revenue per ad campaign per Class Member across Facebook for Sponsored Stories ads; and that additional

revenue can then be allocated to the Class Members based on the number of impressions in which they have appeared.

9. [redacted] By taking an appropriate random sample from this data, it is possible to obtain accurate estimates of the CTRs for all ads, and for various subgroups of ads.

10. A random sample, or probability sample, is a subset of a population that is selected according to a random mechanism. There are various methods for choosing a random sample used by statisticians, and the particular method used for selection determines the appropriate formulas for making calculations and projections from the sample data that is collected to the population (i.e. class). The simplest and most common method for selecting a random sample is known as a simple random sample. Some examples of other random sampling methods include stratified random samples, cluster samples, systematic samples. The type of random sample used in a given situation depends on how the sample is selected, and not the results of the sample.

11. A simple random sample of size n is selected when every subset of n items in the population has the same probability of selection. For example, suppose it is desired to select a simple random sample of 20 items from a population of 1000 items, which are numbered 1, 2, 3, ..., 1000. One could write the numbers 1, 2, 3, ..., 1000 on 1000 identical slips of paper, mix them up, and then reach in blindfolded and pick out 20 slips. Equivalently, a computer random number generator could be used to sort the 1000 numbers in random order, and then just take the first 20 numbers from the randomized list as the sample.

12. A basic principle in statistics is that when a random sample is drawn from a population, then there is a measurable degree of reliability in making predictions from that sample to the population as a whole. The formulas for computing reliability and margin of error depend on the details of the how the random sample was drawn. Employing a random mechanism in selection of the sample insures that the mathematical laws of probability will apply in reaching conclusions about the population, based on the sample results. If the sample is not drawn according to some specifiable random mechanism, then there is no mathematics to justify projecting the results of the sample to the population. "In short, a good survey defines an appropriate population, uses an unbiased method for selecting the sample, has a high response rate, and gathers accurate information on the sample units. When these goals are met, the sample tends to be representative of the population: the measurements within the sample describe fairly the characteristics in the population."[1]

13. Statisticians specify the accuracy of estimates in terms of a confidence interval. A confidence interval consists of 3 components: a point estimate of some population quantity, a margin of error or bound associated with the point estimate, and a degree of confidence. The degree of confidence indicates how certain it is that the plus or minus margin of error attached to the point estimate will include the true (unknown) population value being estimated. For example, if the CTR for a set of ad campaigns is estimated, the the CTR measured from the random sample would be the point estimate. Appropriate statistical formulas would be applied to determine the margin of error, for a particular selected degree of confidence. For example, if the CTR from the sample data is .02, and a 99% level of confidence is used, then a statement

[1] "Reference Manual on Scientific Evidence", 2nd Edition, page 102.

such as the following might be made: "There is 99% confidence that the CTR is within the interval 0.0200 plus or minus 0.0001 (i.e. 0.0199 to 0.0201)."

14. The size of the margin of error depends on the sample size and the variation in the data. In situations where the random sample consists of selecting and examining paper documents or interviewing individuals, it is often difficult and costly to take large samples of, say, thousands of documents or people. However, in this case, given that the sample would be drawn from a computer datafile, and the compilation of data would be carried out by an appropriate computer program, it would be possible to take a very large sample (say, in the millions or tens or millions), and obtain relatively very accurate estimates.

15. There are various possible ways to draw a random sample in this case. The choice of type (or types) of samples that could be implemented depends (at least in part) upon the size of the datafiles, how they are organized, the level of aggregation of the raw data, and the subsets for which separate estimates are desired. Two types of random samples that could be drawn are the following: a) A random sample of ad impressions, and b) a random sample of ad campaigns. The random sample of ad impressions could be designed as a stratified sample, where appropriate subgroups would be defined and a simple random sample of ad impressions selected from each subgroup, such as type of ad, or size of company purchasing the ad. Such a sample would produce an accurate estimate of the CTR for each subgroup, and overall. The random sample of ad campaigns would be a form of cluster sampling, where the clusters are the ad campaigns. Then, either a complete enumeration of each cluster would be used, or a simple random sample from each cluster could be drawn. This type of sample would produce accurate estimates of the CTR for particular ad campaigns, for groups of ad campaigns, or overall. After

more information is obtained through discovery, the details of the sample design(s) can be specified.

16. Based on the information above, and the type and quantity of data retained by Facebook with respect to their advertising campaigns, there are various appropriate methods for calculating click through rates, which can be used in the formulas provided by Mr. Torres.

I declare under the penalty of perjury and under the laws of the United States of America that the forgoing is true and correct and based upon my personal knowledge and/or professional opinions, and that if called upon to testify, I could verify the accuracy of the same. This document was executed in the city of Berkeley, California on March 28, 2012.

Richard Drogin

By:_____

Dr. Richard Drogin