

## **Appendix: Search and Estimation Procedures**

Generally, the tables in the paperback edition are updated versions of the tables published in the original hardcover edition of 2001. The following changes should be noted:

- 1) The counts cover a longer time span: The media mention counts cover from January 1995 through July 2002 and the scholarly citations counts cover from January 1995 to approximately October 2002.
- 2) The new searches have eliminated the chronological bias towards the end of the data set by counting hits using a random sample that spanned all of the returns as opposed to the first data collection, which counted a sample using a sample of one hundred from two specific time periods (media mentions) or thirty (scholarly citation) hits. For many people, such as Zbigniew Brzezinski, the random sample and the samples used for the 2000 tables probably would have returned similar results since the name is fairly unique. However, for people with more common names, such as John Leonard, different John Leonards could have had their “15 minutes of fame” in an unexamined span of the results. For example, John Leonard, the critic for CBS News Sunday Morning, had a much higher return of media mention hits under the random sample. Examining the results, there were approximately over 350 hits of a John Leonard who made much news in August 1996 and August 1999 (when he initially filed suit and, later in 1999 when the judge threw out the case) after he sued Pepsi for false advertising. Clearly, missing these two bubbles of bad data would bias the final counts. Undoubtedly, this change in estimation effected some intellectuals more than others; and, reviewing the numbers will show that media mentions and scholarly citations for some have actually decreased, despite the longer time period, sometimes dramatically.
- 3) The third improvement in the data was the removal of the web hit counts. Initially devising a program that retrieved random samples of all hits and observing the results, we found a myriad of problems with the information on the Web. Even since 2000, when the first data collection was taken, the amount of information on the Web has increased exponentially and the search retrieved results ranging everything between library receiving lists to e-mails posted on the Web to street names or addresses. Even with the random sample, the deluge of information created so much error that an approximate estimate was almost impossible. Moreover, many times, servers or (uncached) Website links that worked one day did not the next.

*Media Mentions.* Media mentions were counted using the three Lexis/Nexis databases – Major Newspapers, Magazine Stories (Combined), and Transcripts. The procedure for searching in each database was the same. Common variations or alias for the subjects were also used (i.e. Bell Hooks and Gloria Watkins or William and Bill Moyers). For example, Bill Moyers was searched under “sing (William or Bill) pre/2 Moyers.” “Sing” returns singular versions of Bill or William, avoiding hits on “Williams” or “Bills.” Using “(William or Bill)” returns mentions of both names and “pre/2” means that “(William or Bill)” must appear within two words before “Moyers.” This allows for middle initials or names.

Searches were run for the seven and half years between January 1995 through July 2002. If the individual had fewer than one thousand mentions (the upper limit that Lexis displays), the first one hundred were examined and the percentage correct was then applied to the total number of hits to produce the estimate of the total used in the tables. If he or she had more than one thousand mentions, a random sample of one hundred hits was examined in order to determine the percentage of correct hits. This involved breaking down the years searched from “January 1, 1995 through July 31, 2002” to smaller increments. Dividing the search into three periods was often sufficient enough, however, some searches, because of their large number of mentions, had to be broken down into months, weeks, or even days. This illustrates the sensitivity of the media search to time. *A* and *B* might have the same number of media mentions for the five-year period, but if *A* happened to have a “hot” week when his mentions were counted, this might create a misleading impression of his celebrity relative to *B*. At the same time, random sampling also tries to improve our estimation procedure in that “hot weeks” for *A* or for people with similar names to *A* do not bias the sample by looking at all time periods as opposed to possibly including one of those “hot weeks,” which would result in overestimation or underestimation.

Because the Transcripts database consists of transcriptions of the spoken word, spelling was erratic. An effort was made to account for the most likely misspellings by, for example, searching for “Stephen or Steven” and for “Rosenberg or Rosenburg.” Undoubtedly, however, some mentions were not picked up by the search.

Format was also a problem with this search. “Thurman pre/2 Arnold” also brings up mentions of Uma Thurman and Arnold Schwarzenegger, because they also fits the “Thurman pre/2 Arnold” criteria. But, these errors were corrected for during the sampling procedure.

*Scholarly Citations.* The counts of scholarly citations were performed using the *Science Citation Index*, the *Social Sciences Citation Index*, and the *Arts and Humanities Citation Index*, all compiled by the Institute for Scientific Information (ISI). These three databases were accessed using the Dialog Corporation’s

DialogLINK software. Scholarly citations were counted only for those individuals listed in Table 5.1 having full-time academic employment for a significant part of their career. All of these, regardless of field, were searched for in the *Social Sciences Citation Index*, on the theory that any academic who is a public intellectual and thus writing on issues of political or ideological character might be doing at least some scholarly work as a social scientist or of interest to social scientists. Academics whose fields are anthropology, biology, chemistry, computer science and technology, linguistics, medicine, and/or physics were also searched for in the *Science Citation Index*; those in architecture, classics, law, literature, philosophy, and/or theology, and writers were also searched for in the *Arts and Humanities Citation Index*. Persons in borderline fields – education, journalism, politics, and publishing – were searched for in either the science or the humanities index (or both), as seemed appropriate given the nature of the individual’s interests, as well as in the social science index.

The Dialog search embraced the publication years 1995<sup>1</sup> to approximately October 2002. Dialog search criteria follows a simple rule of B.E.S.T; whereby, “B” begins the search by opening files to be examined (i.e. the citation indexes), “E”<sup>2</sup> expands the search term by acting as a database thesaurus retrieving similar items or names, “S” selects the results by creating the set of records containing specific terms, and “T” selects the *type* of records that we’d like to observe. Hence, the inquiry for each subject began in their relevant database(s) as searches for hits as cited authors. The search also removed self-citations, excluding any citations by Nussbaum to her own work; and, again, the search was limited to the citing work’s publication years after and including 1995. And, for those searched in more than one citation index, the search also removed duplicates (with the command “rd”), since some journals are listed in more than one of the ISI databases. Thus, as an example, a typical search command would have been:

“B 439, 7; S ca= (Nussbaum M or Nussbaum MC) not au=(Nussbaum M or Nussbaum MC) and py>=1995; rd.”

One problem noted during the searches is the problem of citation format. For most academics, a middle initial was used in order to retrieve results; however, for some, the unavailability of a middle initial and a common last name, such as Kenneth Anderson, returned an extremely high number of results, but a low chance of finding correct hits. Thus, numbers from the 2000 tables and these new tables may show this discrepancy.

Once a set of citations to a given academic is created with the appropriate parameters (publication years, the removal of self-citations and duplicates), the set returning the correct records is then sampled. The search sampled thirty to

---

<sup>1</sup> Dialog includes in 1995 those journals entered into its system during that year, and so includes some late issues from 1994, while the late issues from 1995 appear in the 1996 category.

<sup>2</sup> Given that we knew exactly what we were searching for – hits that contained citations to academics listed in the table – the “E” command was not used.

forty citations that ranged the full time period by observing a random sample<sup>3</sup>. Thus, an example of the final command search would be:

“T s2/free,k/ 2, 69, 98, 111, 254, 311, 350...”

“s2” indicates that the records should be retrieved as a subset of the preceding search and “free” and “k” would return information about the citing article that could help us decipher whether the citation is a correct reference. After sampling the data and determining the correct percentage of records, the numbers are converted using the total number of returned records.

Typographical errors in Dialog or in the ISI databases may have resulted in some legitimate hits being missed. Those errors, however, are likely to be very small. Because sampling was limited to the most recent eight years, subjects whose academic careers and careers as a public intellectual did not coincide might come up misleadingly short on either media mentions or scholarly citations, depending on the rate at which citations depreciate. For example, a prominent scholar in a field in which scholarly citations depreciate rapidly (perhaps because it is a highly progressive field, such as physics) who switched into public intellectual work might have few scholarly citations yet be highly renowned as a scholar – that renown might have been what enabled him to make the switch! However, it is unclear that errors such as these are biased in favor of any of the hypotheses tested in the analysis.

---

<sup>3</sup> Because the records are arranged chronologically, observing the first thirty samples, as opposed to a random thirty throughout the list of records would have created a chronological bias towards the present.