

NET Institute*

www.NETinst.org

Working Paper #08-31

October 2008

**What's in a (Missing) Name?
Status and Signaling in Open Standards Development**

Tim Simcoe
University of Toronto

Dave Waguespack
University of Maryland

Lee Fleming
Harvard Business School

* The Networks, Electronic Commerce, and Telecommunications (“NET”) Institute, <http://www.NETinst.org>, is a non-profit institution devoted to research on network industries, electronic commerce, telecommunications, the Internet, “virtual networks” comprised of computers that share the same technical standard or operating system, and on network issues in general.

What's in a (Missing) Name? Status and Signaling in Open Standards Development *

Tim Simcoe

J.L. Rotman School of Management, University of Toronto

Dave Waguespack

Robert H. Smith School of Business, University of Maryland

Lee Fleming

Harvard Business School

October 2008

*We gratefully acknowledge financial support from the NET Institute (www.netinst.com) and the Kauffman Foundation. We also thank Ajay Agrawal and Scott Bradner for helpful comments. Address for correspondence: Joseph L. Rotman School of Management, 105 St. George Street, Toronto, ON M5S 3E6, Canada. E-mail: timothy.simcoe@rotman.utoronto.ca

What's in a (Missing) Name? Status and Signaling in Open Standards Development

Abstract

How much are we influenced by an author's identity? If identity matters, is it because we have a "taste for status" or because it offers a useful shortcut — a signal that is correlated with the likely importance of their ideas? This paper presents evidence from a natural experiment that took place at the Internet Engineering Task Force (IETF) — a community of engineers and computer scientists who develop the protocols used to run the Internet. The results suggest that IETF participants use authors' identity as a signal or filter, paying more attention to proposals from high-status authors, and this has a surprisingly large impact on publication outcomes. There is little evidence of a "taste" for status. JEL Codes: L1, O3

1 Introduction

How much are we influenced by an author’s identity as opposed to the actual quality of their work? If identity matters, is it because we have a “taste for status” or because it offers a useful shortcut — a signal that is correlated with the value of the work?

In a widely cited paper, Merton (1968) argued that reputation and identity are important factors in scientific careers. He provided qualitative evidence that famous scientists get a disproportionate share of the credit for joint discoveries, and labeled this phenomenon the “Matthew Effect” (after a Biblical reference to increasing returns).¹ One of Merton’s anecdotes is based on a story about Lord Rayleigh, a British mathematician and physicist who won the 1904 Nobel Prize in Physics:

“Rayleigh’s name ‘was either omitted or accidentally detached [from a manuscript] and the Committee turned it down as the work of one of those curious persons called paradoxers. However, when the authorship was discovered, the paper was found to have merits after all.”

This paper presents evidence of the Matthew Effect based on a large-scale replication of Rayleigh’s experience. We focus on a natural experiment that took place at the Internet Engineering Task Force (IETF), a community of engineers and computer scientists who develop the protocols used to run the Internet.

Between 1999 and 2004, many authors’ names were replaced by “et al” in a series of announcements that are an important part of the IETF publication process. We show that the use of “et al” was driven by administrative overload, particularly around IETF meeting dates, when there was typically a flood of new submissions. This created a situation where some authors’ names were randomly removed from announcements, while others’ were left in place. When “et al” obscures the name of an IETF Working Group chair (i.e. a high-status author, rough equivalent to a journal editor) there is a statistically significant drop in the likelihood of publication. There is little change when “et al” obscures the name of a relatively unaccomplished author. The magnitude of this result is striking, given that “et al” only removes authors’ names from an email announcement; it does not result in a blind review process.

In the literature on discrimination, the Matthew Effect is an example of what is called disparate treatment. Findings of disparate treatment lead immediately to the question of whether the observed discrimination was *statistical* or *taste-based* (or both). Statistical discrimination occurs when members of some group are treated differently because of differences in underlying

¹ “For unto every one that hath shall be given, and he shall have abundance: but from he that hath not shall be taken away even that which he hath.” (Matthew 25:29)

ability or past performance. For example, high-status authors may receive more attention because we expect them to produce better work. Taste-based discrimination occurs when groups receive different treatment without such justification.

We use citations to conduct an “outcomes test” for taste-based discrimination. The idea behind this test is that published work by high-status authors will receive fewer citations if these authors face a lower publication-quality threshold. We find no difference in the *ex post* citation rate of high and low-status authors. If citations are an accurate measure of quality, this is inconsistent with the hypothesis that there is a large “unjustified” asymmetry in the screening process that leads to publication.

If there is no taste-based discrimination, how could our relatively weak treatment produce a substantial change in outcomes? We use email data to suggest that one possibility is the importance of capturing attention early in the publication process. In particular, we show that when “et al” obscures the name of a high-status author, there is a substantial decline in the amount of “conversation” generated by the proposal on the IETF’s email listservs (where much of the work actually takes place). We do not identify a specific link between attention and outcomes. However, our results suggest that even in the absence of taste-based discrimination, there can be a large Matthew Effect (i.e. increasing returns to past-success). In particular, statistical discrimination may help those who are already successful with the often difficult task of “building an audience.”

This paper is related to earlier work on discrimination in the academic publishing process. For example, Blank (1991) conducted a randomized experiment on the impact of double-blind refereeing at the American Economic Review. She found that double-blind procedures led to true anonymity for only 54 percent of the referees, and increased the acceptance rate for authors from the most highly ranked institutions. Smart & Waldfogel (1996) also found evidence of reverse discrimination by comparing published papers’ citations to their observed editorial treatment (e.g. page length and placement within the journal). Peters & Ceci (1982) conducted a field experiment by re-submitting twelve papers published by “prestigious” authors to the psychology journals where they were originally published. Nine of the re-submissions went undetected, and of those eight were rejected on re-submission.

Our work also fits into the broader literature on discrimination. This paper is one of a handful that combine data on selection and performance. Other examples include Ayres & Waldfogel (1994) on discrimination in bail-setting, and Hellerstein & Neumark (2004) on labor-market discrimination. Unlike many papers that focus on the selection process, we identify a natural experiment — the random deletion of names by “et al” in announcements — that allows us to get around the problem of omitted variables. Randomization of names has also

been used to study racial discrimination in field experiments, such as Bertrand & Mullainathan (2004). Finally, the fact that “et al” has a large impact on the chance of publication — even though it is not particularly hard to find a complete list of authors — makes our paper relevant to the behavioral literature on “salience” or attention costs (see DellaVigna, 2008, for an overview).

The remainder of the paper proceeds as follows: Section 2 describes the IETF publication process and the “et al” natural experiment. Section 3 describes our data, measures and statistical methods. Section 4 presents results. Section 5 concludes.

2 Setting and Experiment

This section describes the IETF and the process that led to exogenous removal of author names from new proposal announcements.

2.1 The IETF Publication Process

The IETF is a non-profit organization that creates compatibility standards used to run the Internet. These standards provide a set of rules for new product design. By adhering to IETF standards, engineers can ensure that new products will be able to acquire an Internet address, encrypt a message, display a web page, play a video, or identify an e-mail attachment.

Anyone who is interested can become an IETF member simply by choosing to participate, though most members are engineers or computer scientists. While the IETF holds three plenary meetings each year, most work takes place on the Internet, where new ideas are introduced and debated on a series of email listservs. These listservs are often administered by an IETF Working Group (WG) that is developing new standards for a particular technology area.

Figure 1 provides an overview of the IETF publication process. Anyone can submit a proposal. These proposals are called Internet Drafts (IDs). An ID usually goes through a series of revisions before it is either discarded or accepted for publication. Published drafts are referred to as Request for Comments (RFC).

There are two ways for a draft to become an RFC. The first is to become the “consensus” recommendation of an IETF Working Group. This implies support from a solid majority of WG participants (though not unanimity). In practice, the Working Group chair usually determines whether there is a consensus. If a chair decides there is consensus support for an ID, it is forwarded to the Internet Engineering Steering Group (IESG) — a sort of editorial review board. The IESG then issues a last call to the entire IETF. If the last call raises serious issues, the Internet Draft is sent back to the Working Group. Otherwise it becomes an RFC.

The second way for a draft to become an RFC is through the “independent submission” process. In this case, an individual posts a new Internet Draft to a public web server maintained by the IETF. If the draft fits into the agenda of an existing Working Group, or there is sufficient interest in starting a new Working Group, the proposal might become a Working Group draft (though this rarely happens in practice). If the draft does not get picked up by a Working Group, but there is enough interest within the IETF community, the individual may submit it to the RFC editor and ask that it be published as an individual (non Working Group) RFC. The RFC editor typically sends these drafts out to subject matter experts for review. If the reviewers and RFC editor agree that the ID should be published, it is sent to the IESG to ensure that it does not conflict with the work of any IETF Working Groups. If the IESG approves, the draft is published as an RFC.

The Working Group process is used to publish more important RFCs. For example, only WG drafts can be formally designated as IETF Standards. Individual submissions are classified as either Experimental or Informational RFCs. Many WG drafts are essentially “commissioned” works, and when an individual submission is recognized as particularly important, it will generally be adopted by a Working Group. Because WG drafts are more likely to be commercially significant, the approval process can be political.

Still, RFCs published via the individual submission process can be influential. They often propose new uses for IETF technology, describe lessons learned from implementation, and pose new problems for IETF members to work on. The IETF receives a large number of unsolicited IDs, and RFCs published via the individual submission process are cited frequently. We focus on individual submissions in the statistical work below.

Internet Drafts that have not been revised in more than 6 months are said to have “expired” and are removed from the web page where all current IDs are posted for review. While it seems clear that some drafts on both the WG and individual-submission track are essentially rejected (i.e. informed the revision is a waste of time) we can only observe expiration — which might reflect either outright rejection or a decision by the author’s to abandon the project.

2.2 The “et al” Experiment

When a new Internet Draft is submitted to the IETF, it is posted on a web page where anyone can download it. With the initial posting (and for each subsequent revision) an e-mail announcement is sent to the entire IETF via the “ietf-announce” listserv. The top panel in Figure 2 shows a typical message sent to the ietf-announce list. The message contains a title, author list, filename (which contains information about WG affiliation and revision history), date, and abstract.

Before 1999 every announcement included the name of every author who contributed to the proposal. However, beginning in 2000, there was a rapid increase in the use of “et al” in the author list when IDs had more than one author. Our identification strategy exploits the fact that the names obscured by “et al” were occasionally prominent members of the IETF community. The lower panel in Figure 2 shows an ietf-announce message where one of the authors whose name does not appear is a former Working Group chair.

The decision to use “et al” was made by the IETF Secretariat — an administrative body that manages the ietf-announce listserv. The secretariat had a small staff that would process incoming drafts by typing the relevant information into a standardized form. Conversations with the director of the IETF Secretariat suggest that the decision to begin using “et al” for co-authored IDs was simply a response to the rapidly growing volume of submissions during that time period. Figure 3 shows the dramatic increase in ID submissions — particularly individual submissions which more than doubled between 1998 and 2000.

The decision whether to use “et al” on a particular draft was left up to individual clerical staff who typed in the ietf-announce messages. These individuals suggested that they tried to include every name. However, they would resort to “et al” when things became extremely busy. This typically occurred during the time-period immediately before an IETF meeting, when they often received a flood of new proposals. Figure 4 provides strong evidence to support these claims.

The top panel in Figure 4 is based on a kernel-density regression of “et al” usage on time (the unit of observation is a new ID submission with more than one author). The vertical bars represent IETF meeting dates. It is obvious that there is both an increasing trend towards using “et al” and a strong cyclical component that is tied to the meeting dates. The second panel in Figure 4 overlays the predicted “et al” probabilities with the results of a second kernel-density regression showing the rate of new submissions over time. From this figure, it is quite clear that submission volume is also driven by the existence of meeting-related deadlines. However, it is also clear that between late 2000 and mid-2004, any proposal with 2 or more authors had a 30 to 60 percent chance of receiving an “etal” in the announcement process.

We study the impact of individual status on publication outcomes by assuming that the use of “et al” produces an exogenous shock to the information set of ietf-announce readers. In particular, we study what happens when a “high status” name is removed from the message announcing a new draft. Note that this is an extremely weak treatment. In particular, any interested reader can obtain a complete list of authors by simply downloading a copy of the proposal and glancing at the front page. In most cases, the ietf-announce message contains a link to the draft, so this is very easy to do.

Our primary measure of status is a dummy variable that indicates whether an author has

ever chaired an IETF Working Group. The position of WG chair is similar to journal editor in many respects. Chairs decide whether and when a Working Group has reached consensus (though this decision is subject to IESG review). Consequently, these individuals have high visibility within the IETF.² Like journal editors, chairs are usually awarded their position on the basis of past publication success. In particular, Fleming & Waguespack (2008) show that past RFC publications increase the likelihood of becoming a chair.

3 Data and Methods

This section provides an overview of the data set and discusses the empirical methods used below.

3.1 Data

Using public sources, we collected information on every Internet Draft submitted to the IETF between 1992 and 2004. This sample contains 12,342 Internet Drafts. For each proposal we have a submission date, a complete list of authors, and an outcome: expiration (failure) or publication as an RFC.

For each author, we have a complete list of the Internet Drafts they submitted to the IETF, their place in the list of draft authors, an email address, and an indicator of when and whether the author ever served as a Working Group chair. We use authors' email addresses to construct a number of additional variables. For example, we can observe whether a proposal has co-authors from two different institutions (e.g. cisco.com and Harvard.edu), or whether any of the authors is affiliated with a "non-profit" institution (e.g. umd.edu or eff.org). Table 2 provides a complete list of variables and definitions.

Table 3 provides summary statistics by publication-year cohort to show how the "et al" experiment evolved over time. The second column in this table illustrates the large increase in total submissions between 1998 and 2002. This increase is especially large for individual submissions. The third column in Table 3 shows that the increase in submissions led to a corresponding decline in the probability of publication. It also shows that Working Group proposals are roughly 6 times more likely to become an RFC. Our analysis will focus on individual submissions.³ Column four shows the probability that an individual submission has one or more high-status authors. High status authors make up only 3 percent of the individual submissions. Column five in Table 3 shows that the use of "et al" in ietf-announce messages

²Chairs are not allowed to be an author on drafts submitted to their own Working Group.

³We did analyze Working Group drafts and find little evidence of a Matthew Effect in that sample.

increased from one percent of all proposals in 1998 to roughly 25 percent of proposals by 2000, and remained there until the practice was halted in late 2004. For a draft to receive an “et al” it must have more than one author. During our sample period, most drafts have two or more authors, and column seven shows that conditional on receiving an “et al” this figure is much higher. The final column in Table 3 shows that by 2001, the IETF secretariat had developed an informal norm for deciding where to place the cutoff when using “et al” in place of a complete list of names: for individual drafts, an “et al” message would only list the name of the first author.

In spite of the large number of proposals in our sample, Table 3 shows that we have a small numbers problem. In particular, just under 3 percent of all individual proposals have one or more Working Group chair authors. Moreover, only twenty percent of those submissions receive the “et al” treatment, and our design requires that in these cases, the Working Group chair cannot be the first author listed on a draft. The analysis below suggest that the Matthew Effect is large, but our estimates are often fairly imprecise.

3.2 Methods

3.2.1 Identifying the Matthew Effect

We have data from a census of Internet Drafts (indexed by i). Our main dependent variable RFC_i is a dummy variable that equals one if draft i is ultimately published as an RFC. The variable $Chair_i$ is a dummy that equals one if Internet Draft i has one or more authors who are a current or former WG Chair. We construct a variable $MissChair_i$ that equals one for Internet Drafts where there is a chair author, but that author is not listed in the draft announcement. If we assume that “et al” is exogenous, a very simple statistical test for the Matthew Effect would consider the null hypothesis:

$$E[RFC|Chair = 1, MissChair = 1] = E[RFC|Chair = 1, MissChair = 0]$$

for drafts with the same number of authors where the “high status” Working Group chair author occupies the same position in the list of all authors. We begin with a non-parametric test based on this idea, for drafts where there are two or three authors, one of whom is a chair, where the chair is not listed first. Since there are only 296 of these proposals (of which 54 are treated by the use of “et al”) the results are noisy, though statistically significant.

We then turn to a difference-in-differences model that estimates the impact of having a chair-author on the probability of publication, rather than conditioning it out. In particular, we estimate the following regression, where λ_t are a set of publication-year fixed effects, $EtAl_i$

is a main effect, and X_i are a set of Internet Draft characteristics (including a constant):

$$RFC_i = MissChair_i\beta_1 + Chair_i\beta_2 + EtAl_i\beta_3 + \lambda_t + X_i\delta + \varepsilon_i \quad (1)$$

In this specification, β_2 measures the impact of having a high-status author on the proposal. The Matthew Effect is captured by β_1 , which measures the change in publication rate when a chair’s name is removed from the ID announcement (relative to the change when a non-chair’s name is removed). Note that $MissChair_i$ is not quite an interaction between $EtAl$ and $Chair$; in particular, it equals zero for some proposals where $EtAl_i = 1$ and $Chair_i = 1$ but it is possible to see the name of the Working Group chair because the “et al” cutoff is relatively high.

Most studies of discrimination run a model like (1) without the “et al” interaction and test the hypothesis that β_2 is zero. The standard critique of that approach is that it is not possible to correct for spurious correlation between $Chair_i$ and ε_i (e.g. because of omitted variables). In our model, we assume that “et al” administers an exogenous shock to the information set of IETF participants. If this shock is truly random, the “interaction term” $MissChair_i$ and the main effect $EtAl_i$ will be uncorrelated with ε_i . If IETF members treat “et al” as a low-status author, then our specification should produce an unbiased estimate of the Matthew Effect. If they form a guess by allowing that there is some probability that “et al” is a high-status author, we estimate a lower bound on the true effect.

3.3 Explaining the Matthew Effect

While the methods described above are an appealing way to identify the Matthew Effect, they are not particularly good at providing an explanation for it. Specifically, β_1 is an estimate of the the impact of changing IETF members’ beliefs about proposal authors. To understand *why* these beliefs matter we would like to run a different experiment in which we randomly assign individuals to be Working Group chairs (or, less plausibly, hold the set of WG chairs fixed and randomly assign individual characteristics).

We can take one step in this direction by interacting $EtAl_i$ with author or proposal-level observable characteristics X_i , in a specification such as (1). This is akin to asking *which* authors or proposals produce the Matthew Effect. While this can provide useful evidence when there are a rich set of observable characteristics that a reader might use to discriminate between proposals, it immediately brings us back to the omitted variables problem — it is impossible to interact $EtAl_i$ with all of the potentially relevant author attributes. Put differently, even if we find a sizable Matthew Effect, it is not possible to discern whether IETF participants are screening on X or some (unobserved) alternative that is highly correlated with X .

An alternative approach to explaining the Matthew Effect is to search for evidence of the underlying mechanisms that lead to apparent discrimination. This seems particularly important in our case, where the treatment (omitting an authors name from an email announcement) is arguably quite weak. We suspect that a key mechanism in our setting is the amount of “attention” that a draft receives early in the submission and review process. A draft that receives very little attention seems more likely to be abandoned for a number of reasons: attention is needed for a Working Group to become interested in a proposal; authors may interpret lack of interest as an assessment of quality; and the quality of drafts that do receive early feedback may improve as a result. Given the rapid increase in total proposals, attention from prominent IETF members also seems likely to have been a scarce resource during our sample period.

We measure “attention” by counting the number of times an ID is mentioned in email messages to IETF listservs in the time-period immediately after it is announced. This measure focuses on replies to an initial message in order to distinguish a true conversation among IETF members from various automated messages that will mention a new draft. We use the difference in difference specification (1) to ask whether proposals where a high-status name is obscured by “et al” receive less attention than they otherwise would have.

A third approach to explaining the Matthew Effect is to use data on *ex post* outcomes to look for systematic differences across high and low status groups. The use of so-called “outcome tests” for discrimination was strongly advocated by Becker (1993), and Ayres (2002) provides an overview of the approach. In our setting, a natural measure of an ID’s *ex post* performance is the number of citations it receives. Intuitively, if WG chairs get a “free pass” in the review process, the marginal chair-authored RFC should receive fewer citations than a non-chair authored RFC.

To be specific, suppose that cites are a valid measure of proposal quality and that the IETF’s objective is to identify and publish high-quality proposals. The publication process produces a set of proposal-quality thresholds c_s , where $s \in H, L$ identifies an author’s status. An unbiased screening process would generate a common set of quality thresholds $c_H = c_L$, so the distribution of cites conditional on publication is identical across status-groups. Taste-based discrimination in favor of high-status authors would lead to a lower quality-threshold for that group, so that

$$E[Cites|s = H] = E[Cites|q > c_H] <= E[Cites|q > c_L] < E[Cites|s = L] \quad (2)$$

While straightforward, the outcomes test has been criticized for several reasons. Most often, critics point to the fact that it makes predictions about the quality of a *marginal* as

opposed to an *average* proposal. In order to operationalize the outcomes test, we must assume that proposals from high and low-status actors are drawn from the same underlying quality distribution. This assumption may not be very plausible. For example, if high-status authors produce work that is better on average, the outcomes test may show that average quality (conditional on acceptance) is equal even if there is discrimination against the low status types (i.e. $c_H < c_L$).⁴

A second, more obvious, criticism of outcomes tests (that has nevertheless received less attention) is that it is hard to determine whether an outcome measure is itself contaminated by taste-based discrimination.⁵ One approach to this problem is to use prices as an *ex post* performance measure (as in Ayres & Waldfogel, 1994) — on the assumption that markets are unbiased (at least at the margin). Lacking an obvious market-based performance measure, we consider an alternative approach. Specifically, we compare citation-based performance measures from three different sources: future RFCs, US patents and academic publications. We would expect academic publications and especially patents to be less biased, since they are produced by individuals who are (on average) less involved in the IETF process.

However, our approach does raise questions about whether there is really a one-dimensional quality measure. In particular, the objective function that IETF members apply to the screening process may not coincide with the quality perceptions of academic paper-writers or patenters. Nevertheless, the use of multiple performance measures brings additional information to bear on the question of whether the Matthew Effect is statistical or taste-based when there is a potential for similar biases in the citation process.

4 Results

We begin by illustrating our main result using both non-parametric and parametric approaches. We then consider three extensions. First we interact the “et al” dummy with other variables to look for the specific factors that drive screening behavior. Then we examine how status influences attention as measured by emails. Finally, we conduct an outcomes test for taste-based discrimination using citations as a measure of publication quality.

⁴Knowles, Persico & Todd (2001) developed a formal model of discrimination in a search process which generates the prediction (2) even when there is unobserved heterogeneity — including differences in the quality distribution across groups. The key to their model (and the resulting “hit rate” test) is that high and low status authors respond to the incentives created by the search process when deciding whether to submit a proposal.

⁵For example, in the literature on discrimination in police searches, most authors use convictions as the outcome. Of course, this is problematic if the judicial system is prejudiced against certain defendants.

4.1 The Matthew Effect

Table 4 present non-parametric evidence of discrimination in favor of high-status Working Group chairs. We begin with a sample of all individual Internet Drafts submitted to the IETF between 1999 and 2004 with either two or three authors, where the first author is *not* a Working Group chair. We then split this sample into proposals with one or more Working Group chair authors (left panel) and proposals that have no chair-authors (right panel). Finally, we compare means for the treatment sample, where the chair-author’s name is obscured by “et al” and the control sample, where the ietf-announce message contains a complete list of author names.

The first row in the left panel of Table 4 illustrates our main result. Proposals where the second- or third-author chair is not listed in the ietf-announce message are published as an RFC 4 percent of the time, compared to a 13 percent publication rate when the chair-author is listed. While there are only 54 proposals where a chair-author’s name is obscured by “et al” this large difference is statistically significant at the 1 percent level. The right panel shows that there is no change in the publication rate of “et al” proposals if they are not submitted by a high-status author. Moreover, the publication rate for low-status proposals (5 to 6 percent) is very close to the 4 percent publication rate for proposals where the high-status author is obscured in the ietf-announce message.

The second row in Table 4 examines the proportion of individual proposals that are picked up by a Working Group (which raises publication rates dramatically). These differences are small. The next three rows show that because “et al” drafts have more authors, they tend to have more of any variable that measures the diversity of authorship. There is no difference in Name length, which suggest that the use of “et al” is not discriminating against ethnic minorities or authors with particularly complicated names.

We now turn to a parametric model of the Matthew Effect, based on equation (5). Table ?? presents marginal effects from a logit regression (calculated at the means of X) along with t-statistics based on robust standard errors. The regression in the first column is based on the same sample as the left panel in Table 4: proposals with 2 or 3 authors, one of whom is a chair (not in the first position). While the average publication rate for these proposals is 10 percent, there is an 8 percentage point reduction when the chair-author’s name is obscured by “et al.” This is a remarkably large estimate (especially given given the weak nature of the treatment effect) which suggests that revealing an author’s identity can increase publication rates by a factor of four! We suspect that the Matthew Effect is particularly large in this particular setting, where a large number of proposals are being submitted, very few are likely to succeed and the rapid the growth of the Internet is placing demands on the attention of IETF participants.

One way to check that our estimates of the Matthew Effect are not spurious is to examine whether the use of “et al” also causes a large drop in the likelihood of publication when it does not obscure the name of a high-status author. The second column in Table 5 presents results from a difference-in-differences model that is analogous to comparing the right and left panels in Table 4. Estimates of the Matthew Effect are somewhat more reasonable in this model. The main effect of having a high-status author is to increase the publication rate from 6 to 13 percent. However, when a chair’s name is not revealed, that increase is only half as large (from 6 to 9 percent). Reassuringly, we find that the main effect of “et al” is neither negative nor statistically significant.

The last column in Table 5 extends this analysis to all proposals that do not have a Working Group Chair as the first listed author. This increases the number of observations where a chair’s name is obscured, but also adds substantial heterogeneity in the number of authors (which can grow quite large). In this regression the Matthew Effect coefficient declines again. The point estimate suggests that revealing an author’s identity explains roughly one third of the seven percent increase in publication rate associated with having a chair author. While this result is no longer statistically significant at conventional levels ($p = 0.11$), the point estimate seems more reasonable than the result in column one. Moreover, it suggest the sensible result that the Matthew Effect declines as the Working Group chair whose name is obscured moves farther down the list of all authors.

4.2 Explaining the Matthew Effect

4.2.1 What’s in a Name?

The previous sub-section showed that in spite of our relatively small natural experiment, there is statistically significant evidence of a fairly large Matthew Effect. We now ask what might be causing this result? As a first step in this direction, we turn to the specification in equation (1), and interact the “et al” dummy with several variables that IETF participants might use to screen proposals. Since it is difficult to interpret interaction coefficients in a nonlinear model (Ai & Norton, 2003) we estimate linear probability models. The sample includes all individual proposals that have between two and six authors and a non-chair as the first listed author.

The first column in Table 6 replicates our previous results on the Working Group Chair effect. The second column replaces the Working Group chair terms with a count of RFCs published by all authors on the draft, and an interaction term that counts RFCs published by authors who are not listed on the ietf-announce message. We find that a one log-point increase in RFC publications increases the publication rate by 4 percent. However, the effect is only half as large (i.e. declines by 2 percent) when those RFCs are produced by unlisted authors.

The third column of Table 6 considers another piece of information that “et al” may be obscuring: the number of different organizations who sponsor the proposal. We find that drafts with authors from more than one institutional affiliation are 6 percent more likely to be published. However, when “et al” obscures the name of all but one organization, this effect disappears.

The last column in Table 6 adds all of these factors to a single model. While none of the individual effects is statistically significant, all have the predicted sign and we can reject the joint null hypothesis that unlisted chairs, sponsors and past RFC publication are all zero at the 7 percent level.

To be clear, none of this information is available to readers of the ietf-announce messages, which only list the author’s names. Thus, we should interpret these results as providing evidence on the average reader’s knowledge of individual author attributes (based on their name) in addition to whatever quality inferences the reader draws based upon those attributes.

One of the things that is surprising about our findings is that readers appear to know a great deal about certain authors, and use this information to make informed guesses about proposal quality. Yet they do not bother to obtain complete author information (which is readily available in the actual proposals) even though they are generally “one click” away from learning this information. One interpretation of this finding is that in the early rounds of sorting through a large volume of papers that might be read, individuals have extremely high search costs. Our next set of regressions examine this idea more closely.

4.2.2 Status and Attention

Table 7 looks for evidence of a direct link between removing the name of a high-status author from the ietf-announce message, and the amount of attention that a proposal receives. We measure attention by counting the number of email replies that mention the Internet Draft across all of the IETF’s email listservs. On average, a new proposal generates about four replies. We count these replies over a 3, 7 and 14 day window.

Focusing on the first column in Table 7, we can see that there is a 33 percent increase in email response when a high-status author is listed on the ietf-announce message. However, two-thirds of that effect disappears when the chair’s name is removed from the announcement. We also find that there is a large negative main effect of having “et al” on the ietf-announce message. This may reflect the fact that “et al” is used when there are a large number of new proposals arriving, so IETF members are less likely to pay attention to any individual new submission.

Not surprisingly, we find similar results across all three e-mail based measures of atten-

tion. One half to two-thirds of the impact of having a high-status author is purely associated with having that author’s name on the draft announcement. While this “attention effect” is plausible — particularly given the large volume and preliminary nature of many of these proposals — it is nevertheless surprising that a difference of one or two email messages leads to a substantial divergence in publication probabilities. We interpret these findings as evidence of strong increasing returns to attention in the early stages of the creative process. It is not clear whether this is driven by unique features of the IETF publication process, or is a more general feature of creative work. however, we believe it to be an interesting subject for future research.

4.2.3 Status and Citation

Our final approach to understanding the mechanisms behind the Matthew Effect is to conduct outcome tests. In particular, we estimate Poisson models of RFC forward-citation rates to test the hypothesis that taste-based discrimination in favor of current or former Working Group chairs leads to a lower publication-quality threshold (and hence a lower forward-citation rate). We use citations from three different sources as a dependent variables: RFC cites, patent cites and cites from academic journal articles.

Table 8 presents results from our Poisson regression models. For all three citation-types, the coefficient on a Working Group chair dummy variable is statistically insignificant. In part this reflect a smaller sample size; since relatively few individually submitted Internet Drafts are published as RFCs. However, the magnitude of these coefficients are also small: a change of less than 10 percent in each case. The coefficients appear particularly small next to the control variables. For example, there are strong relationships between page counts and citations (for RFC and academic cites) and for drafts that have more than one sponsor.

The small and statistically insignificant effect on the chair dummy suggests that the large “et al” effects we find above are driven by statistical rather than taste-based discrimination. In other words, IETF members are using names to screen out proposals (as the results in Tables 6 and 7 suggest) but not applying a different standard to the proposals submitted by different types of author.

5 Conclusion

Many authors have written about the importance of labels and identity. Perhaps the most famous statement of the hypothesis that a name does not (or should not) matter belongs to Shakespeare:

What’s in a name? that which we call a rose

By any other name would smell as sweet;
So Romeo would, were he not Romeo call'd,
Retain that dear perfection which he owes
Without that title.

This paper presents evidence that Juliet was wrong, at least within the context of Internet standards development. We exploit a unique natural experiment created by the fact that “et al” obscures the names of some authors who nevertheless contribute to proposals brought before the Internet Engineering Task Force. We find that when “et al” obscures the name of a high-status author — specifically a current or former IETF Working Group chair — there is a significant drop in the publication rate. When “et al” obscures the name of a low-status author, there is no change in the likelihood of publication.

These results provide statistical evidence of the Matthew Effect first described by Merton, and our natural experiment provides a unique opportunity to separate the role of identity from the quality of an author’s ideas. Our estimates suggest that the Matthew Effect is surprisingly large in this setting, particularly given the relatively weak treatment — while “et al” obscures author names on an email announcement, it is relatively easy to find a complete list by downloading the relevant proposal.

We ask what gives rise to this large Matthew Effect; specifically, do author names serve as a signal of quality, or is there simply a “taste” for status that contributes to some individuals’ success? Evidence from the “et al” experiment suggests that names are primarily used as a signal: the screening process is one of statistical rather than taste-based discrimination. This conclusion is based on two pieces of evidence. First, publication rates respond to other information that gets obscured by “et al” (e.g. whether there is more than one firm sponsoring the proposal) and not just status cues. And second, citations to published RFCs suggest relatively little difference in the average quality of work produced by high and low status authors.

As an alternative to taste-based discrimination in favor of established authors, we suggest that increasing returns to attention may be a mechanism that explains how our relatively weak treatment can produce a rather large change in publication rates. We show that proposals from high-status authors generate more email conversation among IETF participants. There are several ways that increased early attention might contribute to a greater success in academic publishing, whether directly (getting picked up by a working group) or indirectly (improving the quality of the underlying ideas). Perhaps these “audience building” benefits are an important component of the status effects that sociologists observe in a wide variety of settings. This is an interesting question for future research.

References

- Ai, C. & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters*, 80(1), 123–129.
- Ayres, I. (2002). Outcome tests of racial disparities in police practices. *Justice Research and Policy*, 4, 132–142.
- Ayres, I. & Waldfogel, J. (1994). A market test for race discrimination in bail setting. *Stanford Law Review*, 46(5), 987–1047.
- Becker, G. S. (1993). Nobel lecture: The economic way of looking at behavior. *The Journal of Political Economy*, 101(3), 385–409.
- Bertrand, M. & Mullainathan, S. (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *American Economic Review*, 94(4), 991–1013.
- Blank, R. M. (1991). The effects of double-blind versus single-blind reviewing: Experimental evidence from the american economic review. *The American Economic Review*, 81(5), 1041–1067.
- DellaVigna, S. (2008). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, forthcoming.
- Fleming, L. & Waguespack, D. (2008). Scanning the commons? evidence on the benefits to startups of participation in open standards development. *Management Science*, forthcoming.
- Hellerstein, J. K. & Neumark, D. (2004). Production function and wage equation estimation with heterogeneous labor: Evidence from a new matched employer-employee data set. Working Paper 10325, National Bureau of Economic Research.
- Knowles, J., Persico, N., & Todd, P. (2001). Racial bias in motor vehicle searches: Theory and evidence. *The Journal of Political Economy*, 109(1), 203–229.
- Merton, R. K. (1968). The matthew effect in science. *Science*, 159, 56–63.
- Peters, D. & Ceci, S. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. *The Behavioral and Brain Sciences*, 5, 187–255.
- Smart, S. & Waldfogel, J. (1996). A citation-based test for discrimination at economics and finance journals. Working Paper 5460, National Bureau of Economic Research.

Tables and Figures

Table 1: Variable Definitions

Variable Name	Definition
Published as RFC	Indicator Variable = 1 if Drafts is published as an RFC
Became WG Draft	Indicator Variable = 1 if Individual Draft is picked up by IETF WG
Et Al Dummy	Indicator: One or more author names not listed in announcement
Chair Author	One or more authors is a past or current WG chair
Chair First	Chair Author = 1 <i>and</i> WG chair is listed as first author
Unlisted Chair	WG Chair Author is not listed in ietf-announce message
log RFCs	Log Count of cumulative RFCs Published by all draft authors
Multi-sponsor	Indicator: Draft authors affiliated with more than one organization
Multi-domain	Indicator: Draft has authors from both commercial and non-commercial TLD's
Name Length	Number of Letters in longest surname of any author
Publication Year	Year when draft is first submitted to IETF
Authors	Number of authors on draft
log Filesize	Log of draft file size in Kilobytes
RFC Cites	Citations from future RFCs to focal RFC
Patent Cites	Citations from US patents to focal RFC
Article Cites	Citations from academic journal articles to focal RFC
X-Day Replies	Count of e-mail replies containing ID filename to IETF listservs in X days

Table 2: Summary Statistics

Variable	Mean	S.D.	Min	Max	N
Published as RFC	0.21	0.41	0.00	1.00	9356
Et Al Dummy	0.20	0.40	0.00	1.00	9356
WG Dummy	0.29	0.46	0.00	1.00	9356
Publication Year	2001.26	1.84	1998.00	2004.00	9356
Authors	2.32	2.07	1.00	42.00	9356
log Filesize	10.19	0.79	7.50	15.47	9356
Affiliations	1.76	1.41	1.00	22.00	9356
Name Length	7.54	2.29	0.00	19.00	9356
Chair Author	0.30	0.46	0.00	1.00	9356
Chair First	0.19	0.40	0.00	1.00	9356
UnListed Chair	0.04	0.19	0.00	1.00	9356
log RFCs	0.99	1.15	0.00	4.94	9356
3-Day Replies	2.73	6.15	0.00	105.00	6215
7-Day Replies	3.41	7.58	0.00	122.00	6215
14-Day Replies	3.80	8.50	0.00	130.00	6215
RFC Cites	5.88	14.40	0.00	229.00	1943
Article Cites	0.84	5.40	0.00	187.00	1943
Patent Cites	0.68	3.49	0.00	91.00	1943

Table 3: Summary Stats by Cohort and Working Group Status

Publication Year	Drafts (Count)	Published (Percent)	Chair Auth (Percent)	Et Al (Percent)	Authors (Count)	Et Al ID Authors	Et Al Cutoff*
Individual Author Submissions							
1998	436	13.99	3.44	0.92	2.08	4.25	4.25
1999	644	11.96	2.80	1.55	2.40	9.80	4.60
2000	1,023	8.41	3.03	11.34	2.61	6.27	2.63
2001	1,211	7.93	2.48	29.89	2.55	4.75	2.16
2002	1,207	8.53	1.99	24.03	2.15	4.20	2.16
2003	1,068	7.12	2.81	26.12	1.97	3.66	2.20
2004	1,011	7.52	3.36	20.28	1.77	3.78	2.14
Total	6,600	8.71	2.76	19.18	2.23	4.40	2.23
Working Group Submissions							
1998	369	39.02	7.32	1.08	2.14	2.25	3.25
1999	434	44.24	6.91	5.07	2.55	7.55	7.55
2000	420	51.90	8.81	16.67	2.84	5.99	3.46
2001	491	44.60	8.76	40.73	3.28	5.35	2.25
2002	358	55.87	8.66	34.64	2.61	4.56	2.33
2003	356	58.71	5.34	31.74	2.18	3.93	2.23
2004	328	56.71	7.62	21.34	1.77	3.40	2.07
Total	2,756	49.64	7.69	21.88	2.54	4.83	2.58

*The number of authors listed on the announcement before “et al”

Table 4: Non-parametric Comparisons

	Individual Drafts 2 or 3 Authors (1 chair)			Individual Drafts 2 or 3 Authors (No chair)		
	Chair Unlisted	Chair Listed	P-value	Et Al	No Et Al	P-value
Published as RFC	0.04	0.13	0.01	0.06	0.05	0.64
Became WG Draft	0.02	0.03	0.51	0.01	0.01	0.97
Authors	2.87	2.36	0.00	2.25	2.74	0.00
log Filesize	10.42	10.07	0.00	10.17	10.30	0.00
Sponsors	2.13	1.88	0.02	1.55	1.80	0.00
Name Length	8.04	8.04	1.00	7.91	7.92	0.96
Multi-Sponsor	0.24	0.23	0.84	0.13	0.17	0.07
Publication Year	2002.50	2000.94	0.00	2000.98	2002.33	0.00
Observations	54	242		1,498	346	

Each observation is an Internet Draft submitted to the IETF between 1999 and 2004, with either 2 or 3 authors (one of whom is a WG Chair) and excluding all drafts where the WG chair is listed as the first author. Columns compare sample means for Drafts where the chair is listed in the ietf-announce message to cases where the chair is unlisted.

Table 5: The Matthew Effect

Logit Models of Internet Draft Publication			
Unit of Observation = Internet Draft			
Dependent Variable = Published as RFC			
Sample: Excludes WG Chair First-Authors			
Sample	2-3 Authors With WG Chair	2-3 Authors All Drafts	2-4 Authors All Drafts
Unlisted Chair	-0.08 (-2.02)**	-0.04 (-2.86)***	-0.02 (-1.59)
Chair Author		0.07 (3.11)***	0.07 (3.74)***
Et Al Dummy		0.02 (1.16)	0.02 (1.55)
log Filesize	-0.02 (-0.66)	-0.00 (-0.42)	0.00 (0.06)
Authors	0.01 (0.21)	-0.01 (-1.07)	-0.00 (-1.15)
Publication Year	-0.01 (-1.09)		
Pr[RFC] at Means of X	0.104	0.061	0.062
Observations	296	2,140	3,087

*10% significance; **5% significance; ***1% significance. Marginal Effects (at sample means) listed with robust t-stats in parentheses.

Table 6: What's in a (Missing) Name?

Linear Probability Models of Internet Draft Publication				
Unit of Observation = Internet Draft				
Dependent Variable = Published as RFC				
Chair Author	0.09 (7.64)***			0.03 (2.24)**
Unlisted Chair	-0.05 (-1.88)*			-0.02 (-0.59)
log RFCs		0.04 (8.90)***		0.03 (4.64)***
Unlisted RFCs		-0.02 (-2.15)**		-0.01 (-1.10)
Multi-Sponsor			0.06 (4.93)***	0.02 (2.06)**
Unlisted Sponsor			-0.05 (-3.28)***	-0.03 (-1.50)
Et Al Dummy	0.04 (2.44)**	0.04 (2.81)***	0.05 (3.19)***	0.05 (3.11)***
log Filesize	-0.00 (-0.38)	0.00 (0.03)	-0.00 (-0.52)	0.00 (0.00)
Authors	-0.03 (-0.86)	-0.03 (-0.93)	-0.04 (-1.05)	-0.04 (-1.04)
Authors Squared	0.00 (0.32)	0.00 (0.36)	0.00 (0.46)	0.00 (0.44)
Constant	0.22 (2.41)**	0.19 (2.05)**	0.23 (2.58)***	0.18 (2.02)**
Obs.	3213	3213	3213	3213
R-squared	0.03	0.04	0.02	0.04
Joint F-Test				P = 0.07*

+10% significance; *5% significance; **1% significance. Robust t-stats in parentheses.

Table 7: Status and Attention

Poisson Regression Models			
Sample = Individual Proposals: 1999-2004			
DV	3 Day Replies	7 Day Replies	14 Day Replies
Chair Author	0.33 (7.42)***	0.27 (6.81)***	0.23 (5.91)***
Unlisted Chair	-0.20 (-2.55)**	-0.13 (-1.84)*	-0.13 (-1.97)**
Et Al Dummy	-0.25 (-5.62)***	-0.26 (-6.29)***	-0.24 (-6.20)***
log Filesize	0.12 (5.46)***	0.10 (5.32)***	0.09 (4.75)***
Authors	0.22 (1.89)*	0.17 (1.57)	0.15 (1.50)
Authors Squared	-0.04 (-2.06)**	-0.03 (-1.82)*	-0.03 (-1.82)*
Multi-sponsor	0.07 (2.27)**	0.07 (2.53)**	0.10 (3.54)***
Constant	-2.45 (-7.92)***	-2.11 (-7.49)***	-1.72 (-6.45)***
Observations	1,794	1,794	1,794
ID Year Effects	Y	Y	Y

*10% significance; **5% significance; ***1% significance. Robust t-stats in parentheses.

Table 8: Status and Performance

Poisson Regression Models			
Sample = Non Working Group RFCs: 1999-2004			
Dependent Variable	RFC Cites	Patent Cites	Article Cites
WG Chair Author	0.07 (0.37)	-0.04 (-0.11)	-0.07 (-0.18)
log(Future RFCs)	0.14 (1.67)*	0.06 (0.33)	0.14 (0.69)
log(Pages)	0.64 (6.14)***	0.06 (0.28)	0.94 (3.81)***
Multi-Sponsor	0.62 (2.73)***	0.97 (1.91)*	-0.01 (-0.02)
Authors	-0.19 (-1.79)*	0.29 (1.92)*	-0.10 (-0.63)
Constant	0.49 (1.04)	-2.17 (-1.96)**	-1.63 (-1.86)*
Observations	429	429	429
RFC Year Effects	Y	Y	Y
RFC Type Effects	Y	Y	Y
ID Year Effects	Y	Y	Y

*10% significance; **5% significance; ***1% significance. Robust t-stats in parentheses.

Figure 1: The IETF Publication Process

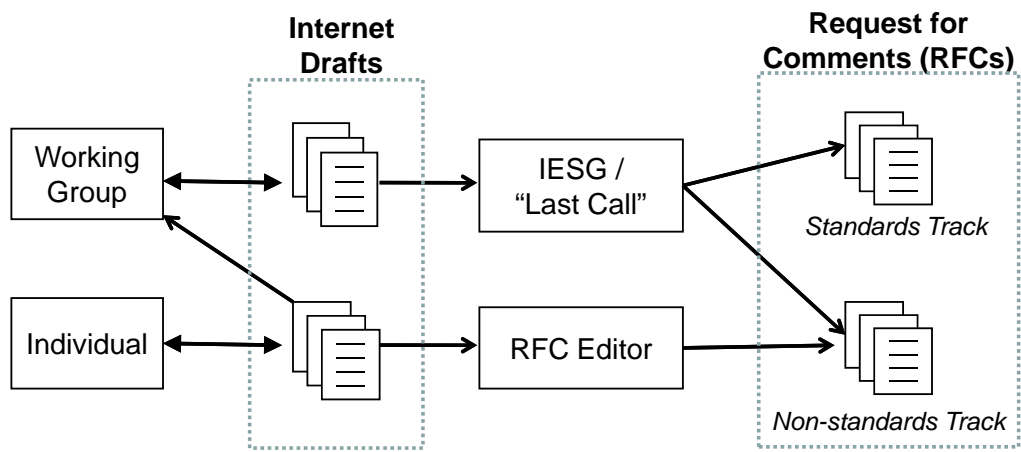


Figure 2: Typical ietf-announce messages

This message is the first version of a proposal (xpf2) published by the “xmldsig” Working Group. The 3rd author, J. REAGLE, is a Working Group chair. The proposal was eventually published as an RFC.

```
A New Internet-Draft is available from the on-line Internet-Drafts
directories.
This draft is a work item of the XML Digital Signatures Working Group of
the IETF.

      Title           : XML-Signature XPath Filter 2.0
      Author(s)       : J. Boyer, M. Hughes, J. Reagle
      Filename        : draft-ietf-xmldsig-xpf2-01.txt
      Pages           : 18
      Date            : 2003-8-22

XML Signature [XML-DSig] recommends a standard means for specifying
information content to be digitally signed and for representing the
resulting digital signatures in XML. Some applications require the
ability to specify a subset of a given XML document as the information
content to be signed. The XML Signature specification meets this
requirement with the XPath transform. However, this transform can be
difficult to implement efficiently with existing technologies. This
specification defines a new XML Signature transform to facilitate the
development of efficient document subsetting implementations that
interoperate under similar performance profiles.
```

This message is the second version of a proposal (iodef) published by the “inch” Working Group. The 2nd author (R. DANYLIW) is a Working Group chair whose name does not appear in the announcement. This series was not published.

```
A New Internet-Draft is available from the on-line Internet-Drafts
directories. This draft is a work item of the Extended Incident Handling
Working Group of the IETF.

      Title           : The Incident Data Exchange Format Data Model
and XML Implementation Document Type Definition
      Author(s)       : J. Meijer et al.
      Filename        : draft-ietf-inch-iodef-02.txt
      Pages           : 85
      Date            : 2003-9-30

The purpose of the Incident Data Exchange Format (IODEF) is to define
data formats for information related to computer security incidents
typically exchanged between collaborating Computer Security Incident
Response Teams (CSIRTs). The IODEF satisfies the requirements specified
in RFCXXX [1] This Internet-Draft describes a data model for
representing commonly exchanged incident information exported from
incident handling systems managed by CSIRTs. An implementation of the
data model in the Extensible Markup Language (XML) is presented, an XML
Document Type Definition is developed, and examples are provided.
```

Figure 3: New Internet Draft Submissions (1992-2004)

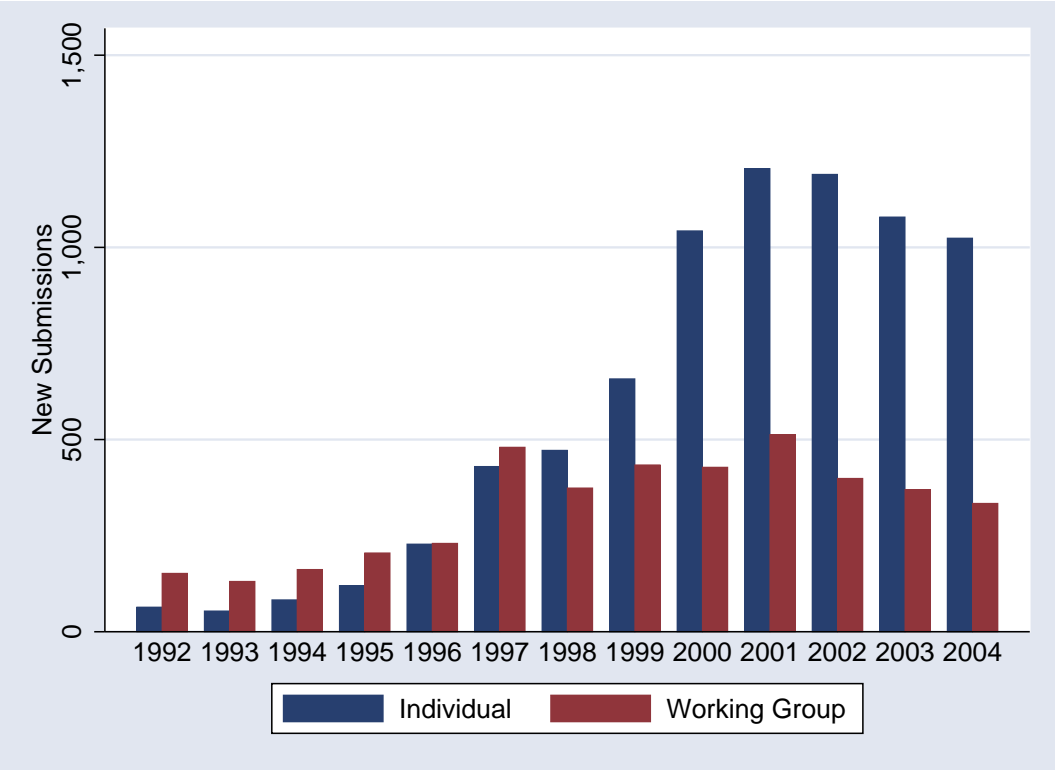


Figure 4: “Et al” Frequency, Meeting Dates, and Draft Submissions

