

# **Open Access and Accuracy: a comparison of authors' self-archived manuscripts and published articles**

David Goodman, Sarah Dowson and Jean Yaremchuk

## Abstract

Some approaches to Open Access (OA) use authors' manuscript copies for the OA version, in the form accepted after peer review but prior to full editing. Advocates of such approaches are certain that these versions differ only trivially from the publishers' versions; many of those who oppose them are equally certain that there can be major discrepancies. In a pilot study, we have examined the actual differences in a small number of such article pairs in the social sciences and in biology. Using an operational classification of the extent of error, we have determined that neither pronouncement is likely to be correct.

We found numerous small differences that affect readability between open access and publishers' versions. We also found a low frequency of potentially confusing errors, but sometimes it was the publisher's and sometimes the manuscript version that was more accurate. We found two cases where errors introduced by the publisher omit technical details that are necessary to evaluate the validity of the conclusions. However, we found no error that actually affected the validity of the data or results.

Interestingly, we did find problems with the stability of the document locations on authors' sites, and, in some cases, with their disappearance in PubMed Central after initial placement there.

## Introduction

There has been substantial conjecture as to whether Open Access (OA) repository versions of published articles differ in important ways from their formally published counterparts, to the extent that differences between them decrease the scientific validity of one or other version. Discussions (1) have included debate of the value of increased quality as authors' versions progress towards publication, through peer review and copy-editing, with those involved in the editing process offering anecdotal evidence in support of significant value added. Some argue that variations between publishers' and authors' versions of articles are not sufficiently significant to justify the existence of both. Others are concerned that the existence of OA versions presents the possibility of serious, perhaps even fatal, errors being presented in the absence of peer review and editing. (2), (3) The hypothetical example of a potential error in medical dosage data has been cited, (4) although no actual instance was demonstrated.

There seems to be consensus that articles can change as the move through the publishing processes, and version control is necessary. (5) Plans for 'Green OA' depend on authors' versions for open access and the published versions for 'the record'; (6) however, it is clear that publishers' versions must offer substantially more value to justify purchase. There is, however, no published quantitative study which assesses whether there are differences between open access and published pairs that are significant enough to affect meaning.

## Methods

### *Social Sciences*

Towards this end, we have undertaken two range-finding studies, focusing on distinct disciplines. The first was completed in late 2005 and encompassed twelve scholarly articles from eight journals in the social sciences, including the fields of economics, sociology and political science. The articles were selected from a list that Kristin Antelman (personal communication) prepared randomly from social science journals included in the *Social Science Citation Index*. The articles chosen for analysis included tables, graphs and equations, which present increased opportunity for errors. All had easily found authors' versions – in all cases examined as PDFs.

Two readers (DG and SD) performed independent reviews of the authors' and published versions of these articles. Introductions, abstracts and conclusions were read word-for-word. The body of the articles and the footnotes were read quickly to see whether the first and last phrases in the paragraphs were identical. If there were differences, the text was read closely, and differences marked. Equations, charts, tables, graphs, footnotes, and references were always compared closely, character by character. We then compared the results of our readings, and agreed upon the classification of the differences.

### *Biochemistry*

The comparison of social sciences articles was then replicated for twelve biochemistry articles, chosen from those with authors' manuscripts in PubMed Central (PMC) (7), and examined for differences between the author's manuscripts and published versions, using the PDF versions archived at PMC. The methodology was modified in that one reader (JY) followed the complete process of reading introductions, abstracts and conclusions word-for-word, skimming the body of the articles and the footnotes, and reading the full text if differences were found in first or last sentences. Equations, charts, tables, graphs, footnotes, references were always compared closely. All numerical values and units, including those in materials and methods sections, were also compared closely. Differences were noted and categorized as in the first study, and marked copies were reviewed by a second reader (DG).

### *Classification of Differences.*

These were categorized as follows:

*Type 1.* Errors normally adjusted by proofreading. They include typos, caps versus lower case, minor placement differences in tables, simple grammatical fixes, and standardizations for house style – e.g. ‘Fig.’ for ‘Figure’. The total figure for these was a minimum, as probably not all errors were detected; they are not reported in detail here. Corrections for capitalization and spelling affected scientific terms as well as ordinary English, but no case was found where there was any ambiguity in what the author had intended. No mistakes were found or corrected involving errors in scientific units, or in decimal places.

*Type 2.* These are normally adjusted by copy-editing without reference to the author or editor. These include substitution of words or phrases, location on the page, and minor additions, clarifications or edits in the published version. Readers with rudimentary subject knowledge would have no trouble with either version. Again, the total figure was a minimum and they are not reported in detail here; the distinction between Type 1 and Type 2 is not important to our results. However, many errors of both types were found in plain text as well as in the tables and figures, although in such context, the meaning was normally unambiguous.

*Type 3.* These include omissions or changes that a copy-editor would not necessarily have been able to correct. While these might confuse non-expert readers of these papers, such as first-year graduate students, they do not affect the validity of the data or the conclusions. (In practice, however, these differences would result in two readers, each with a different version of the article, having difficulty following line by line when communicating via email or telephone.) We did not class additions to the references as type 3 errors, because the normal updating of references just before publication made this difficult to analyze. The type 3 errors found here are reported in detail in Appendices A and B. In addition, we found one paper where the paper posted as the author's version seemed quite different from the final manuscript approved by peer review.

*Type 4.* More severe differences (e.g. in data or conclusions), which could lead two readers, each with a different article version, drawing different conclusions. Differences of this magnitude would be those affecting the actual data or their interpretation. An error of this severity in the author's version should have been caught in peer review (and thus been corrected in the manuscript), but was at least found and corrected by the publisher subsequently. It is also possible that errors in the author's version remained uncorrected in the publisher's version, but we did not examine for this. An error of this type which was introduced in the publisher's version would be a production failure, and might require a published correction. We found two such possible errors, both in data relevant to the statistical validity of the conclusions. Both are missing lines in tables, apparently omitted in error by the publisher; they are discussed in Appendix A.

*Type 5.* Differences affecting the validity of the paper, to the extent of justifying a public retraction. Authors' errors of this type should certainly have been caught in peer review (and thus been corrected in the manuscript); a difference such as this represents a major failure of peer-review. However, no errors of this type were identified in the study.

For social science, the authors relied on their common experience of social science research papers, supplemented by that of an economics specialist (TK); for biochemistry, on DG.

## Findings: social science

No comparison revealed differences affecting the validity of either the author's version or the published version of the same article.

The differences found in social sciences article pairs ranged from none (apparently exact PDF copies) to extensive rewrites in the published version. For seven of the articles, no differences above type 2 were found: in one pair, the author's version was an exact PDF of the journal article; in another set the articles were virtually identical; and for a third, the published article solely provided more references. The other four pairs in this group contained editorial changes, standardization of charts and abbreviations, corrections of typographical errors, and some minor rewrites in the published versions.

Of the remaining five pairs with higher grade differences, three authors' versions of articles were fuller than the published versions; one contained a line of copy in a table omitted in the published article; another contained an additional informative footnote; and the third had explanatory notes (which were omitted from the published version) under three figures.

In one of the remaining two pairs, the publisher's version contained additions to and reworking of tables, which were not present in the author's version. The other pair had two appendices in the author's version, one of which was truncated in the publisher's version, which directed the reader to another URL for further information in a PDF containing the appendices.

While a sample size of twelve is not sufficient to draw strong conclusions, the range of social sciences articles examined, as well as the range of differences found, suggests that the variance between open access and subscription access versions does not impact the conclusions of the articles. However, this pilot study also demonstrates that there is no clear *a priori* choice of a superior version. For some pairs, the author's version was the more accurate, while for others the published article was preferable. See Table 1.

Table 1. Comparison of Social Science Article Pairs

<b>Paper</b>	<b>Type</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>Comments</b>
<b>1</b>	Global Sourcing (PDF)	0	0	0	0	0	Author's version is exact copy of publisher's version (not included in count)
<b>1</b>	Global Sourcing (NBER working paper)	>25	10	3	0	0	V of infinity vs. '1-V'; differences in k and z variable; publisher's version generally clearer
<b>2</b>	Poverty and Access to Roads in Papua New Guinea	>10	6	1	1	0	Money definitions better in publisher's version but t statistics line in table was omitted by publisher
<b>3</b>	The Rise of the Regulatory State	9	4	0	0	0	Two type 2 errors in author's version; still uncorrected in publisher's version
<b>4</b>	Partner Selection in Public Goods Experiments	2	12	3	0	0	Extensive rewrites in publisher's version; major difference in one figure
<b>5</b>	The Genesis and Evolution of Urban Policy	4	7	0	0	0	Author's manuscript had additional references
<b>6</b>	The Place of Locational Analysis	7	0	0	0	0	No significant differences found.
<b>7</b>	What is Field Theory?	7	15	0	0	0	Publisher's version added seven references and deleted two
<b>8</b>	Constructing Social Problems in an Age of Globalization	8	21	3	1	0	Author's version fuller; confusing difference in one table; publisher omits line with N values
<b>9</b>	Cyclical Dynamics in Idiosyncratic Labor Market Risk	13	13	3*	0	0	Tables differ considerably; publisher's version has extensive additional data
<b>10</b>	Differentiated Products Demand Systems From a Combination of Micro and Macro Data	21	13	0	0	0	Publisher's version clearer in places
<b>11</b>	Liquidity Constraints, Household Wealth and Entrepreneurship	16	11	3	0	0	Author's version has more information in legend
<b>12</b>	Judicial Checks and Balances	24	26	1	0	0	Definitions, and appendices A and B, published as supplement in publisher's version

\*These represent a major difference in the content - see appendix A.

### Findings: biochemistry

As was found in the review of social sciences articles, the majority of the biochemistry pairs had no type 3 or higher differences. Of the twelve article pairs reviewed, eight were virtually identical, although some had minor editorial changes in the published versions. In two article copy-editing had brought about significant improvements in the publishers' versions by correcting errors, including the addition of relatively unimportant missing lines in tables; even there, however, the quality of the authors' manuscripts was in all cases sufficient for scientific purposes. Two of the pairs had significant differences in overall quality, with extensive editing in the published versions. Not coincidentally, the authors of both of these are non-native English speakers; the differences indicated that the publishers had gone to some effort to correct the style, but did not completely rewrite unclear or non-idiomatic passages to the same standard as other articles in the same journals. Interestingly, the two journals which provided the greatest improvement in quality from copy-editing were top-tier publications by society publishers: *Biochemistry* (American Chemical Society) and *Blood* (American Society of Hematology). See Table 2.

Table 2. Comparison of Biochemistry Article Pairs

Item	Title Type	1	2	3	4	5	Comments
1	Peptide Hairpins with Strand Segments Containing $\alpha$ - and $\beta$ - Amino Acid Residues	10	6	1	0	0	Some typos corrected in publisher's version, but new ones added
2	RAGE and Amyloid Beta Interactions	0	0	0	0	0	Exact copy
3	DNA Strand Arrangement within Sfil-DNA Complex	21	12	2	0	0	Non-native English language author; extensive edits in publisher's version
4	Identification and Characterization of Mechanistically Distinct Inducers of $\gamma$ -globin Transcription	24	0	0	0	0	Publisher's version edited for house style
5	Enzymatic Characterization of Human Mitochondrial C <sub>1</sub> -tetrahydrofolate Synthase	3	0	0	0	0	Publisher's version edited for house style
6	Extremely Low Penetrance of Hearing Loss in Four Chinese Families with the mitochondrial 12S rRNA A1555G mutation	0	0	0	0	0	Same text and tables in both versions, the only difference was that the author's version was in PMC standard format.
7	Integrin $\alpha_D\beta_2$ , an Adhesion Regulator Upregulated on Macrophage Foam Cells, Exhibits Multiligand-binding Properties	18	1	1	0	0	Non-native English language author; extensive minor edits; no abbreviation list in publisher's version
8	Two Visual Processing Pathways are Targeted by Gonadotropin-releasing Hormone in the Retina	0	0	0	0	0	Same text in both versions, even when editing would have been warranted
9	Roles of Cell-intrinsic and Microenvironmental Factors in Photoreceptor Cell Differentiation	2	0	0	0	0	Identical text; published charts easier to read
10	Electrical Stimulation Directly Induces Pre-angiogenic Responses in Vascular Endothelial Cells by Signaling through VEGH Receptors	4	0	0	0	0	Several edits
11	Instability of the Biotin-protein Bond in Human	2	1	0	0	0	Typos corrected, line of clarification added to table in

	Plasma						publisher's version
<b>12</b>	Two Isoforms of a Divalent Metal Transporter (DMTI) in <i>Schistosoma mansoni</i> Suggest a Surface-associated Pathway for Iron Absorption in Schistosomes	2	0	0	0	0	Minor grammatical edits in publisher's version

## Conclusions

There was a large number of errors in punctuation, spelling, or sentence construction in some authors' manuscripts that were corrected in the final publishers' versions, though a few obvious errors were missed, and a few new ones introduced. Twenty-three differences were found that had the potential to cause confusion. Usually the publisher's version was clearer, but in some instances the author's version was clearer. In almost all cases the conclusions could not reasonably have been affected, and any reader other than a raw beginner in the subject should have been able to tell what was intended. The one error that might conceivably have affected the conclusions, a missing statistic, was introduced by the publisher. The errors were not evenly distributed; 14 out of the 24 papers showed no Type 3 or 4 errors (see Table 3).

Table 3. Distribution of Type 3 or 4 errors

Number of errors per paper	Number of papers	Total errors
0	17	0
1	3	3
2	2	4
3	4	12
4	1	4
5 or more	0	0
Total	24	23

A number of uncorrected errors from the authors' manuscript copies remained in the published version, and a few additional errors were introduced (see Table 4). In some cases, the author's manuscript was identical to the published version; we do not know to what extent the respective publishers may have permitted self-archiving of the final published version in each particular instance. (8)

In several instances, the corrections made by publishers to the authors' versions that affected data or other substantial matter were not reflected in the authors' self-archived versions. These include issues that must have been referred to the authors as queries, so that they would have had the

opportunity to amend their versions. Again, it may be that the publishers did not permit this. We initially wondered whether there would be any instances affecting scientific quality, but we did not find any. In the one case where truly major differences were found, it appears that the version of the manuscript that the author posted was different from the final peer-reviewed version; additional tables has apparently been added to the one sent for publication.

Table 4. Source of Type 3 or 4 errors

	Number of errors	Examples
Publisher	3	A1, A2
Author	13	A1, A2 ,A4, A9, B1, B3
Unclear or Joint	7	A8 , A9, A11

A: Appendix A, Social science

B: Appendix B, Biochemistry

A recent study by Wates and Campbell (9) examined the author queries raised by copy-editors working for Blackwell Publishing. They examined corrections for 189 articles in a wide range of subjects, and found that only 6 of them might have been significant. This is an error rate two orders of magnitude smaller than our own findings. The differences may be accounted for by a number of factors (quite apart from the very small size of our own sample). First, the authors writing for the Blackwell journals may have prepared their copy more carefully than some of the authors whose manuscripts we examined here, and thus there may have been fewer errors to correct. Second, the previous copy preparation by Blackwell's peer-reviewers and editors, before the copy-editing stage, may have been more careful than the average we found. Third, we examined errors that were missed by publishers' copy-editing as well as those that were found. There may have been additional errors that were not queried by Blackwell's copy-editors because they were not noticed; in our study, we found several such errors. Fourth, errors may also have been introduced in the process of copy-editing and preparing the manuscript for publication; in our study, we found several of these.

In our study, there was no question that the work of the copy-editors improved the readability of most papers, and in some cases the improvement was very noticeable. However, the added value

seems to be primarily in readability, rather than in the correction of scientifically significant errors. The process of copy-editing corrected some minor errors which could have caused confusion to beginners, but occasionally introduced others. None of the errors found and corrected by copy-editing were major; either these papers contained no major errors after peer review, or the publisher did not find them. The two scientifically relevant errors found, in each case the omission of a statistic from a table, were introduced by their respective publishers.

Our own study was on a very small scale, and it would be necessary to confirm the results by more extensive studies, covering more subjects and publishers, and different types of papers. Our confidence in our findings is greatly increased by the findings of the much larger-scale Blackwell study, which found even fewer minor author errors (corresponding with our Type 3), and – like our own study - no major ones (corresponding with our Type 5). However, only one step in the information chain was studied. The implications of the findings respecting the suitability of repository versions require further interpretation.

## Implications

These findings give a preliminary indication that repository versions of articles may be of adequate quality for practical purposes; further refinement, while desirable, is not necessary if it introduces cost or decreases access.

The research also indicates that no version will be free from errors. It is generally understood that any error-correction procedure is capable of introducing new errors, and one unambiguous case was found (Article A1). Similarly, the resetting of tables and other matter that cannot be read as sentences offers opportunity for error, and introduced errors were indeed found; omission of lines of text is possible in any transcription process, and we found at least two examples, in one of which the tables in question had clearly been recomposed. In summary, we would say that the publisher's version is generally preferable, unless the publisher introduces significant errors.

On the other hand, authors appeared relatively indifferent to minor differences in presentation of equations. At least in the instances represented here, the articles were obviously not prepared using specialised software, such as TeX, and the publishers were clearly expected to reset the equations; in all cases, they appear to have done so accurately. There was a similar degree of skill in improving their quality, and a number of cases of small and large ambiguities of presentation were resolved (see Table 5).

Table 5. Preferred version, with regard to (a) major errors and (b) readability

	Number of papers	Examples
Publisher's version	16	A1, A4, A6, A7, A9 , A10 , A11, B1, B3, B4, B5, B7. B9, B10, B11, B12
Author's version	2	A2, A8,
Different, neither version unequivocally preferred	2	A3, A12
Equally good	4	A5, B2, B6, B8

Other problems are possible with self-archived versions. A systematic check of the stability of archiving showed disquieting results: of the 11 social science papers that had been self-archived on faculty websites, only five were available at the same location 15 months later. The other six remained available at a relatively easily findable URL - five of them were elsewhere at the university web site, although they could only be located by searching for the article, either through the university's search engine or by a general-purpose search; none of the original locations had links to the new location. The last article was no longer on the university site of the first author, but was found posted on the site of one of the other authors. The one paper which had been self-archived on the disciplinary archive Social Science Research Network (SSRN) also had a changed URL without any direct link to it. This finding is particularly disturbing in view of the recent confirmation that researchers consider self-archiving on their own web page to be sufficient, and do not think that organized institutional archiving is necessary. (10)

For the 12 papers on PubMed Central, three were no longer findable on 1 March 2007, though they had been there in June 2006. According to PMC, they were removed 'because ... deposit conflicted with a pre-existing agreement between the author and the publisher' (automated message, March 8, 2007). This seems a very high proportion, but perhaps it is due to their relatively early posting, before authors clearly understood publishers posting policies.

This research indicates the general feasibility of the Green (self-archiving) Open Access approach, as far as the scientific accuracy of the articles is concerned, although readability is generally markedly improved in the published version, particularly if the author was not a native English speaker; however, it appears that 100% error-free publishing is not being achieved. Other aspects were not systematically examined, but it appears that stability of self-archived material may be a problem, whether articles are deposited in centralized repositories or on the authors' own sites. Interestingly, none of the papers found here was located in an organized institutional repository.

While differences were indeed found in article pair comparisons, none was serious enough to invalidate significant data, conclusions or overall validity of the findings; none would warrant a published correction or retraction. Larger samples, and comparison across additional fields of

study, are needed to draw stronger conclusions, but hopefully the door to quantifiable review has been opened.

*A preliminary version of this paper was presented at the 2006 Charleston Conference, on Thursday, November 9, 2006 under the title 'Open Access and Accuracy'*

David Goodman

Previously, Palmer School of Librarianship at Long Island University, and Princeton University Library

dgoodman@princeton.edu

Sarah Dowson

Palmer School of Librarianship at Long Island University

Jean Yaremchuk

Palmer School of Librarianship at Long Island University

#### Acknowledgement

*We acknowledge with thanks the assistance of Thomas Krichel of the Palmer School of Librarianship at Long Island University in interpreting some of the equations in the social science papers, and the assistance of E. M. L. Goodman of New York City Technical College of the City University of New York, in final copyediting.*

## **Appendix A: Type 3 and 4 differences between social science article pairs**

(These articles were originally accessed in December 2005, with notes on access added on 1 March 2007.)

**Article A 1:** Antras, P and Helpman, E. 2004. Global Sourcing. *Journal of Political Economics*. 112: 552-580

<http://post.economics.harvard.edu/faculty/antras/papers/GlobalSourcingFinal.pdf> (exact copy) (confirmed, 1 March 2007) (note: not counted in tabulation of errors)

<http://www.nber.org/papers/w10082> (NBER working paper) (confirmed, subscription access)

### **Three type 3 differences**

1) Equation 13, page 567 in the journal (page 15 in the author's version). The journal has the mathematical expression, 'v of infinity' while the NBER working paper-print has '1-V'. Either form is acceptable, but the difference would confuse a beginner. The journal's form is the more usual.

2) Different uses of k and z variables: on page 561 in the journal (pages 9 and 10 in the author's version), the variable k is defined as 'ownership structure k'. In both versions, the variable k is properly used in equations 6 and 7. However, for Equation 16 (page 570 in the journal, page 19 in the author's version), the journal uses a 'z' variable for a Pareto function to distinguish it from the 'k' used earlier, while the author's version uses the same designation 'k' in this different sense. This difference is repeated on each of several following pages. The journal version is obviously clearer.

3) On page 572, the journal's equation 18 uses z; the author's version on page 21 uses k. It seems the journal was wrong to use z here; the previous correction was in this instance wrongly applied. In correcting the preceding problem, the journal made matters worse in this case.

**Article A 2:** Gibson, J and Rozelle, S. 2003. Poverty and Access to Roads in Papua New Guinea. *Economic Development and Cultural Change*. 52: 159-285 .

[http://www.mngt.waikato.ac.nz/depts/econ/staff/johng/PDF\\_Files/Published/EDCCRoadsPoverty.pdf](http://www.mngt.waikato.ac.nz/depts/econ/staff/johng/PDF_Files/Published/EDCCRoadsPoverty.pdf) (Not at this URL on 1 March 2007, but at

<http://www.mngt.waikato.ac.nz/Departments/staff/johng/Published.html>.)

### **One type 3 difference, one probably type 4**

1) On pages 169 and 170, the journal's Figures 3 and 4 have '1 kina = 100 toea' under them. The author's version does not define 'toea' although the y-axis of the Figure 3 and 4 graphs is measured in toea. The publisher's version is clearer.

2) On page 178 in the journal, a line with the t-statistic for a variable in a table is omitted. The author's version on page 31 has the line. The variable is average years of schooling for adults. The publisher omitted this line in error (possibly at a stage subsequent to copy-editing).

In this particular case, this variable is used in the model and therefore the t statistic is necessary to determine the validity of the article; an unexpectedly high value might have cast doubt upon the conclusion. The value (which is not unexpectedly high) was given correctly in the author's version. Peer-reviewers would be expected to catch such an omission; unless provided, it is difficult to evaluate the validity of the conclusion. This probably should be considered a type 4 error, introduced by the publisher. (It might, however, be considered a type 3 error, on the grounds that students would have assumed the value acceptable.)

**Article A3:** Glaeser, E L and Schleifer, A. 2003. The Rise of the Regulatory State. *Journal of Economic Literature*. 41: 401-426.

[http://post.economics.harvard.edu/faculty/schleifer/papers/rise\\_JEL\\_JAN\\_2003.pdf](http://post.economics.harvard.edu/faculty/schleifer/papers/rise_JEL_JAN_2003.pdf)

(Not at this URL on 1 March 2007 but at

[http://www.economics.harvard.edu/faculty/shleifer/papers/rise\\_JEL\\_JAN\\_2003.pdf](http://www.economics.harvard.edu/faculty/shleifer/papers/rise_JEL_JAN_2003.pdf) .)

### **No type 3 differences**

**Article A4:** Coricelli, G, Fehr, D and Fellner, G. 2004. Partner Selection in Public Goods Experiments. *The Journal of Conflict Resolution*. 48: 356-379.

[http://papers.ssrn.com/sol3/Delivery.cfm/SSRN\\_ID467](http://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID467)

(Not present at the original URL on 1 March 2007, but at <http://ssrn.com/abstract=467021>.)

### **Three type 3 differences, 1 serious**

1) The questionnaire and accompanying instructions were published by the journal as supplementary online material with a link on page 356 of the journal; they were presented in the author's version as a 16-page appendix. Either way is acceptable, but the journal version is the usual formal style for publication.

2) On page 362 of the journal (page 8 of the author's version), the journal has the 'Experimental Procedure' paragraph as '...conducted in one session with 24 participants and one session with 12 participants for each treatment.' The author's version has '...6 sessions with 24 subjects each.' Apparently, there were 6 sessions with 24 participants each, adding up to the 144 participants overall. The journal version is more exact. It seems possible that the added data might have been requested during peer review, but not corrected by the author in his version

3) On page 368 in the journal, (page 14 of the author's version), Figure 4 has slightly different data. For one of the variables, the journal shows  $n=20$ , and no outliers are shown; the author's version has  $n=25$ , and there is an asterisk signifying outlier data. The information content is very similar, though potentially confusing. This can be considered more serious than the usual type 3 error; beginners might not have realized that the differences primarily represent a choice of presentation. The publisher's version reflects the preferred presentation.

**Article A5:** Uitermark, J. 2005. The Genesis and Evolution of Urban Policy: A Confrontation of Regulationist and Governmentality Approaches. *Political Geography*. 24: 137-163.

<http://www.justusuitermark.nl/Documents/pg-html.doc> (Not at this URL on 1 March 2007, but at

<http://www.justusuitermark.nl/Documents/pg2005.pdf>)

### **No type 3 differences**

**Article A6:** Barnes, T J. 2003. The Place of Locational Analysis: A Selective and Interpretive History. *Progress in Human Geography*. 27: 69-95.

<http://www.geog.ubc.ca/~tbarnes/place%20final%20revised.doc> (Not at this URL on 1 March 2007, but at [http://www.geog.ubc.ca/~tbarnes/pdf/PAPER\\_place\\_of\\_locational\\_analysis.pdf](http://www.geog.ubc.ca/~tbarnes/pdf/PAPER_place_of_locational_analysis.pdf).)

### **No type 3 differences**

**Article A7:** Martin, J L. 2003. What is Field Theory? *The American Journal of Sociology*. 109: 1-49.

<http://www.csudh.edu/dearhabermas/lewinkurt01.pdf> (confirmed, 1 March 2007)

### **No type 3 differences**

**Article A8:** Benson, R and Saguy, A C. 2005. Constructing Social Problems in an Age of Globalization. *American Sociological Review*. 70: 233-260.

<http://www.soc.ucla.edu/faculty/saguy/constructingsocial.pdf> (confirmed, 1 March 2007)

### **Three type 3 differences, one type 4**

1) Table 2 (page 249 of the journal and page 57 of the author's version) has different figures in the top row. The table concerns the differences between US and French immigration stories, shown as US– France. The journal shows this as 0.06 and (minus) 0.06. The OA version has 0.05 and 0.05, thus not specifying the direction, although it is implied by the other data in the article. The author intended this as an absolute difference; the journal interpreted it as a numerical subtraction. The difference in the numbers appears to be a rounding error. The meaning is the same, though both presentations are somewhat unclear.

2) Footnote 12 (page 252 of the journal, page 59 of the author's version) differs. The journal's footnote 12 leaves out a paragraph of peripheral but interesting information included in the OA version, and uses a less specific conclusion. The journal editor apparently removed this as

unrelated, but the author disagreed. The journal version would be what most editors would prefer.

3) In Table 4, the journal on page 253 has, under the 'US – France' category, (minus) '-0.25'. The OA version on page 58 has, (plus) '0.25', interpreting this as an absolute difference not a numerical subtraction. The meaning is the same, though both presentations are somewhat unclear.

4) Also in Table 4, the journal on page 253 omits the values for the number of subjects for the three determinations in the second half of the table, while the values are given in the OA version (p. 58). The table is substantially rearranged in the published version, and the N values were apparently omitted. Such values are essential to evaluate the reliability of such data. Peer-reviewers would be expected to catch such an omission; unless it is present it is difficult to evaluate the validity of the conclusion. This probably should be considered a type 4 error, introduced by the publisher. (It might, however, be considered a type 3 error, on the grounds that students – who might have been expected to notice the omission - would have assumed the value acceptable. )

**Article A9:** Storesletten, K, Telmer, C L and Yaron, A. 2004. Cyclical Dynamics in Idiosyncratic Labor Market Risk. *Journal of Political Economy*. 112:695-717.  
<http://savage.wharton.upenn.edu/papers-yaron.sty1A.pdf> (Not at this URL on 1 March 2007, unless linked to from <http://savage.wharton.upenn.edu/yaron-papers.html> )

### **Three type 3 differences, two severe**

1) In the journal, Figure 1, p. 702, Cross-sectional variance of earnings, is the same as Fig. 1 in the author's version (p. 23); However, Fig 2 in the journal, p. 707, Cross-sectional variance of consumption, does not appear in the author's version at all. In the section in which it appears, 'Consumption Data' is presented in both versions by the same table, Table 3, but only the publisher's version has Figure 2. The very complicated multi-part figure is exactly analogous to Fig. 1; it cannot have been omitted by accident, because the term 'Fig. 2' is used in the author's version to refer to its own very different Fig.2, which corresponds to Fig. 3 of the published

version (see below). The entire section is rather brief in both papers, and appears to represent a condensed version of a fuller treatment elsewhere. We count this as a type 3 error, but it warrants special consideration.

2) 'Figure 3' in the journal (page 710), with the caption 'Simulated cross-sectional standard deviation', corresponds to 'Figure 2' (page 24) in the author's version.

3) Table 3 on page 708 of the journal, 'Idiosyncratic Consumption Process: Parameter Estimates' is more extensive than Table 3 in the author's version (page 22). It contains two additional sections, not present in the author's version. This is analogous to the missing graph – the Consumption part of the article is not given as fully in the author's version. We count this as a type 3 error, but it warrants special consideration.

The title page of the article has the note: 'This paper circulated previously as a component of the paper "Asset Pricing with Idiosyncratic Risk and Overlapping Generations".' The graphs referred to are not present in the author's version of that paper at <http://ideas.repec.org/p/hhs/iessp/0703.html> (accessed 5 March, 2006). It is obvious that the author's version does not exactly represent the manuscript of the paper actually published. No reasonable student could be expected to figure out the differences. This type of error was not anticipated by our classification. It would be tempting to classify this as the equivalent of a class 4 error by the author: indifference to providing an exact author's version.

**Article A10:** Berry, S, Levinsohn, J and Pakes, A. 2004. Differentiated Products Demand Systems From a Combination of Micro and Macro Data: The New Car Market. *Journal of Political Economy*. 112:68-105.

<http://www.econ.yale.edu/~steveb/microblp.pdf> (confirmed, 1 March 2007)

### **No type 3 differences**

**Article A11:** Hurst, E and Lusardi, A. 2004. Liquidity Constraints, Household Wealth, and Entrepreneurship. *Journal of Political Economy*. 112: 319-347.

[http://www.dartmouth.edu/~alusardi/final\\_entrepreneurship\\_jpe\\_sept2003.pdf](http://www.dartmouth.edu/~alusardi/final_entrepreneurship_jpe_sept2003.pdf) (Not at this URL on 1 March, but only from [http://faculty.chicagogsb.edu/erik.hurst/research/final\\_entrepreneurship\\_JPE\\_sept2003.pdf](http://faculty.chicagogsb.edu/erik.hurst/research/final_entrepreneurship_JPE_sept2003.pdf); not at Lusardi's web site, or anywhere at Dartmouth.)

### **Three type 3 differences**

1) Figure 1 in the journal, on pages 326-328, lacks the author's version's explanatory note (page 34). The same information is in the text of each version. The author's version might be more helpful, though the editors probably found it redundant, since normal journal style is not to include the information twice.

2) The journal's Figure 2 (page 331) does not have a headline, and does not have the explanatory note under the graph present in the author's version (page 35). The journal, on page 329 at the bottom, includes the phrase, '...we restrict our sample to those not owning a business in 1989.' The author's version has this information in the note below Figure 2; the journal omits the information from the Figure, but has some of the information in the note on page 330. The basic information is the same in both.

3) In Figure 3, on page 335, the journal omits the headline and the explanatory note present in the author's version (page 36).

**Article A12:** La Porta, R, Lopez-de-Silanes, F, Pop-Eleches, C and Shleifer, A. 2004. Judicial Checks and Balances. *Journal of Political Economy*. 112: 445-470.

[http://www.columbia.edu/~cp2124/papers/JudicialChecksFinal\\_JPE.pdf](http://www.columbia.edu/~cp2124/papers/JudicialChecksFinal_JPE.pdf)

(confirmed, 1 March 2007)

### **One type 3 difference**

1) The journal does not include Appendices A and B, but refers to a link on its cover page for the data used in the article. The author's version has Appendices A and B.

## **Appendix B: Type 3 differences between biochemistry article pairs**

(These articles were originally accessed in June 2006 and checked on 1 February 2007.)

**Article B1:** Roy, R S, Gopi, H N. Raghothama, S, Gilardi, R D, Karle, I L and Balaram, P. 2005. Peptide Hairpins with Strand Segments Containing  $\alpha$ - and  $\beta$ - Amino Acid Residues: Cross-Strand Aromatic Interactions of Facing Phe Residue. *Biopolymers* 80:787-799. PMID 15895435; PMCID 1380308

### **One type 3 difference**

1) A line of data in Table II was missing in the author's version, but added to the publisher's version; the nature of the chemical bond for this data point makes it clear to an experienced reader that the missing data could not be significant; however, a beginner might not have realized (note: Checked by DG). Note that for this paper in general, the editing in the publisher's version substantially improves on the carelessly prepared manuscript version.

**Article B2:** Chaney, M O, Blaine, K, Tyler A, Kuo, Y-M, Esh, C, Rahman, A, Luehrs, D C, Schmidt, A M, Stern, D, Yan, S D and Roher, A E. 2005. RAGE and amyloid beta interactions: atomic force microscopy and molecular modeling. *Biochimica et Biophysica Acta*. 1741:199-205.

PMID 15882940; PMCID 1168693948 (Document not available at PMC in 1 March 2007, but was available in June 2006.)

### **No type 3 differences**

**Article B3:** Lushnikov, A Y, Potaman, V N, Oussatcheva, E A, Sinden, R R and Lyubchenko, Y L. 2006. DNA Strand Arrangement within Sfil-DNA Complex: Atomic Force Microscopy Analysis. *Biochemistry*. 45: 152-158.

PMID 16388590; PMCID 1352315

### **Two type 3 differences**

1) Substitution of 'helices' for 'strands'

2) An added sentence in the publisher's version more clearly describing the analysis. Both changes occurred several times in the paper, but we count them as only two differences, for consistency with the social science analysis.

Note: There was extensive language editing involving type 1 and 2 differences in the publisher's version, including added words for clarification (some of which might otherwise have confused beginners), and in some cases highlighting findings more prominently, e.g., 'select between the two models' became 'ascertain which of the two models is correct'. In one case, 'and' was changed to 'or'. This extensive editing made the publisher's version more readable and clearer.

**Article B4:** Haley, J D, Smith, D E, Schwedes, J, Brennan, R, Pearce, C, Moore, C, Wang, F, Petti, F, Grosveld, F, Jane, S M, Noguchi, C T, Schechter, A N. 2003. Identification and characterization of mechanistically distinct inducers of  $\gamma$ -globin transcription. *Biochemical Pharmacology*. 66:1755-1768.

PMID 14563486: PMCID 1351252

### **No type 3 differences**

**Article B5:** Walkup, A S, and Appling, D R. 2005. Enzymatic characterization of human mitochondrial C<sub>1</sub>-tetrahydrofolate synthase. *Archives of Biochemistry and Biophysics*. 442: 196-205.

PMID 16171773; PMCID 1265403875

Document not available at PMC by January 2007, but was available in June 2006.

### **No type 3 differences**

**Article B6:** Young, W-Y, Zhao, L, Qian, Y, Wang, Q, Li, N, Greinwald, J H. Jr., Guan, M-X. 2005. Extremely low penetrance of hearing loss in four Chinese families with the mitochondrial 12S rRNA A1555G mutation. *Biochemical and Biophysical Research Communications*. 328: 1244-1251.

PMID 15708009: PM CID unknown. (No longer available on PMC, but was available in June 2006.)

### **No type 3 differences**

**Article B7:** Yakubenko, V P, Yadav, S P and Ugarova, T P. 2006. Integrin  $\alpha_D\beta_2$ , an adhesion receptor upregulated on macrophage foam cells, exhibits multiligand-binding properties. *Blood*. 107:1643-1650.

PMID 16239428, PMCID 1367263

### **One Type 3 difference**

1) Discussion section includes major rewording to make the conclusion stronger and more understandable. Beginner might not have fully understood significance of result. Publisher's version was clearer.

**Article B8:** Grens, K E, Greenwood, A K and Fernald, R D. 2005. Two visual processing pathways are targeted by gonadotropin-releasing hormone in the retina. *Brain, Behavior and Evolution*. 66:1-9.

PMID 15821344, PMCID 1167600

### **No type 3 differences**

**Article B9:** Bradford, R L, Wang, C, Zack, D J and Adler, R. 2005. Roles of cell-intrinsic and microenvironmental factors in photoreceptor cell differentiation. *Developmental Biology*. 286: 31-45.

PMID 16120439, PMCID 1351328

### **No type 3 differences**

**Article B10:** Zhao, M, Bai, H, Wang, E, Forrester, J V and McCaig, C D. 2004. Electrical stimulation directly induces pre-angiogenic responses in vascular endothelial cells by signaling through VEGF receptors. *Journal of Cell Science*. 117: 397-405 .  
PMID 14679307, PMCID 1459284

**No type 3 differences**

**Article B11:** Bogusiewicz, A, Mock, N I and Mock, D M. 2004. Instability of the biotin-protein bond in human plasma. *Analytical Biochemistry*.. 327: 156-161.  
PMCID 1464055, PMID 15051531

**No type 3 differences**

**Article B12:** Smyth, D J, Glanfield, A, McManus, D P, Blair, D, Anderson, G J and Jones, M K. 2006. Two isoforms of a divalent metal transporter (DMT1) in *Schistosoma mansoni* suggest a surface-associated pathway for iron absorption in schistosomes. *Journal of Biological Chemistry*. 281: 2242-2248.  
PMCID 1459476, PMID 16267047

**No type 3 differences**

## **References**

- (1) Goodman, D. 2004. The Criteria for Open Access. *Serials Review*, 30: 258-270. Retrieved August 1, 2006 from <http://dlist.sir.arizona.edu/798/>
- (2) Morgan, C. Re: Errors in author's versions.  
<http://www.library.yale.edu/~license/ListArchives/0607/msg00256.html> (accessed 16 Jan 2007)
- (3) PubMed Central. Author manuscripts in PMC.  
<http://www.pubmedcentral.nih.gov/about/authorms.html> (accessed 16 Jan 2007)
- (4) Banks, P. Re: Errors in author's versions.  
<http://www.library.yale.edu/~license/ListArchives/0607/msg00256.html> (accessed 16 Jan 2007)
- (5) NISO/ALPSP Working Group on Versions of Journal Articles.  
[http://www.niso.org/committees/Journal\\_versioning/JournalVer\\_comm.html](http://www.niso.org/committees/Journal_versioning/JournalVer_comm.html) (accessed 1 Aug 2006)
- (6) Suber, P. *Open Access Overview* (last revised 1 March 2006).  
<http://www.earlham.edu/~peters/fos/overview.htm> (accessed 4 Jan 2007)
- (7) PubMed Central. Author Manuscripts in PMC (2005/6).  
[http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pmc&cmd=search  
&term=author%20manuscript\[filter\]](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pmc&cmd=search&term=author%20manuscript[filter]) (accessed 22 July 2006)
- (8) Antelman, K. 2004. Do open-access articles have a greater research impact? *College & Research Libraries*, 65: 372-382. Available at <http://eprints.rclis.org/archive/00002309/> (accessed 1 August 2006)
- (9) Wates, E & Campbell, R. 2007. Author's version vs. publisher's version: an analysis of the copy-editing function. *Learned Publishing*, 20: 121-129.
- (10) Davis, P M & Connolly, M J L. 2007. Institutional Repositories: Evaluating the Reasons for Non-use of Cornell University's Installation of DSpace. *D-Lib Magazine* 13. Available at <http://www.dlib.org/dlib/march07/davis/03davis.html> (accessed 16 March 2007)

