How should editors respond to plagiarism?
COPE discussion paper

Elizabeth Wager, 26th April 2011

Summary
This paper aims to stimulate discussion about how editors should respond to plagiarism. Different types of plagiarism are described in terms of their: extent, originality of the copied material, context, referencing, intention, author seniority, and language. Journal responses to plagiarism are also described including: educating authors, contacting authors' institutions, issuing corrections, and issuing retractions. The current COPE flowcharts recommend different responses to major and minor plagiarism. Possible, more detailed, definitions of these are proposed for discussion. Decisions about when to use text-matching software are also outlined. The appendix describes other systems for classifying plagiarism and links to related documents and resources.

- **Questions for discussion**
- Should we attempt to define different types of plagiarism?
- If so, is the distinction between major and minor plagiarism useful or do we need more categories?
- What types of plagiarism should prompt journals to inform authors' institutions?
- What other sanctions should journals impose on authors for plagiarism?
- How should journals handle cases of the various forms of plagiarism in submitted and published work?

Contents

Aims and scope

This discussion paper aims to describe different forms of plagiarism to enable editors to discuss the appropriate responses to them. It does not aim to provide guidance at this stage, but we hope it may lead to the development of more detailed guidance which might supplement the COPE flowcharts in the future. It does not cover 'self-plagiarism' since we feel this is better considered separately as redundant publication.

We welcome comments from journal editors and researchers working in all fields and languages, whether or not they are COPE members.

Background

COPE (the Committee On Publication Ethics) aims to help editors and publishers of scholarly journals to handle ethical issues. It is an independent, not-for-profit organization (run from the UK as a registered charity) funded by membership fees. Over 6400 journals from a wide range of academic disciplines and world regions belong to COPE. Many major publishers have signed up their journals as members. All members are expected to follow COPE's Codes of Conduct for Editors.

The COPE Code of Conduct states that editors should 'maintain the integrity of the academic record' and 'should take all reasonable steps to ensure the quality of the material they publish'. The more aspirational Best Practice guidelines propose that editors should have 'systems in place to detect … plagiarised text.. either for routine use or when suspicions are raised'. COPE already has flowcharts to guide editors in cases of suspected plagiarism but comments from our members suggest more guidance would be welcome, especially regarding the use of text-matching tools that are now available to many journals.

1. Introduction: why do we need a discussion paper?

The COPE flowcharts (www.publicationethics.org) recognise that an editor's response to plagiarism should depend on the type and extent of the copying. They suggest different responses to 'Clear plagiarism' (described as 'unattributed use of large portions of text and/or data, presented as if they were by the plagiarist') and 'Minor copying of short phrases only' with 'no misattribution of data' (giving an example of copying 'in [the] discussion of [a] research paper from [a] non-native language speaker'). The flowcharts also distinguish plagiarism (ie copying from others) from redundancy or 'self-plagiarism' (ie copying from one's own work). The flowcharts also suggest that the editor's response might vary according to the seniority of the author (with editors simply writing an educational letter to very junior researchers but considering informing the institution of more senior authors) as well as whether the authors are writing in their native language.

The availability of powerful text-matching software and systems such as CrossCheck (which enables editors to compare text to a large database of published academic literature as well as against material freely available on the internet – see www.crossref.org/crosscheck.html) has made the detection of text duplication both easier and more sensitive. However, editors now have to decide when to use such systems and how to interpret their output. Journals that routinely screen all submissions for matching text (rather than checking only papers they plan to accept) also need to consider what to do when they find matched text in a paper they plan to reject. Therefore, although the dictionary definitions of plagiarism may be uncontroversial,

editors need a more detailed taxonomy to distinguish the different forms of plagiarism so they can decide on an appropriate, proportionate and consistent response.

2. Types of plagiarism

Any original creation may be plagiarised. Although most discussions focus on text (and this type of copying is the easiest to detect using software) it is important to recognise that ideas, images, creative works (eg musical compositions or choreographies), and data can also be plagiarised.

The following factors may be helpful in distinguishing types of plagiarism (see Table 1):

- extent
- originality of copied material
- position / context
- referencing / attribution
- intention
- author seniority
- language

Table 1: Features of different types of plagiarism

Feature	**Least severe type**				**Most severe type**
Extent	A few words	A few sentences	Whole paragraph	Several paragraphs	Whole paper
Originality of copied material	Widely-used phrase / idea		Phrase / idea used by a small number of authors		Original phrase / idea
Position / context / type of material	Standard method		Describing another worker's findings		Data / findings
Referencing / attribution	Source fully and clearly referenced		Source partially / inaccurately referenced		Unreferenced
Intention	No intention to deceive				Intention to deceive

2.1 Extent

The most blatant forms of plagiarism involve the copying of entire papers or chapters which are republished as the work of the plagiarist. Such cases usually involve not only plagiarism but also breach of copyright. Whole articles or chapters may also be plagiarised in translation. The COPE retraction guidelines recommend that such articles should be retracted and the flowcharts on plagiarism suggest that editors should consider contacting the author's institution in such cases. However, the COPE retraction guidelines state that 'if only a small section of an article (e.g. a few sentences in the discussion) is plagiarised, editors should consider whether readers (and the plagiarised author) would be best served by a correction (which could note the fact that text was used without appropriate acknowledgement) rather than retracting the entire article which may contain sound, original data in other parts'.

Scholarly works often summarize the work of other researchers. It may be difficult to draw a line between legitimate (and accurate) representation of other studies and copying original material. Researchers may also feel that little harm is done if they use similar language to another publication so long as the source is properly cited. If the original authors summarized their findings clearly and succinctly it could be argued that little is gained by forcing other authors to paraphrase this. However, others will argue that any verbatim copying should be indicated by using quotation marks, otherwise they would consider it to be plagiarism.

Most text-matching software detects strings of several words, since duplication of just a few words can occur by chance. However, academic papers and reports may contain technical language that involves standard phrases that are longer than the strings used by software. For example, a Google search for the phrase 'smokers with chronic obstructive pulmonary disease' produces >58,000 results, suggesting that this is a widely used phrase, but such a 6-word string may also trigger a match on text-matching software. Therefore extent alone cannot be taken as a benchmark.

2.2 Originality of copied material

Originality needs to be considered in conjunction with extent. The example given above indicates the difference between a standard phrase (such as 'smokers with chronic obstructive pulmonary disease') and original usage (such as Winston Churchill's 'the end of the beginning' or Shakespeare's 'the winter of our discontent' – both of which contain fewer than 6 words and would therefore probably not be detected by text-matching software yet are usually considered sufficiently original to be attributed to the original writer). While publishers of poems and song lyrics tend to guard their copyright fiercely, and permission is required to quote even a single line, technical publications may contain descriptions of standard techniques which will tend to be described in similar or identical ways. Therefore the originality of the copied material should be considered as well as the extent.

Table 2: Examples of language of low originality used in reports of medical research

	Hits for exact phrase (January 2011)	
Phrase	**Google**	**Google Scholar***
P<0.05 was considered statistically significant	588,000	70,600
computer-generated random number list	5120	354
double-blind, double-dummy, placebo-controlled	56,800	882
randomised in a 1:1 ratio	8510	1020
numbered, opaque, sealed envelopes	12,200	912
performed according to the Declaration of Helsinki	410,000	1860

*Google Scholar searches for academic publications only

2.3 Position / context

Certain sections of research reports may be more likely to include non-original material. In particular, the Methods section may describe widely-used techniques. Standardized descriptions of public data sources, proprietary techniques, questionnaires or equipment may even be regarded as good practice. For example, analysis of the UK General Practice Research Database (www.gprd.com) has resulted in over 750 publications. All these publications probably include a description of the database and these are likely to use similar language. Similarly, it may be better if the original description of an assay provided by a company or the supplying laboratory is copied rather than reworded by each user, since the original wording may be the most accurate. Therefore editors may view text similarity in Methods sections differently from that in other parts of a paper.

The editor of a mathematics journal has noted 'statements of the mean value theorem from calculus book to calculus book are virtually identical; there's really only way to state Schur's Lemma. Probably, automated software would detect these instances as plagiarism. They're not, of course. Sometimes, there's really only one way to define something or give the "usual examples".' [Lance Small, personal communication].

The type of publication may also affect judgements about the acceptability of text similarity. While research reports may describe standard methods, editorials may be expected to represent the author's opinion and original deliberations on a topic and it would therefore be inappropriate to use the same words as another author except in direct and properly attributed quotations. Similarly, review articles, and the discussion sections of research papers, are expected to provide an original synthesis of, and commentary on, previously published work. Therefore, apart from quotations, the words may be expected to be the author's own.

However, editors may also consider the consequences of the copying and its potential to mislead readers. In this respect, copying a few sentences from the Discussion section of another researcher's paper may be considered less harmful, and less deceitful, than the theft of data (which may constitute not only plagiarism but also data fabrication since the work was not done by the copier). Thus, if an editor finds a paper that appears to describe legitimate, original research, but includes some sentences taken from the Discussion of another author's paper on a related topic, the editor may simply ask the author to indicate that these are direct quotations, or to paraphrase the copied text, before publication. If the copying is discovered after publication, the editor may suggest that it can be rectified by a correction rather than a retraction and may not feel that the author's institution should be informed.

When using software to detect text similarity, editors should not forget that reference sections will contain large amounts of copied text in the form of titles of cited articles. Some software systems, such as iThenticate / CrossRef, allow these sections of the paper to be excluded from the search, together with any text enclosed in quotation marks.

If systematic reviews or databases are regularly updated over many years, the original authors may retire and be replaced by others. An updated review or database will, naturally, contain large sections from the previous versions and this may appear to be plagiarism if the authors have changed (and automatic systems will not recognise acknowledgements to previous versions).

2.4 Referencing / attribution

Academic publications are expected to reference other works and may also quote from them. Inexperienced or poorly trained authors may mistakenly believe that so long as another work has been cited, parts of it can be reproduced in their own work. While copying parts of cited work is probably not intended to deceive the reader in the same way as copying unattributed material, the practice is generally considered to be poor scholarship and inappropriate for an academic journal. Editors may have a role in educating authors if they discover this type of copying, especially if it is detected before publication.

2.5 Intention

Intention to deceive is often considered a factor distinguishing misconduct from careless work or honest error. However, it is usually impossible to prove intent and therefore may be less useful in practice than in theory. Extreme forms of plagiarism, such as copying an entire paper and submitting it under a different author's name to another journal can only be

deliberate. Editors must use their own judgement to determine whether authors' explanations for less extreme forms of copying are plausible or could have occurred through honest error. When confronted with identical text, authors may counter with explanations such as having a photographic memory or inadvertently copying notes or preliminary material into a publication. When a senior researcher at Stanford University was found to have incorporated large chunks of text from a well-known textbook into a chapter he had prepared for another book, he told an inquiry that 'when he cut-and pasted the material into his manuscript, he added handwritten notations detailing where the text came from. These notations were supposed to have been printed in the body of his chapter…..'. Nevertheless he was found guilty of 'grossly negligent scholarship' and resigned as chairman of the Department of Medicine (Science 1984;224:35-7).

Authors who admit intentional copying may nevertheless insist that this is acceptable in their discipline or culture and that, rather than representing academic theft or laziness, it is, in fact a form of flattery or 'homage' to the original author. They may also suggest that quotation marks are unnecessary because specialist readers (for whom they are writing) will immediately recognise the quotations and be aware of their source.

2.6 Author seniority

Since editors may believe that some forms of plagiarism result from poor mentorship or supervision rather than intentional misconduct, their response may vary according to the seniority of the authors involved. Editors may apply different sanctions to junior authors who they believe genuinely did not know they were doing something inappropriate from those applied to experienced researchers who are expected to know better. Thus, an editor may respond to the copying of a paragraph from a cited paper by asking a junior author to paraphrase (if detected before publication) or issue a correction (if detected after publication). However, for a similar degree of copying by a senior author, the same editor might reject or retract a submission and consider informing the author's institution.

Informing an author's institution is generally considered to be a relatively serious action to take, since it may have serious consequences for the researcher concerned. Editors therefore tend to be reluctant to inform institutions except in serious cases of misconduct and when they feel they have well-founded suspicions of wrong-doing. However, if contacting an institution is viewed, not as a potential punishment for the author, but as an attempt to engage the institution in dialogue and work together to prevent future problems, one might argue that editors should contact institutions more often and definitely in cases where they feel junior researchers have received inadequate training or guidance, since this is something the institution may be able to remedy.

If an editor detects copying in a manuscript that is going to be rejected, then contacting a head of department or dean might prevent the authors from simply submitting the manuscript, unchanged, to another journal.

2.7 Language

Text matching software will only detect text copying in the same language. However, republication of an unattributed translation of another person's work is also plagiarism, although it is harder to detect and may be harder to prove unless extensive.

Just as editors' responses may depend on the authors' seniority, they may also depend on whether authors are writing in their native language since editors recognise the difficulties

that non-native writers face in correctly paraphrasing another author's work. In some cases, researchers may actually have been encouraged, when learning a language, to adapt sentences and 'borrow' structures from published works. This may result in so-called 'patch' (or 'patchwork') writing. This form of copying will only be detected by sensitive text-matching systems and those that employ a degree of 'fuzzy' matching, since authors are likely to have changed some words in adapting the sentence for their own use. Authors who use this technique usually copy from a wide range of sources, often with individual sentences coming from different publications. This may result in a high total similarity 'score' for the article from an anti-plagiarism detection system such as CrossCheck, but the matched text will be found to come from multiple sources, and each copied section will be short (with few or no substantial chunks of copied text). However, few, if any of the sources of the copied text are likely to be cited in the publication, since they may be on unrelated topics.

Some editors may see little harm in authors who describe their own methods and findings accurately, but using sentence structures taken from other publications. Others may regard this as a sign of poor scholarship or a form of minor plagiarism. The acceptability of 'patch' writing probably depends on the originality of the writing being copied. While it may be entirely unacceptable for works of creative fiction, it may be considered acceptable when describing widely-used methods which, as already shown, may use a degree of standardized text. If the copied structures are clear and grammatically correct, some editors may even feel that this method of writing will benefit readers and journals, since methods will be accurately described and the manuscript will require less copy editing to correct grammatical mistakes.

However, others may have concerns that authors will be tempted to copy inappropriate phrases that do not correctly describe their own research, especially if they do not completely understand the phrases being copied. To misrepresent research methods (for example by stating that a study was prospective or randomized when, in fact, it was not) is generally considered a serious form of misconduct. Editors may therefore be concerned that, if they tolerate 'patch' writing, such misrepresentation may be a consequence.

3. Detecting and responding to plagiarism

This paper aims to stimulate discussion among editors and researchers (ie authors) to determine whether there is consensus about which responses are most appropriate for the various forms of plagiarism, or, if that is not possible, at least determine if any sanctions are inappropriate.

The original COPE guidelines on good publication practice (published in 1999) noted that 'plagiarism ranges from the unreferenced use of others' published ideas … to submission under "new" authorship of a complete paper, sometimes in a different language.' However, these guidelines did not describe what editors should do if they encountered these different forms. The guidelines did offer general guidance on the sanctions that editors might take against authors (see text box).

Possible responses to misconduct
(from COPE Good Publication Practice, 1999)

The following [sanctions] are ranked in approximate order of severity:
(1) A letter of explanation (and education) to the authors, where there appears to be a genuine misunderstanding of principles.
(2) A letter of reprimand and warning as to future conduct.
(3) A formal letter to the relevant head of institution or funding body.

(4) Publication of a notice of redundant publication or plagiarism.
(5) An editorial giving full details of the misconduct.
(6) Refusal to accept future submissions from the individual, unit, or institution responsible for the misconduct, for a stated period.
(7) Formal withdrawal or retraction of the paper from the scientific literature, informing other editors and the indexing authorities.
(8) Reporting the case to the General Medical Council, or other such authority or organisation which can investigate and act with due process.

The COPE flowcharts on plagiarism (published in 2006) recommend different responses for 'clear plagiarism' and 'minor copying' but provide only rather general indications of how editors might distinguish these two phenomena.

4. Screening for plagiarism

The availability of powerful tools such as CrossCheck makes it possible to screen submissions for matching text and some journals are now doing this routinely. However, screening carries costs (in the form of charges for using the tools, and in terms of editorial time) and therefore editors and publishers need to decide the best ways of employing it. The options include:

- screening all manuscripts on receipt
- screening manuscripts that are sent out for external peer review
- screening manuscripts that are provisionally accepted
- screening a random sample of manuscripts
- using the software only in cases when plagiarism is suspected.

We know that some COPE members started by screening only accepted manuscripts but switched to screening all submissions because of the frequency of problems they discovered. The Editor of *Anesthesia & Analgesia* notes in an editorial that 'I have screened every submitted manuscript for many months. Approximately 1 of every 10 submissions has had unacceptable amounts of text taken verbatim and without attribution from another source.'

5. Defining plagiarism

Editors also need to decide how to interpret and respond to findings of text similarity. It is important that authors receive fair and consistent treatment from journals but devising a detailed policy on responses to plagiarism is difficult given the many forms that plagiarism can take. Because text-matching software has only become available relatively recently, cases of plagiarism are likely to be uncovered in back issues of the journal. Editors therefore need a clear policy for responding to plagiarism in material published recently and in the past.

Clear-cut cases of serious plagiarism (eg whole articles or large sections of text) may warrant retractions. Since the general concept of plagiarism is not new, and large-scale plagiarism has been identified as a serious form of misconduct for decades, most editors would agree that this is the correct course of action. However, identification of 'patch writing' or 'micro-plagiarism' has only become possible with the availability of specialised software. Some editors may therefore feel uncomfortable about applying sanctions to authors retrospectively. One solution to this problem would be to announce an amnesty for older publications (ie an agreement that the journal will not take action if minor plagiarism is found in previous issues) but warning authors that text similarity in future submissions will not be tolerated.

To devise workable policies for both submitted and published articles, editors will need to consider the thresholds for deciding when to:

- educate authors and ask them to rewrite copied text
- reject an article
- issue a correction (for a published article)
- issue a retraction (for a published article)
- inform an author's institution

Returning to the COPE flowcharts, perhaps we need to provide more guidance about how to distinguish major from minor plagiarism. One possibility would be to produce definitions based on the characteristics described above. **(Please note, the following definitions are simply for discussion, they do <u>not</u> represent official COPE guidance! We particularly welcome comments on how to define significant sections of text. In these examples, we have used the number of words, since both sentences and paragraphs can vary considerably in length. However, we recognise the problems in proposing arbitrary limits and would welcome other suggestions.)**

For example / for discussion
<u>Major plagiarism</u> could be defined as:
Any case involving

- unattributed copying of another person's data / findings, *or*
- resubmission of an entire publication under another author's name (either in the original language or in translation), *or*
- verbatim copying of >100 words of original material in the absence of any citation to the source material, *or*
- unattributed use of original, published academic work, such as the structure, argument or hypothesis/idea of another person or group where this is a major part of the new publication and there is evidence that it was not developed independently.

<u>Minor plagiarism</u> could be defined as:

- verbatim copying of <100 words without indicating that these are a direct quotation from an original work (whether or not the source is cited), unless the text is accepted as widely used or standardized (eg the description of a standard technique)
- close copying (not quite verbatim, but changed only slightly from the original) of significant sections (eg > 100 words) from another work (whether or not that work is cited)

<u>Use of images without acknowledgement of the source</u> could be defined as:

- republication of an image (photograph, diagram, drawing, etc.) generated by another person without acknowledging the source

Journal responses could then be matched to these, for example:
(again, these proposals are for discussion):

Minor plagiarism in submitted article – write to author and request reworking or (if article is being rejected) point out that minor plagiarism has been detected and advising the authors that this should be corrected before resubmission

Minor plagiarism in published article – contact author and discuss findings, issue a correction and apology

Major plagiarism in submitted article – present findings to all authors and ask them to respond; ask the authors if all or only some of them are responsible for the plagiarised sections, decide if any authors were unaware of the plagiarism and, if so, whether they are in any way responsible for the behaviour of the other authors (eg in a supervisory capacity); explain that plagiarism is unacceptable and that you plan to inform their institution; contact the institutions of authors you consider were directly involved with, or should take responsibility for, the plagiarism

Major plagiarism in published article – as for submitted article, then retract article

Use of images without acknowledgement of the source – if the image contains data from another person's research (eg a graph), and this is shown as if it were the work of the copyist, this should be treated as data copying (ie major plagiarism). For images that do not contain original data (eg diagrams showing processes, maps, illustrative photographs) the author of a submitted paper should be told to seek permission for republication from the copyright holder, remove images for which permission is not granted and insert appropriate acknowledgements for images for which permission has been granted; if such images have already been republished, the editor should contact the author and issue a correction giving the appropriate acknowledgements.

6. Next steps

We hope this paper will stimulate discussion. We encourage journal editors and publishers to let COPE know if they are developing or revising their policies on plagiarism, and, if so, what they have decided. We also encourage comments, especially on the suggested possible definitions and responses, from editors and publishers (whether or not they are COPE members), and from researchers / authors and academic institutions. If we feel there is sufficient agreement on what constitutes good practice, we will review the COPE flowcharts or publish further guidance.

Please send comments to
Natalie Ridgeway, COPE Operations Manager
cope_opsmanager@publicationethics.org

7. Other relevant documents and plagiarism classifications

Growing use of text matching software has led several editors to review their policies on plagiarism. We would welcome further examples, especially from outside the biomedical field.

Anesthesia & Analgesia (Shafer SL, in press, March 2011)
The journal's editor, Steven Shafer, proposes categorising plagiarism into four categories and also discusses self-plagiarism (which he notes is an oxymoron because it 'implies that an author has stolen from himself').

Category	**Definition**	**Action:** submitted paper	**Action:** published paper
Intellectual theft	Deliberate copying of large blocks of text without attribution	Reject paper Inform author's institution Impose sanctions (ban author)	Inform author's institution Retract paper Impose sanctions (ban author)
Intellectual sloth	Copying of 'generic' text, eg a description of a standard technique, without clear attribution	Either reject paper or Instruct authors to rewrite plagiarized text	Retract paper
Plagiarism for scientific English	Copying of verbatim text often from multiple sources	Instruct authors to rewrite plagiarized text	Retract paper
Technical plagiarism	Use of verbatim text without identifying it as a direct quotation but citing the source	Instruct authors to credit verbatim text / identify direct quotations properly	Retract paper

Association of Computing Machinery statement on plagiarism
A 5-page statement setting out the Association's policies on plagiarism including the penalties that may be imposed on those who have committed plagiarism
http://www.acm.org/publications/policies/plagiarism_policy

European Science Editing 2010;**36**:62-66
Mary Ellen Kerans and Marije de Jager, who are experienced translators and technical editors provide helpful definitions of terms including copy-paste writing, micro-plagiarism, and patch (or mosaic) writing. They point out that patch writing is unlikely to be detected by peer reviewers or readers, although it may be apparent to authors' editors or copy editors. They discuss strategies for dealing with different types of plagiarism before publication and warn against the 'uncritical use of detection software'.

Journal of Second Language Writing 2003;**12**:317-45
Diane Pecorari reports on attitudes to plagiarism in 17 postgraduate students and their supervisors in a European university. She concludes that although student writing may contain elements 'which could be described as plagiarism' there was no intention to plagiarize.

The Lancet 2011;**377**:281-2
Announcement of the journal's policy on plagiarism and text recyling and intention to use CrossCheck to screen certain categories of papers.

Learned Publishing 2010;**23**:9-14
Helen Zhang describes how the *Journal of Zhejiang University – Science* used CrossCheck to screen over 600 submissions before they were sent for external review and again just before publication and discovered that 23% contained unacceptable amounts of copied text.

Office for Research Integrity guide, http://ori.hhs.gov/education/products/plagiarism/
Miguel Roig offers guidance for writers on avoiding plagiarism 'and other questionable writing practices' in this 63-page document. He quotes a definition of plagiarism from the American Association of University Professors as: 'taking over the ideas, methods, or written words of another, without acknowledgment and with the intention that they be taken as the work of the deceiver' and describes several different forms of plagiarism.

Nature Publishing Group
Instructions for authors contains a page with links to various articles on plagiarism
http://www.nature.com/authors/policies/plagiarism.html

www.plagiarism.org
A website run by iThenticate, the company which produces the text matching software used in Turnitin and CrossCheck. Provides examples of 11 different kinds of plagiarism, six in which sources are not cited and five in which sources are cited.

www.plagiarismadvice.org
A website run by iParadigms Europe Ltd which supplies Turnitin and iThenticate and runs a biennial International Plagiarism Conference. Includes presentations from the conferences and a 'plagiarism reference tariff' for the application of penalties for plagiarism by students in higher education.

System 2008;**36**:337-52
Qing Gu and Jane Brooks describe sociocultural and psychological aspects of plagiarism from a study of evolving perceptions of plagiarism among 10 Chinese postgraduate students and their English tutors.