

Responding to misconduct (and preventing it from occurring): the role of journals and institutions

Elizabeth Wager PhD

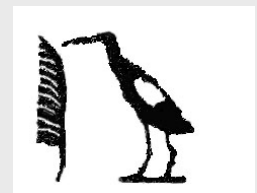
Publications Consultant, Sideview

Editor-in-Chief, *Research Integrity & Peer Review*

Visiting Professor, University of Split, Croatia

Former Chair, Committee on Publication Ethics (COPE)

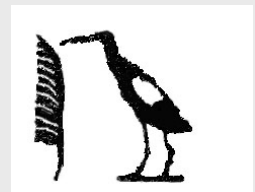
liz@sideview.demon.co.uk



How common is misconduct?

- Systematic review (screened 3207 papers)
- Meta-analysis (18 studies)
 - surveys of fabrication or falsification
 - *NOT* plagiarism
- **2%** admitted misconduct themselves
(95% CI 0.9-4.5)
- **14%** aware of misconduct by others
(95% CI 9.9-19.7)

Fanelli *PLoS One* 2009;4(5):e5738



How common is plagiarism?

Evidence from Croatia

At universities:

- Essays by 198 medical students

No plagiarism 17 (9%)

<10% copied 51 (26%)

>10% copied 130 (65%)

- Strict warnings to students had no effect on plagiarism rate

Bilić-Zulle *Croat Med J* 2005;**46**:126

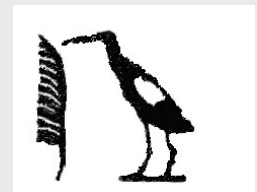
In journals:

- *Croatian Medical Journal* (2009-10)

- Screened 754 manuscripts

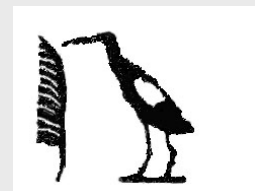
- 11% (85) contained >10% plagiarized text

Baždarić et al *Sci Eng Ethics* 2012;18:223



Does peer review detect misconduct?

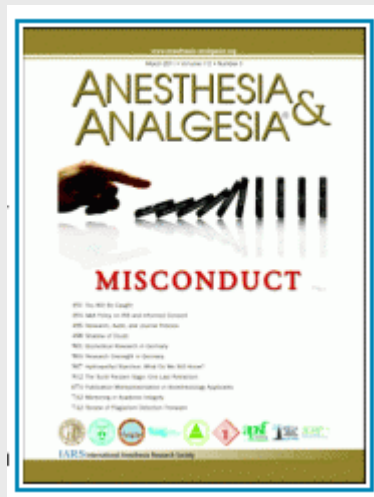
- Obviously not in all cases
- Reviewers sometimes spot:
 - plagiarism (*especially of own work*)
 - redundant publication (*from checking refs*)
 - multiple submission (*from seeing same paper*)
 - ?fabricated data *probably very rarely*





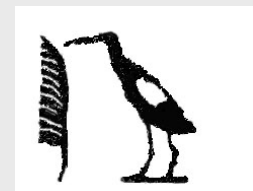
Schön's retracted papers

- 8 in *Science* (published 2000-1)
- 6 in *Physics Review* (4 from 2001)
- 7 in *Nature* (published 1999-2001)



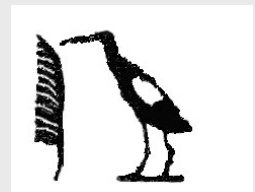
Yoshitaka Fujii

- 183 retractions in anaesthesia journals



Tools for detecting misconduct

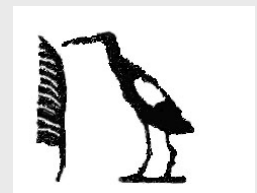
- Anti-plagiarism software
(eg CrossCheck, Turnitin)
- Screening images (PhotoShop)
- Data review (digit preference)
- Replication (for basic research)





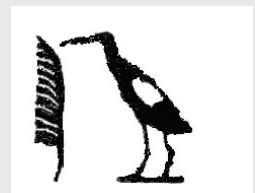
CrossCheck

- Based on Turnitin software
- Compares text against publishers' d-base
- D-base run by CrossRef (doi system)
- D-base currently contains 30 million papers
- Shows % concordance + source
- Can exclude “quotes” and references



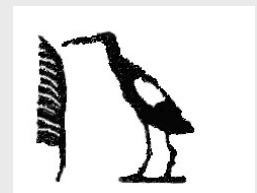
What's the place of text-matching software?

- Valuable tool for screening
- Can be used in conjunction with training (e.g. paraphrasing practice)
- Valuable tool for assessing possible misconduct
- BUT %-match must be interpreted carefully



Who's using text matching software?

- 117 million submissions to Turnitin
- >1.25 million instructors
- Many publishers using CrossCheck
- Elsevier now screening ALL submissions



Anglia Ruskin
Good practice with GradeMark
What's New
Help
Paper 2 of 3

Originality
GradeMark
PeerMark

Julius Caesar essay
BY JAMIE GREEN

turnitin
SIMILARITY
68%
INDEX
GRADE
--
OUT OF 0

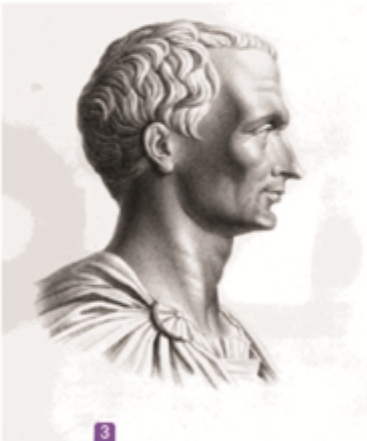
Match Overview

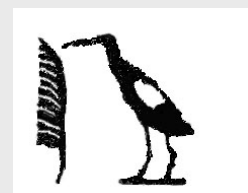
1	www.sparknotes.com Internet source	30%
2	www.overdrive.com Internet source	14%
3	Acumen PI- Title: How ... Publication	14%
4	en.wikipedia.org Internet source	10%

3
Consider how portents, omens and dreams add to the dramatic tension in Shakespeare's Julius Caesar.

4
The Tragedy of Julius Caesar is a tragedy by William Shakespeare, believed to have been written in 1599, and portrays the conspiracy against the Roman dictator Julius Caesar in 44BC, his assassination and the defeat of the conspirators at the Battle of Philippi. It is one of several Roman plays that Shakespeare wrote, based on true events from Roman history. Although the title of the play is Julius Caesar, Caesar is not the central character in its action he appears in only three scenes, and is killed at the beginning of the third act. The protagonist of the play is arguably Marcus Brutus.

3
Since Shakespeare lived the Elizabethan period he would have had the conception of the divine order of the universe being mystical. For example, the Elizabethans would see storms as a warning to some sort of disaster or calamity to be visited upon men, by God. Also many Elizabethans were superstitious about many things and saw the most commonplace events as portents and omens of events to come. Shakespeare exploits this Elizabethan sensibility to great dramatic effect in Julius Caesar.



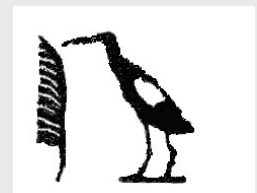


Challenges with software

- Only checks text – NOT data, tables, figures
- Works in single language only
- Cannot spot translated plagiarism
- Cannot identify plagiarism of ideas

For academic publishing:

- Need to check authors to distinguish plagiarism from redundant publication ('self-plagiarism')



Clear plagiarism from one source

iThenticate Sample Article BY TEST AUTHOR

Quotes Included 75%
Bibliography Included 75%
SIMILAR

Match Overview

1	Internet 4091 words crawled on 04-Nov-2008 www.carleton.ca	59%
2	CrossCheck 262 words Gitte Lindgaard, "Utility and Experience in the Evolution of Usability", Human-Computer Interaction Series, 2008	4%
3	Internet 175 words crawled on 02-Feb-2010 www.apo.org.au	3%
4	Internet 127 words crawled on 30-Apr-2010 eprints.qut.edu.au	2%
5	Internet 101 words crawled on 03-May-2013 diepienaars.co.za	1%
6	Internet 75 words crawled on 16-Aug-2011 australiaart.wordpress.com	1%
7	Internet 75 words crawled on 24-Sep-2010 courses.coe.uh.edu	1%
8	CrossCheck 41 words Gitte Lindgaard, "Integrating aesthetics within an evolutionary and psychological framework", Theoretical Issues in	1%

interpreted the meaning of the stimuli (LeDoux 1994). Thus the amygdala does not depend entirely on signals from the neocortex as originally believed, and Zajonc's early findings are beginning to converge with more recent theoretical explanations of human emotion as well as with empirical evidence.

The mere exposure effect begins to wane once the stimulus exposure time exceeds 50msec, when the organism begins to take more detailed information into consideration. Therefore, if visual appeal is appraised within that window of 50 msec, as indeed a series of experiments in our lab showed clearly to be the case (Fernandes, Lindgaard & Dillon 2003; Lindgaard et al. 2006), the judgment involves the amygdala over which the neocortex has no control. The organism's response can thus truly be said to be visceral (Norman 2004). This has important implications for web design and budget resource allocation, because the value of textual information is likely to be assessed in terms of the site's immediate visual appeal. So, an e-

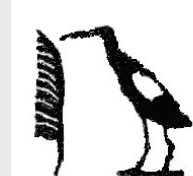
©Australian Journal of Emerging Technologies and Society 2007 5
<http://www.swin.edu.au/ajets>

AJETS Vol. 5, No. 1, 2007, pp: 1-14

commerce web site representing a business that is in competition with numerous others on the Internet and that fails to meet users' expectations both in terms of aesthetics and in terms of its informative content is unlikely to be successful at converting browsers to customers even if its quality of products or services is superior to its competitors. Either people will instantly click on to the next site, or they will interpret even the slightest usability flaw negatively to confirm their initial emotional impression. By contrast, a visually appealing site will be forgiven

PAGE: 9 OF 25

Text-Only Report



Low total but some long copied passages

06-Feb-2014 10:21AM 5475 words • 24 matches • 20 sources

iThenticate® In Google We Trust BY G BILDER

Quotes Included: 13%
Bibliography Included: 0%

Volume 12, Issue 1, Winter 2006
DOI: <http://dx.doi.org/10.3998/3336451.0009.101>
Permissions

Abstract

Trust, authority, and reputation are central to scholarly publishing, but the trust model of the Internet is almost antithetical to the trust model of academia. Publishers have been so preoccupied with the brute mechanics of moving content to the online world that they have virtually ignored the challenge that the Internet trust model poses to the scholarly publisher. Publishers can learn much about approaches to handling Internet trust from the actions of major online players outside the publishing industry. Publishers should also benefit from watching the trust models that are being experimented with in the nascent realm of social software applications. Publishers once led the way in establishing the apparatus of trust during the transition from manuscript to print culture in early modern Europe. Ultimately, publishers should again take the lead in helping to establish new mechanisms of trust in what could reasonably be described as "the early modern Internet."

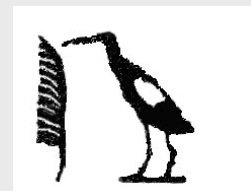
Every day, Internet users are pelted with spam, hoaxes, urban legends, and scams - in other words, untrustworthy data. The Internet is largely without any infrastructure to help users identify authoritative and trustworthy content. Indeed, the history of the Internet is littered with examples of how technologists have underestimated the crucial role that social trust and authority play in communication.

Authority is the *sine qua non* of academic publishers and librarians, and the trust model of the Internet is almost completely antithetical to the model of scholarly trust that has evolved following the development of academic publishing in early modern

Match Overview

1	Internet 164 words crawled on 13-Sep-2012 www.citeulike.org	3%
2	Internet 103 words crawled on 01-Jul-2006 cites.boisestate.edu	2%
3	Internet 93 words crawled on 01-Sep-2011 www.crossref.org	2%
4	Internet 86 words crawled on 27-Jul-2007 www.breakawayrepublic.com	2%
5	Internet 59 words crawled on 10-Sep-2010 strathprints.strath.ac.uk	1%
6	Publications 47 words "How has the flora and fauna of the rainforest adapted ... o their environment?", <i>gcse/geography/physical geograp</i>	1%
7	Internet 42 words crawled on 25-Oct-2002 www.eff.org	1%
8	Internet 25 words crawled on 13-Sep-2011 blogs.library.ualberta.ca	<1%

PAGE: 1 OF 13 Text-Only Report



Overlap from legitimate source (citulike)

06-Feb-2014 10:21AM 5475 words • 24 matches • 20 sources FAQ

iThenticate® In Google We Trust
BY G BILDER

Quotes Included 13%
Bibliography Included SIMILAR

of people your colleagues trusted. Outfoxed provides a tantalizing glimpse of what the Internet might feel like if the restrictions of local/global and horizontal/vertical trust can be transcended.

Conclusion

¹¹ In Jorge Luis Borges's short story, "Tlön, Uqbar, Orbis Tertius", Borges is impressed by a friend's obscure quote and asks him to cite it. The friend, Adolfo Bioy Casares, obliges, saying that he encountered it while reading an entry on the country "Uqbar" in the *Anglo-American Cyclopdia*. When they go and check Borge's copy of the same edition of the same encyclopedia, not only do they fail to find the quote, but they fail to find the entry on Uqbar. When the embarrassed Bioy Casares goes home, he looks up the entry in his copy of the encyclopedia and finds that it is indeed there. It seems that the two copies of the encyclopedia are identical, except for the addition of the one entry on Uqbar in Bioy Casares's copy. Investigation reveals that the country Uqbar never existed, and that the fictional encyclopedia entry was the result of a vast conspiracy of intellectuals to test Berkeleian idealism by imaging a fictional world in the hopes of actually creating it.

A mere summary of the story does not do justice to the foreboding atmosphere that Borges creates. Modifying a print reference work could not have been a trivial task - it would have required the collusion of countless influential people who were traditionally the guardians of authority and credibility: editors, composers, printers etc. That the encyclopedia entry was fabricated and inserted into some copies of an otherwise legitimate reference work is shocking, and serves to underscore the power, influence, and cunning of the conspirators.

Today, the much-vaunted Wikipedia makes a virtue of the fact that anyone can add or

Match Overview

1	Internet 164 words crawled on 13-Sep-2012 www.citulike.org	3%
2	Internet 103 words crawled on 01-Sep-2006 cites.boisestate.edu	2%
3	Internet 93 words crawled on 01-Sep-2011 www.crossref.org	2%
4	Internet 86 words crawled on 27-Jul-2007 www.breakawayrepublic.com	2%
5	Internet 59 words crawled on 10-Sep-2010 strathprints.strath.ac.uk	1%
6	Publications 47 words "How has the flora and fauna of the rainforest adapted ... o their environment?", gosea/geography/physical geograp	1%
7	Internet 42 words crawled on 25-Oct-2002 www.eff.org	1%
8	Internet 25 words crawled on 13-Sep-2011 blogs.library.ualberta.ca	<1%

PAGE 11 OF 13 Text-Only Report



Problem: how to define plagiarism?

COPE Discussion Document:

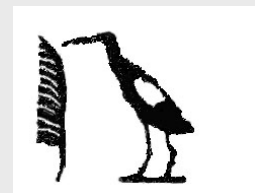
<http://www.publicationethics.org/resources/discussion-documents>

Feature	Least severe				Most severe
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 few authors	Original phrase / idea	
Position / type of material	Standard method		Describing another's work	Data / findings	
Referencing / attribution	Source fully and clearly referenced		Source partially / inaccurately ref'd	Unreferenced	
Intention	No intention to deceive			Intention to deceive	

Originality is important

	Hits for exact phrase	
	Google	Google Scholar
P<0.05 was considered statistically significant	588,00	70,600
Performed according to the Declaration of Helsinki	410,000	1860
Double-blind, double-dummy, placebo-controlled	56,800	882
Randomised in a 1:1 ratio	8510	1020

cf: Shakespeare 'the winter of our discontent'
The Beatles 'It's been a hard day's night'

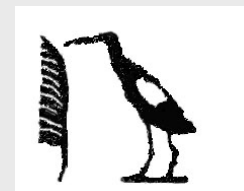


Turnitin is effective for training

Study of 40 engineering students in Turkey

	CONTROL (No training) (n=23)	Turnitin™ training (n=17)
% plagiarism Baseline		8-22%
2 nd essay	2-22%	0-12%

Köse & Arikan *Cont Online Lang Edu J* 2011;1:122

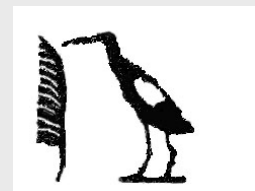


Training can reduce student plagiarism

- 4 year study of postgrad medical student assignments in UK university
- Introduced 40-minute interactive plagiarism seminar
- Assignments checked using TurnitinTM
- % text match >3SD from mean were examined
- Two staff members confirmed plagiarism

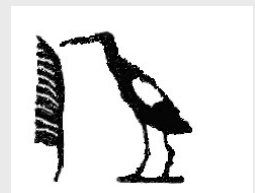
Years without training	Years with training
7/470 (1.5%)	2/615 (0.3%)
	p=0.013

Marshall et al *Med Teacher* 2011;**33**:e375



What should training cover?

- What is plagiarism?
- Awareness of university policy (+ sanctions)
- Good citation practices
- Good writing / paraphrasing practices
- Copyright ... images, etc.



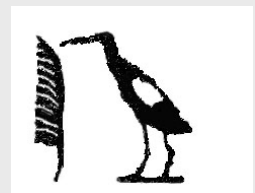
Misconduct

Research misconduct

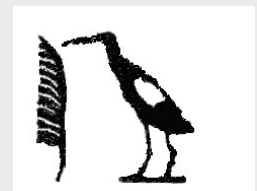
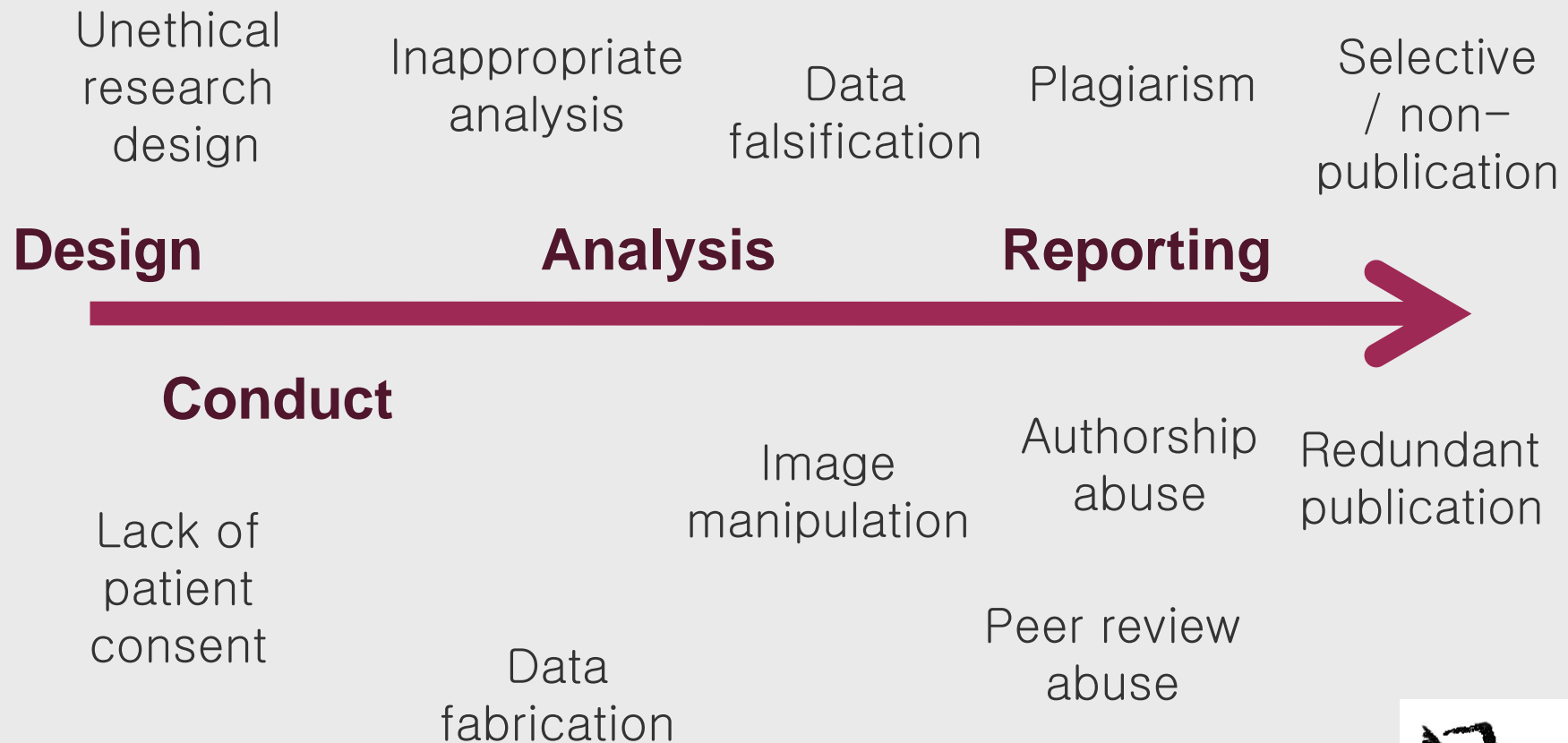
- Fabrication
- Falsification
- Unethical research

Publication misconduct

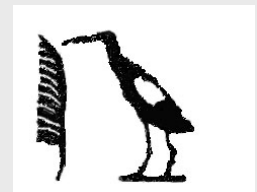
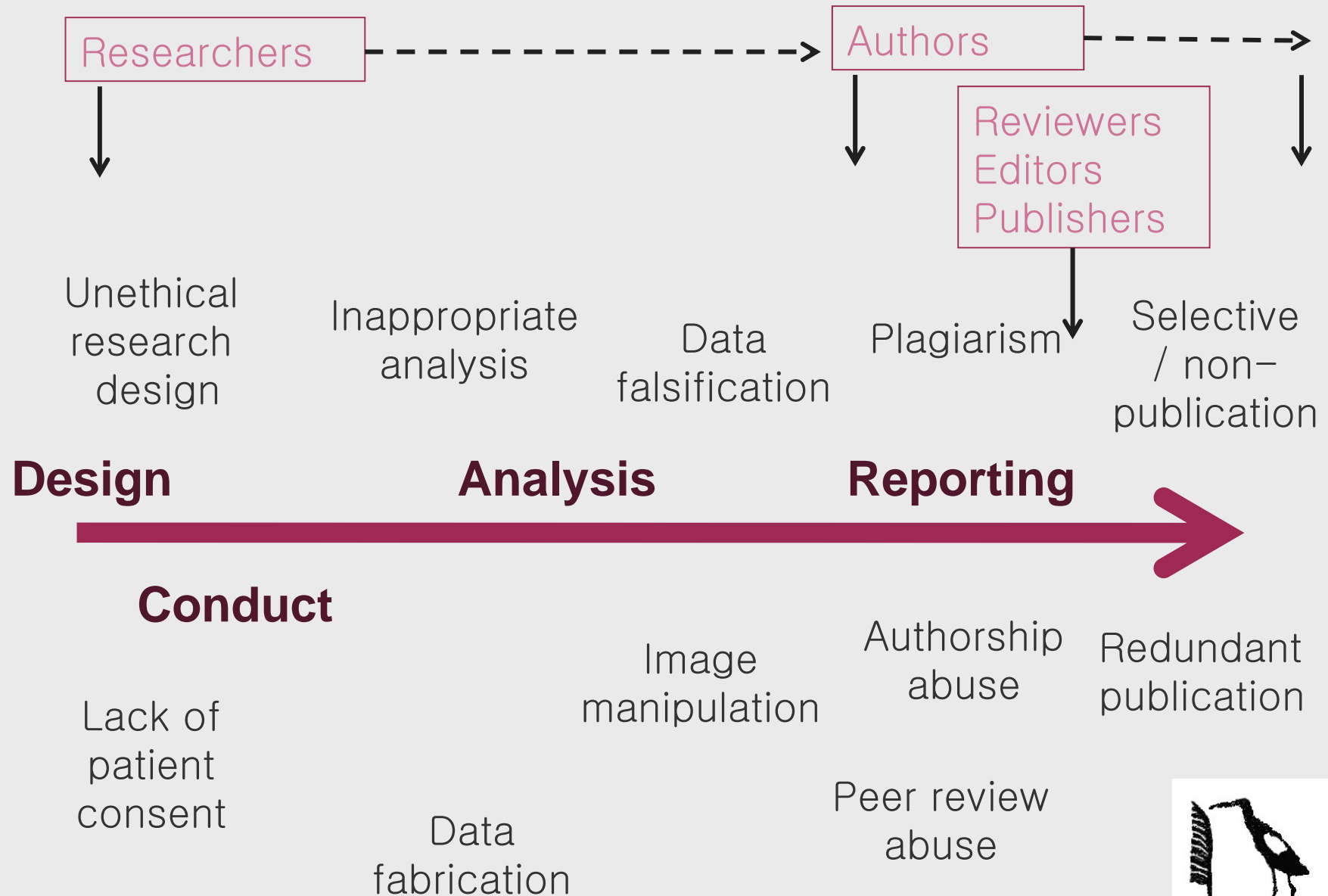
- Plagiarism
- Authorship abuse
- Redundant publication
- Undeclared Competing interests
- Reviewer misconduct



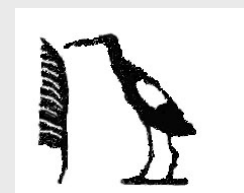
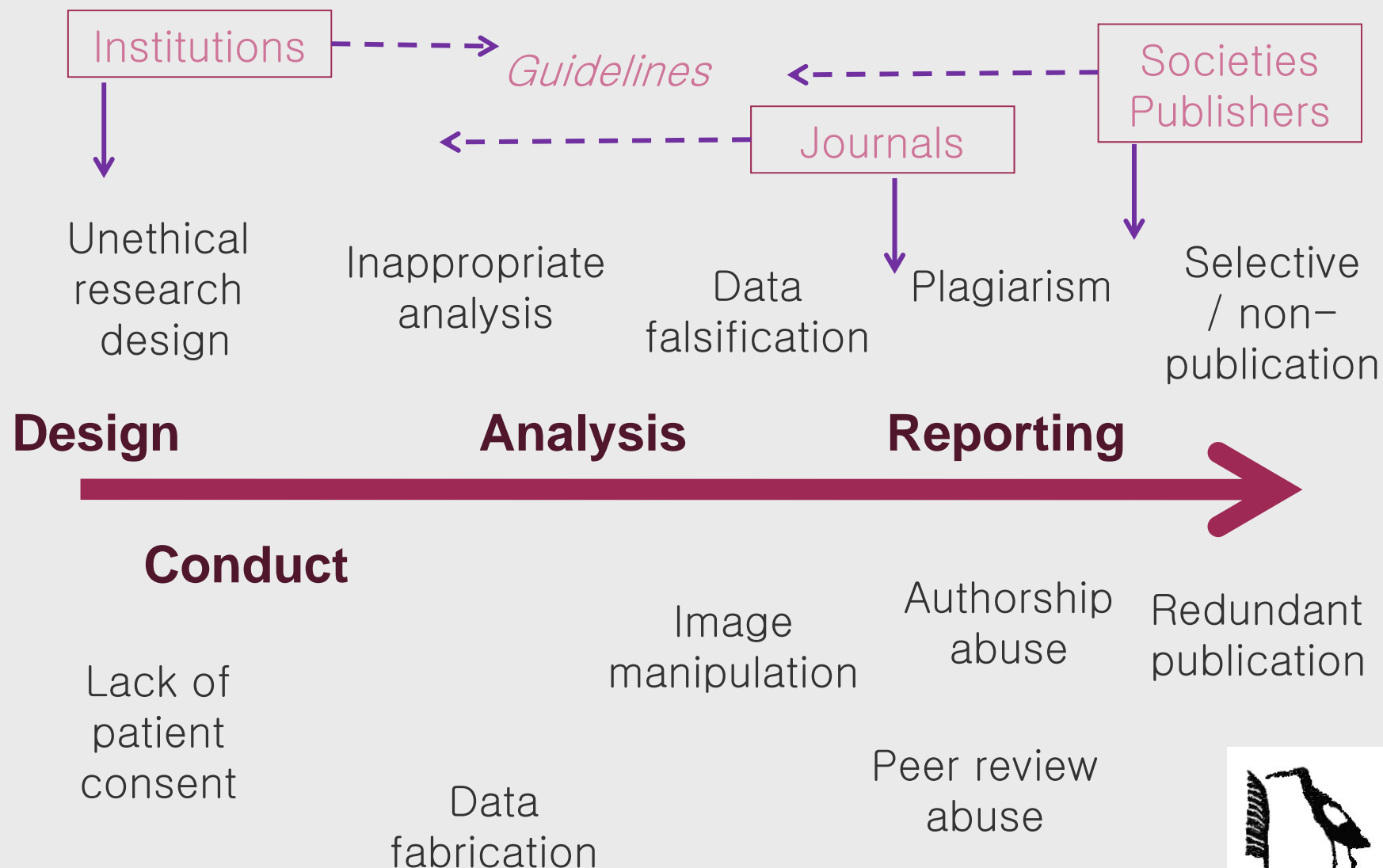
Research and publication ethics are a spectrum



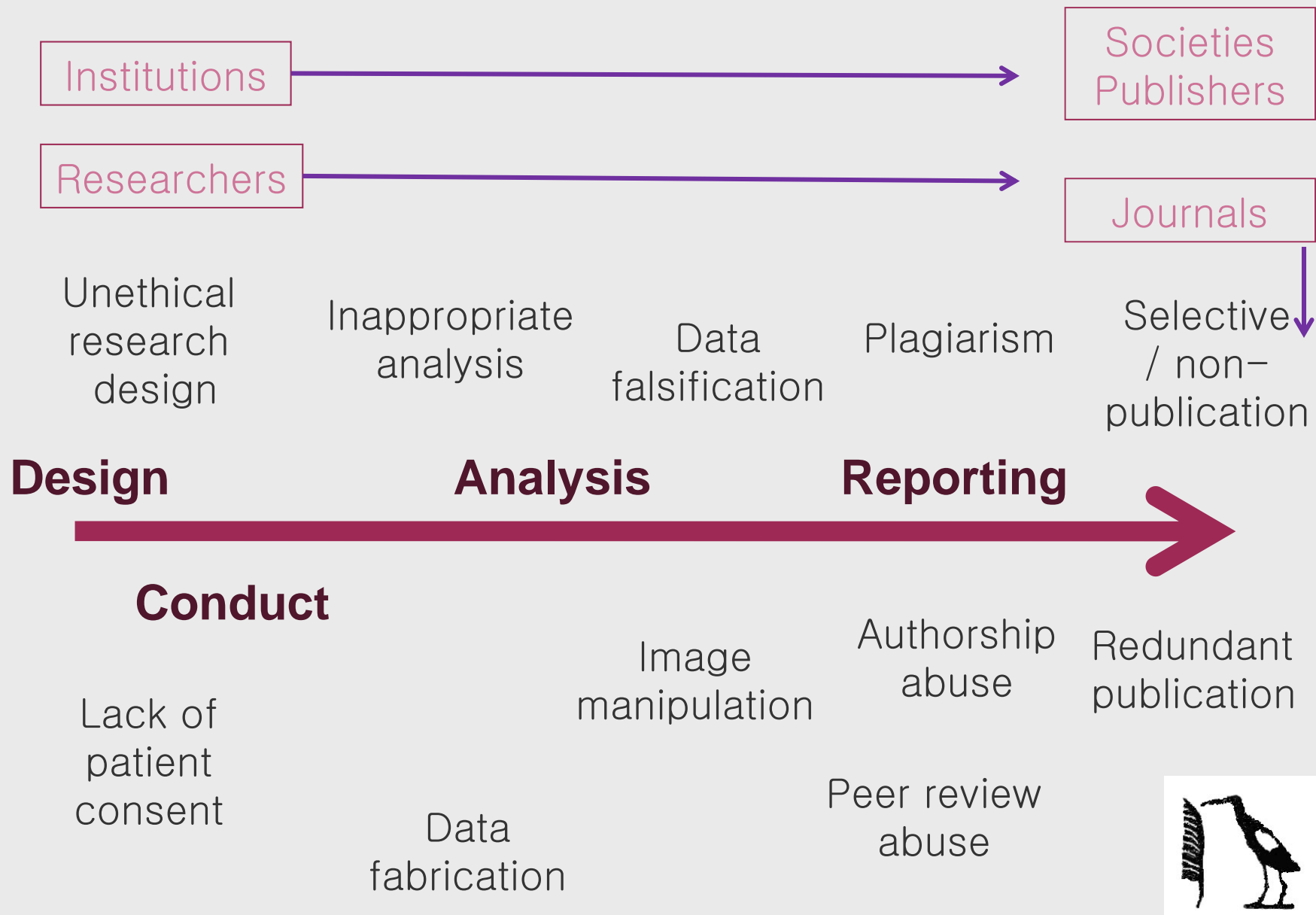
When does misconduct occur?



Who has influence?



Who is responsible?





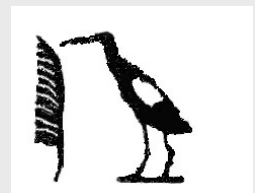
What can institutions do?

- Promote research integrity
- Educate researchers
- Examine effects of policies and incentives
- Detect research and publication misconduct
- Prevent misconduct (eg ethics review, screening)
- Investigate suspected misconduct
- Discipline researchers



From research to practice

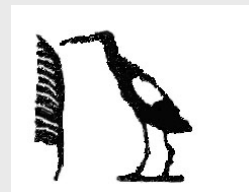
- To have a policy, institutions need a clear definition of plagiarism
- Institutions need fair processes to deal with cases of suspected misconduct
- Staff need support (tools, time, training)



Consistently implemented policies are important

- *“some students plagiarise deliberately and get away with it, and others, who do not intend to cheat are bewildered and frustrated to be punished for what they did not know was wrong, and a third group are angry because they know some students plagiarise and are not caught”*

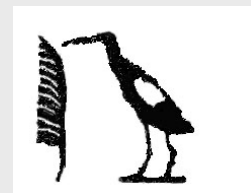
Diane Pecori: *Teaching to Avoid Plagiarism*, McGraw Hill, 2013





What can journals do?

- Promote research integrity (in instructions)
- Educate researchers (eg editorials)
- Examine effects of policies
- Detect **some** research and publication misconduct
- Prevent misconduct (eg ethics review, screening)
- Investigate suspected misconduct
- Discipline researchers



Institutions and journals have complementary roles

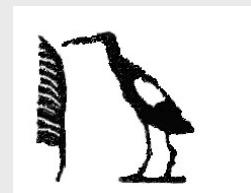
- Institutions and journals should work closely together, eg on cases of suspected misconduct
- COPE guidelines on cooperation



Cooperation between research institutions and journals on research integrity cases: guidance from the Committee on Publication Ethics (COPE)

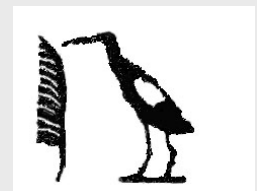
Different opportunities & responsibilities

	Institution	Journal	Society
Education	+	+	+
Promotion	+	+	+
Detection	+	+	
Investigation	++		
Prevention	+	+	+
Correction	+	++	



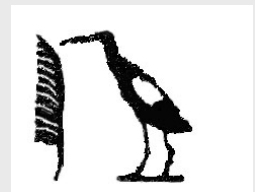
Conclusions

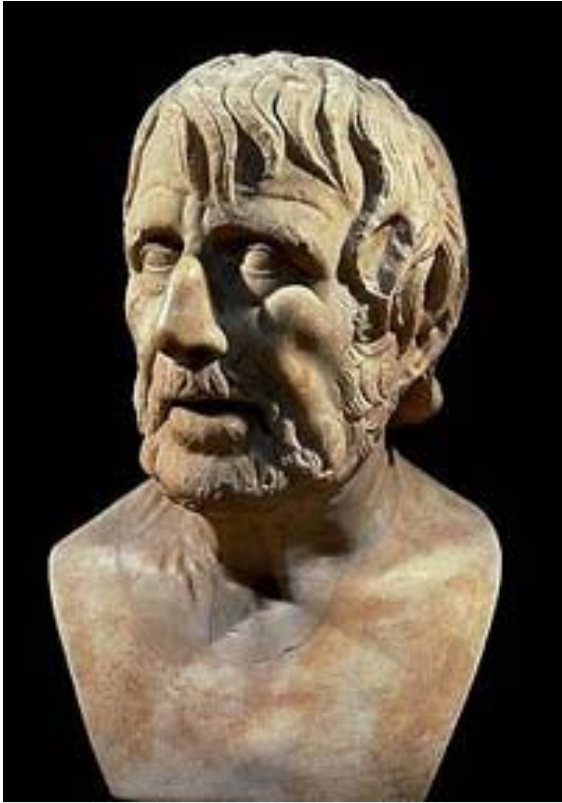
- Plagiarism is a major problem in academia and scholarly publishing
- Text-matching software is useful for screening and assessing but should be used in conjunction with:
 - training (for students and staff)
 - clear policies
 - support (for staff)



Conclusions (2)

- Institutions have a responsibility to encourage research integrity and to prevent, detect and investigate misconduct
- Journals and institutions should cooperate on cases of suspected misconduct





*“It is a vice to trust all,
and equally a vice to trust none”*

Seneca 4 BC – 65 AD

