HOW TRUSTWORTHY?

AN EXHIBITION ON NEGLIGENCE, FRAUD, AND MEASURING INTEGRITY
EXHIBITION CATALOGUE

HOW TRUSTWORTHY?
AN EXHIBITION ON NEGLIGENCE, FRAUD, AND MEASURING INTEGRITY

Edited by
Dr. Thorsten Stephan Beck, Melanie Rügenhagen
and Prof. Dr. Michael Seadle

HEADT Centre Publications
The goal of this exhibition is to increase awareness about research integrity. The exhibition highlights areas where both human errors and intentional manipulation have resulted in the loss of positions and damage to careers. Students, doctoral students, and early career scholars especially need to recognize the risks, but senior scholars can also be caught and sometimes are caught for actions decades earlier. There is no statute of limitations for breaches of good scholarly practice.

This exhibition serves as a learning tool. It was designed in part by students in a project seminar offered in the joint master's programme on Digital Curation between Humboldt-Universität zu Berlin and King's College London. The exhibition has four parts. One has to do with image manipulation and falsification, ranging from art works to tests used in medical studies. Another focuses on research data, including human errors, bad choices, and complete fabrication. A third is concerned with text-based information, and discusses plagiarism as well as fake journals and censorship. The last section covers detection and the nuanced analysis needed to distinguish the grey zones between minor problems and gross negligence in case of fraud.

Detecting fraud is important, and it is equally important for people in decision-making roles to understand when actual long term damage to scholarship has taken place, and when scholars have done something wrong but without negative consequences in the long run to both science and scholarship in general. Students especially need to learn the distinction between error and intent, negligence and gross negligence. Everyone makes mistakes, but undermining the reliability of scholarly results harms the whole scholarly community.
Images play an important role in many scholarly disciplines – in medical research as well as in art history. Images open up the invisible, they represent data and facilitate the understanding of complex contexts. Their power of persuasion is much praised – but the ability to manipulate images digitally has increased the distrust in them.

Faked images can falsify reality, simulate results, or show an idealized world. This part of the exhibition shows the different ways in which images can be manipulated and how difficult it is to detect errors or to expose forgeries. Examples are drawn from medical research, astronomy and art history, as well as from photo competitions. Look closely – what seems deceptively real is not necessarily authentic.
THE “GREATEST” FORGER

In 2010, Wolfgang Beltracchi emerged as a man who, along with three accomplices, forged a vast number of paintings attributed to dozens of renowned artists and sold them for millions to auction houses, collectors and museums.

In 2006, a Maltese company bought the painting Red Picture with Horses, attributed to Heinrich Campendonk. Beltracchi’s activities came to an end when Nicholas Eastaugh, a forensic art scientist, performed an analysis of the painting and discovered traces of titanium white, which was only used from 1916 on, while the painting is dated 1914. Beltracchi later blamed a mislabeled oil color tube for this mistake. The forger was sentenced to six years in prison for 14 of his forgeries. He claims to have created more than 300 forgeries, some of which may still be part of museums, galleries and private collections worldwide.

“So I took a zinc white from a tube, a Dutch product, but unfortunately it didn’t say that it contained a small amount of titanium white.”

Wolfgang Beltracchi, Art Forger and Artist

Red Picture with Horses
Attributed to Heinrich Campendonk (1889 – 1957)
Forged by Wolfgang Beltracchi
Oil on canvas
About 2005
Source: dpa - picture alliance GmbH
THE DARK SIDE OF THE MOON

What makes this copy of Galileo Galilei’s *Starry Messenger* special are the watercolor drawings of the moon. Assuming this would make the book extremely valuable on the market, the owner Richard Lan, a New York book seller, asked art historian Horst Bredekamp for an in-depth analysis of its authenticity in 2005. Bredekamp’s analysis concluded that there was no evidence of forgery, but that the “New York” copy was most likely a print template for all later editions of the *Starry Messenger*.

But some doubts remained. In 2011, Georgia State University art historian Nick Wilding compared the “New York” copy to other known copies of the work and discovered anomalies, such as misspellings, a closed circle in the stamp of the Cesi library, as well as some distorted portions of text, which are not typical for a letterpress print. After more evidence was collected, it became obvious that the “New York” copy indeed must have been a forgery – created from the digital version of another edition.

Given this proof, Bredekamp and colleagues started another research project in which they critically reflected on their own authentication methods. Since it was now clear the book was not an original, more invasive techniques could be applied. The material analysis of a fiber sample eventually revealed that the manuscript contained cotton linters, which were not used in paper production before the 19th century.

“I have learnt, in a bitter way, what I knew before, but not in this concrete sense: that phenomena can be looked at from different perspectives and that from different viewpoints they tell completely different narratives.”

*Horst Bredekamp, Art Historian*
Drawings of the "New York" Copy of the Starry Messenger
Attributed to Galileo Galilei (1564 – 1642)
Ink and watercolor on paper
Forged by Marino Massimo De Caro (before 2005)
Drawings by Dr. Thorsten Beck (2019)
In this image from Hubble Telescope, we see the “Eagle Nebula” – a young cluster of stars in the constellation Serpens with bright blues and greens swirling together in pillars of clouds. It may surprise some to learn that such images were, in fact, originally greyscale-colored and that the colors were added using color filters that illustrate the colors associated with the chemical elements in the nebula.

There is no doubt NASA and ESA/Hubble manipulated these images, but the question arises whether this form of coloring should be mentioned more prominently than is currently done. Scholarly research generally demands transparency, yet one of the largest research facilities does not state clearly that the colors do not reflect the Hubble Telescope image. Uninformed viewers may believe that by looking through a telescope, one sees these amazing and awe-inspiring sights, while the colors are actually rendered after Hubble Telescope took the greyscale image.

“This image shows the pillars as seen in visible light, capturing the multi-coloured glow of gas clouds, wispy tendrils of dark cosmic dust, and the rust-coloured elephants’ trunks of the nebula’s famous pillars.”

Hubble Space Telescope, 2015
Eagle Nebula, M 16, Messier 16
Post-processing space image
2015
Source: NASA, ESA / Hubble and the Hubble Heritage Team
(Accessed September 12, 2018)
NASA, ESA / Hubble and the Hubble Heritage Team
This photograph by Marcio Cabral was one of the winners of the Wildlife Photographer of the Year competition in 2017, but before long, third party accusations were made that the anteater in the photograph could be the taxidermied prop at the entrance of the park where the image was taken. The museum carried out a several week-long investigation.

While an expert panel of investigators concluded that the anteater in Cabral’s image bore too much similarity to the taxidermied specimen to be a coincidence, the photography website, Petapixel, made a case the image is genuine. The museum ultimately decided to rescind Cabral’s prize because, from their perspective, it violated the contest’s rule that “entries must not deceive the viewer or attempt to misrepresent the reality of nature.”

Unless the photographer can provide clear evidence he did not stage the photo, the museum will not reverse its decision, and will bar him from entering the contest ever again.

“After a thorough investigation, and a review of all of the information collated, a decision was made.”

Wildlife of the Year competition office
The Night Raider
Marcio Cabral
Photograph
2017
Source: Biosphoto
(Accessed November 4, 2019)
Marcio Cabral / Biosphoto
WHAT REMAINS OF THE MASTER’S HAND?

The “Savior of the World” is the title of the painting that goes down in history in November 2017 as the most expensive artwork ever sold. The auction house Christie’s scored nearly half a billion US dollars at the auction for a picture that was traded in 1958 for only 45 pounds. The spectacular rediscovery and time-consuming authentication of the image as an original da Vinci is documented on the auction house website. For more than 6 years, well-known experts examined the work of art before finally concluding that this was an original from Leonardo’s hand.

Critical voices, including the American art critic Jerry Saltz, have cast doubt on the authenticity of this work. This is not just about authorship – the work cannot really be meaningfully integrated into Da Vinci’s Œuvre – but rather how much of Leonardo’s original artwork is left. The picture was restored and painted over too often and too comprehensively to speak of a genuine Da Vinci. The claim is a cliché of how to imagine a masterpiece from Leonardo’s pen. The commercially successful auction of this “newly discovered old master” expresses a system in which money matters more than authenticity.

“Why is a Leonardo in a Modern and Contemporary auction?” Before I could say, “Yeah! Why?” he answered, “Because 90 percent of it was painted in the last 50 years.”

Jerry Saltz

Saltz, 2017.
**Salvator Mundi**  
Attributed to Leonardo da Vinci (1452 – 1519)  
Oil on panel  
About 1500  
Source: Christie's  
(Accessed September 12, 2018)  
Christie's
DISAPPEARING DATA

This image was included in a paper that was used politically as a reference for the Italian senate’s opposition to genetically modified food. The research showed that genetically manipulated soya bean could have an impact on goat offsprings whose parents ate those soya beans. Due to its political significance, Elena Cattaneo, an Italian senator and neuroscientist, took a close look and discovered various aspects of image manipulation. She then commissioned an Italian forensic expert with an analysis of the images.

Federico Infascelli – who headed the research team that produced this paper – was accused of self-plagiarism, the erasure of elements from the images and image duplication and recycling. After uncovering the manipulation, Dr. Cattaneo contacted the University of Naples, which conducted a research integrity investigation. As a result, the paper was retracted. Afterwards Cattaneo herself had to face accusations in the Italian senate for acting as a lobbyist for biotech giant Monsanto.

“The case is very important also because these papers have been used politically in the debate on GM [genetically manipulated] crops.”

Elena Cattaneo, Italian Senator
Manipulated Electrophoresis Gel
Federico Infascelli et al.
(1) Fig. 1 from: "Fate of transgenic DNA and evaluation of metabolic effects in goats fed genetically modified soybean and in their offsprings"
(2) Fig. 1 with altered contrasts and brightness (edited by Thorsten Beck)
Published in Animal, 2010
Source: Cambridge University Press
(Accessed September 12, 2018)
Cambridge University Press
Text is ubiquitous in the academic world, including in the natural sciences, where some scholars characterise the descriptions in articles as mere advertising for the real results in labs or databases. Nonetheless verbal descriptions still matter. Citations in academic papers have themselves become an industry, and some universities use citation counts as a factor in hiring, tenure, and promotion decisions. This has tempted some people to increase their citation count by citing themselves frequently, sometimes arguing that they want to avoid self-plagiarism this way. Plagiarism itself has become a widely discussed topic with commercial software systems to discover plagiarism, and with people who set strict rules about copying for cases that once represented a legitimate reuse of the language of a topic. Actual plagiarism is an ethical and potentially a copyright issue, but even serious copying does far less damage to the fabric of scholarship than data falsification.
The drive to publish has created a market for predatory journals. These journals charge for publishing through Article Processing Charges (APCs), and attract authors by promising fast turnaround via an abbreviated or entirely fake review process. The contents of these publications are not necessarily false, but they have skipped the benefit of a review process meant to catch errors and to detect untruths.

At the other end of the spectrum is censorship, which takes a variety of forms. Sometimes it involves local governments rejecting texts for libraries or schools, and sometimes it becomes a ban on works that express unwanted viewpoints or opinions. Not all forms of censorship are bad, such as those that restrict incitement to violence and hate-crimes, but it can also affect information about, for example, evolution, or lead to a ban on novels.
PLAGIARISM DETECTIVES

The plagiarism detection site VroniPlag uses a diagram that looks like a barcode to indicate the number of pages with text that – by their standards – match texts that other authors wrote. The barcode uses five colors: light blue means pages that are not relevant, white means pages with no matching, black where there is some matching (less than 50%), dark red means 50% – 75% of the page had matches, and red means that more than 75% had matches. VroniPlag also gives the total number of pages where they found any matches, and the percentage of pages for the whole work. The barcode graphic, the number, and the percent are persuasive in conveying an accusation, but do not convey a balanced view.

Without changing VroniPlag’s standards for what constitutes a match, more balanced numbers can be found by calculating how much matching occurred simply by using the colours. The example below (Dissertation Thiemo-Marcell Jeck) uses both the maximum percentage for the colour, and the average between the minimum and maximum percent for the colour:

<table>
<thead>
<tr>
<th>Colour</th>
<th>Pages with colour</th>
<th>max % colour</th>
<th>Pages with max % colour</th>
<th>Average % colour</th>
<th>Pages average % colour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>20</td>
<td>50%</td>
<td>10 pages</td>
<td>25%</td>
<td>5 pages</td>
</tr>
<tr>
<td>Dark red</td>
<td>4</td>
<td>75%</td>
<td>3 pages</td>
<td>62.5%</td>
<td>2.5 pages</td>
</tr>
<tr>
<td>Red</td>
<td>2</td>
<td>100%</td>
<td>2 pages</td>
<td>87.5%</td>
<td>1.75 pages</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td></td>
<td>15 pages</td>
<td></td>
<td>9.25 pages</td>
</tr>
</tbody>
</table>

This calculation means that instead of 26 pages (17%) with potential overlap, as VroniPlag claims, there were 15 pages (9.1%) using the color percentages at maximum values, and only 9.25 pages (5.6%) using the colour averages. These numbers could make a significantly different impression on those evaluating whether plagiarism occurred, while still using the same standards.


21. März 2013: Abkehrung → keine öffentliche Mitteilung

Website VroniPlag
Source: VroniPlag Wiki
(Accessed September 12, 2018)
Screenshot, September 12, 2018
A consensus has grown among scientists that climate change is a serious issue. Nonetheless the US Environmental Protection Agency removed its page specifically on climate change <https://epa.gov/climatechange/> in 2017. Other pages on climate change are still available, but Time Magazine notes that “The EPA’s site is now riddled with missing links, redirecting pages and buried information. Over the past year, terms like “fossil fuels“, “greenhouse gases” and “global warming” have been excised.”

There is good reason to believe that the missing information is a deliberate form of censorship by the administration.

Website of US Environmental Protection Agency
Source: United States Environmental Protection Agency (Accessed November 4, 2019)

8 Barron, 2018.
We want to help you find what you are looking for.

You can view an archived version of this content on the January 30, 2017, snapshot.

Other ways to help you find what you are looking for:

- Use our advanced search page
- Browse our curated A-Z index of terms and topics or see our automated list of website topics
- Search frequently asked questions or submit a question
- Go to the EPA home page
- To report problems with our website, or to suggest improvements to our website, please use the "Contact Us" link on the page where you see a problem, or where you would like to suggest an improvement. This link is located at the upper right and bottom of each web page.

You can also search our site.

What are you looking for?  

Search EPA.gov
Fake journals damage the integrity of publishing with a pretense of academic rigor and the goal of enriching themselves from research money.

The Süddeutsche Zeitung submitted an article to the Journal of Integrative Oncology on 20 July 2018, in which a fictitious author, "R. Funden" (equivalent in English to "I. Maginary"), wrote a fake study on "Die kombinierten Effekte von Essigsäureethyl-esterextrakten in Bienenharz auf das Absterben menschlicher Darmkrebsstellen" (English: "The combined effects of ethyl acetate extracts in bee resin on the death of human colon cancer sites"). The journal accepted the article with the claim that a reviewer wanted the label on a graphic improved and asked if the ethics commission had approved the study; with those minor changes the article was accepted and the author asked to pay 1892 € publication fees.

OMICS is the publisher of The Journal of Integrative Oncology, and the US Federal Trade Commission "filed a complaint against the academic journal publisher OMICS Group and two of its subsidiaries, saying the publisher deceives scholars and misrepresents the editorial rigor of its journals".9

There are blacklists (such as Beall's list of potential predatory publishers) and white lists (often maintained by subject specialties within universities), but scholars should also watch for the obvious signs: overly fast peer review without substantive comments, undocumented impact-factor claims, Article Processing Charges (APCs), and grandiose names or names similar to established journals.

Beall’s List of Predatory Journals

9 Straumsheim, 2016.
WHAT COUNTS AS PLAGIARISM?

A year after German Defense Minister Karl-Theodor zu Guttenberg admitted plagiarism, another member of the German government, Minister of Education Annette Schavan, was accused of plagiarism. Schavan never admitted plagiarism, but in 2013 the Universität Düsseldorf took away the doctorate that she had earned in 1980.

Less than the equivalent of 20 complete pages out of Schavan’s 335 page dissertation, are said to be plagiarized, but the accusers made their case by speaking in terms of the high number of pages (97) with plagiarism – this equals 29% of all pages. Here is an example from the blog schavanplag and its list of “unambiguous” plagiarism cases. The words in red are from the Schavan text that match words of Hannah Arendt. The words in blue in the Arendt text are those that do not match those of Schavan.

Schavan

“Durch die Gründung und Erhaltung von politischen Gemeinwesen schließlich schafft menschliches Handeln die Bedingungen für eine Kontinuität der Generationen und damit für Geschichte.”

Arendt

“[…] das Handeln schließlich, soweit es der Gründung und Erhaltung politischer Gemeinwesen dient, schafft die Bedingungen für eine Kontinuität der Generationen, für Erinnerung und damit für Geschichte.”

In this passage, 17 out of Schavan’s 23 words (74%) match those of Hannah Arendt, and 17 out of Arendt’s 26 words match those of Schavan (65%). The overlap is considerable, but not simple copying of whole sentences. The accusers apparently feel the overlap suffices to rule out chance selection.
1.4. Der Mensch – das handelnde Wesen [035]

Typus: Verschleierung

Dissertation: Seite: 015, Zeilen: 29-32

Fundstelle: Seite(n): 037, Zeilen: 09-12

Durch die Gründung und Erhaltung von politischen Gemeinwesen schafft menschliches Handeln die Bedingungen für eine Kontinuität der Generationen und damit für Geschichte.

Anmerkung: Die Urheberin des Gedankens, die politische Theoretikerin und Philosophin Hannah Arendt, wird in der vorliegenden Arbeit nicht nur nicht hier, sondern an gar keiner Stelle erwähnt.

2. Erziehung: Hilfe zur Personalisation [039]

2.2. Erziehung als Hilfe zur Personalisation [040]

Typus: Bauernopfer

Dissertation: Seite: 020, Zeilen: 01-07

Fundstelle: Seite(n): 044, Zeilen: 17-23

Neben der behavioristischen Psychologie, wonach unter "Sozialisierung" jenes Muster von Belohnung und Strafe verstanden wird, das sich z.B. im Verhalten der Mutter gegenüber dem Kind manifestiert, jene "child-rearing-practices", durch die der junge Mensch soziale Verhaltensweisen erlernt, die ihm die Eingliederung in die Gesellschaft ermöglichen [FN 3] [...].


Note: The remark at the bottom of section 1.4. says that Annette Schavan did not give any reference to the philosopher Hannah Arendt in her entire dissertation.


(Accessed September 13, 2018) Screenshot: September 13, 2018
Cornell University professor and former editor of the journal Perspectives on Psychological Science, Robert Sternberg, has been accused of self-plagiarism and an unusually high number of self-citations (150 during his editorship). Earlier in 2018 researchers, prompted a petition to the Association for Psychological Science.

Graduate student Brendan O’Connor opened the public discussion on Twitter. Nick Brown compiled his own findings and those of O’Connor in a blogpost at the end of April 2018. One of the most alarming examples was an article written by Sternberg that was comprised of 95% of two of his own works without referencing either one.

Recycling one’s own text could theoretically be a breach of copyright if authors have surrendered the copyright of their published articles. Some self-plagiarism is almost inevitable, but it becomes problematic when taken to extremes.
Image of Robert Sternberg's Article "WICS: A New Model for School Psychology"

Areas highlighted in green and yellow are recycled from Sternberg's other works.\textsuperscript{12}

Source: Nick Brown's Blog
(Accessed September 13, 2018)
Article published in *School Psychology International*, 2010
*Sage Journals*

\textsuperscript{12} Brown, 2018.
One of the principles of modern scholarship is the ability to build on past research results. This is true for the humanities as well as for the social sciences and especially the natural sciences. Fake or unreliable data undermines any new work that tries to build on it. The cases in this section all involve some form of false data that reached public notice, but not all were intentional fraud. Contamination of samples and disagreements about the interpretation of data played a role in some. More problematic are other cases where process-falsification undermined the credibility of the data, such as fake peer reviews or falsified consent forms. There are also cases where notable researchers just grew weary of the work involved in creating real data and knew enough about how people in the research world thought to reverse engineer the process in order to create data that reflected the desired results.

It is hard to say how much false data is part of the current scholarly record. Replication studies are not popular with journal editors or scholars, and many experiments go untested. Scholarly data are too infrequently made available for re-analysis, and recreating the data themselves is expensive and time-consuming. Some attempts have been made to create fraud detection tools, but the tools tend to be discipline-specific and often rely on statistical expectations whose fit to a particular case must always be examined.
MISIDENTIFIED CELLS

Misidentified cells quickly become a form of de facto contamination in lab sciences. In 1966, a geneticist and professor at the University of Washington proclaimed that cells he was working with appeared to be contaminated. In 2015, Christopher Korch looked into the issue, and concluded that “nearly 5,800 articles in 1,182 journals may have confused HeLa for HEp-2, and another 1,336 articles in 271 journals may have mixed up HeLa with INT 407”. More than 30,000 publications are said to use misidentified cell lines.

There is no reason to think that scientists deliberately misidentify cells, but once the misidentification has taken place, the problem spreads, since the samples themselves are a form of living data. Precisely how best to warn people about the problem is itself unclear. Retractions seem overly harsh in papers where some uncertainty still exists about the scope and effect of the problem. Published “expressions of concern” are one possible solution, but professors and research leaders need also to warn their doctoral students and staff.

Uncertainty about the identity of Hep2 Cell Lines?
Dr. Thorsten S. Beck
Photographic Collage Illustration
2018

13 Oliver, 2018.
This scene of a seating arrangement illustrates one of the studies that psychologist Diederik Stapel claimed to have carried out. Stapel argues that untidy surroundings triggered racist behavior. His setting was a train station in Utrecht (Netherlands), but he did not adjust his descriptions to the exact specifications of the actual station and got the number of seats wrong that he claimed were part of his experiment.

Stapel was a professor who did real research at some point in his past but eventually started to fabricate data that was more elegant and less messy than real data, and was more pleasing for editors. His fraud unraveled when in 2011 two graduate students voiced their suspicions to the department leadership. Almost 60 of his publications have been retracted by now. He lost both his job and his doctoral degree.
Jan Hendrik Schön claimed in 2000 to have used nanotechnology to show superconductivity in organic materials. Researchers failed to replicate Schön’s results, which raised doubts. In 2002, physicists noticed similarities between figures in Schön’s works.

Bell Laboratories, Schön’s employer, started an investigation, which ultimately concluded that he committed misconduct in 16 out of the 24 cases that staff evaluated. Bell Laboratories fired him after the committee reported that Schön misrepresented and manipulated data. Reich (2009) wrote “… in 2002, Schön revealed to his investigators how he had actually done it. He was doing science backwards. He started with the conclusion he wanted and then assembled data to show it.”

Jan-Hendrik Schön:
10 million nano-transistors on the tip of a needle?
Dr. Thorsten S. Beck
Photographic Collage Illustration
2018

Reich, 2009.
FAKE PARKINSON’S STUDY

This image illustrates an approach to treating Parkinson’s disease that neuroscientist Bruce Murdoch and Parkinson’s disease researcher Caroline Barwood promoted based on a study they never conducted. Murdoch had received grant money and forged consent forms to “prove” people agreed to participate. In the end he faced legal consequences.

Murdoch and Barwood described their study in a paper that they published in the European Journal of Neurology in 2011. The hint to raise initial doubts came from a whistleblower. Their employer, University of Queensland (Australia), investigated 92 papers and retracted three of the Murdoch/Barwood co-authored papers. The university ended up re-paying roughly AUS $175,000 in research funds. Murdoch and Barwood were ultimately sentenced to two years in jail. Murdoch admitted to 17 fraudulent cases.

Treatment for Parkinson’s disease?
Dr. Thorsten S. Beck
Photographic Collage Illustration
2018
Ex-Harvard psychology professor, Marc Hauser, claimed that tamarin monkeys reacted to the alternation of music patterns in the 2002 paper “Rule learning by cotton-top tamarins”, published in the journal *Cognition*. If true, this would have been an important discovery.

The research assistant who was asked to examine the research video of the tamarins, however, had a completely different opinion and felt that the monkeys did not have any reaction. This contradictory analysis of the result raised suspicion. Harvard’s report on the case says Hauser skewed his analysis and misrepresented results. The Office of Research Integrity issued a report after that in 2012 that claims that Hauser even fabricated data. In 2011, Hauser had already resigned from his position at Harvard, though he only admits to mistakes, not to research misconduct. The article in question as well as others have been retracted.
The integrity of information is complex and should not be judged in strict black and white / yes or no terms. Fair treatment for those accused of integrity violations requires a nuanced view using greyscale metrics. Greyscales show both the range of integrity levels and the degree to which each level indicates potential malpractice. Those levels allow nuanced and balanced decisions about consequences that affect people's lives.

Plagiarism greyscale metrics for texts without a reference:

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Label</th>
<th>Valuation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Significant matches</td>
<td>Multiple identical sentences in a paragraph</td>
<td>C1</td>
<td>Probable plagiarism</td>
</tr>
<tr>
<td>Some matches</td>
<td>≥5 contiguous words within a sentence of ≥5 words</td>
<td>C2</td>
<td>Potential plagiarism</td>
</tr>
<tr>
<td>Near copying 1</td>
<td>Multiple (&gt;3) exact phrases (&gt;3 words) overlap in contiguous sentences in a paragraph</td>
<td>C3</td>
<td>Potential plagiarism</td>
</tr>
<tr>
<td>Near copying 2</td>
<td>Fewer than 5 contiguous words within a sentence of more than 5 words</td>
<td>C4</td>
<td>Potential plagiarism</td>
</tr>
<tr>
<td>Similarity 1</td>
<td>Several (&lt;3) exact phrases (&gt;3 words) overlap in contiguous sentences in a paragraph or more than 9 words in sequence</td>
<td>S1</td>
<td>Questionable practice</td>
</tr>
<tr>
<td>Similarity 2</td>
<td>Many (&gt;5) of the exact same words (excluding function words) in contiguous sentences in the same paragraph</td>
<td>S2</td>
<td>Questionable practice</td>
</tr>
<tr>
<td>Topic overlap</td>
<td>Topic overlap with facts and standard phrases</td>
<td>S3</td>
<td>Probably not plagiarism</td>
</tr>
<tr>
<td>Standard words or phrases</td>
<td>Widely used words or phrases in this context</td>
<td>SW</td>
<td>Not plagiarism</td>
</tr>
<tr>
<td>General knowledge/facts</td>
<td>Information so broadly known that formulations cannot be considered unique</td>
<td>GK</td>
<td>Not plagiarism</td>
</tr>
<tr>
<td>Implied quotation</td>
<td>Something such as an indentation or colon implies that the passage was quoted, but no reference is supplied</td>
<td>IQ</td>
<td>Matching without an intent to deceive</td>
</tr>
</tbody>
</table>

This table is an extract from Seadle, 2017, S.38, with minor changes made for this exhibition.
The form of integrity violation that does the least damage to scholarship and to science generally is plagiarism, which is a legal and ethical problem, but does not falsify data. The tables on the following pages offer metrics to measure plagiarism problems. The same principles can be used to judge the far more complex situations that arise with data manipulation and falsification.

Plagiarism greyscale metrics for texts with a reference:\17:

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Label</th>
<th>Valuation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing quotation marks</td>
<td>A correct reference is there, but the quotation marks are missing</td>
<td>MQ</td>
<td>Bad practice, but no deception</td>
</tr>
<tr>
<td>Incomplete reference</td>
<td>A partial reference is there, but quotation marks are missing</td>
<td>IR</td>
<td>Bad practice, probably no deception</td>
</tr>
<tr>
<td>Poor reference</td>
<td>An inaccurate reference is there, with or without quotation marks</td>
<td>PR</td>
<td>Bad practice, potential deception</td>
</tr>
</tbody>
</table>

17 This table is an extract from Seadle, 2017, S.39, with minor changes made for this exhibition.

The same principles can be used to judge the far more complex situations that arise with data manipulation and falsification.

Data manipulation or fabrication greyscale metrics:

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Label</th>
<th>Valuation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing provenance</td>
<td>Provenance is necessary to establish data validity.</td>
<td>MP</td>
<td>Serious risk of data fraud</td>
</tr>
<tr>
<td>Inaccurate or unreliable provenance</td>
<td>Provenance should include evidence that is itself genuine.</td>
<td>IP</td>
<td>Possibility of data fraud</td>
</tr>
<tr>
<td>More than 10% of the data manipulated</td>
<td>This suggests tampering for samples $&gt;100$.</td>
<td>RDM</td>
<td>Risk of data fraud</td>
</tr>
<tr>
<td>More than 1% of the data manipulated</td>
<td>The reasons for the manipulation need to be convincing.</td>
<td>PDM</td>
<td>Possibility of data fraud</td>
</tr>
<tr>
<td>Some data (less than 1%) manipulated</td>
<td>Some data manipulation can be legitimate with a reasonable explanation.</td>
<td>SDM</td>
<td>May be acceptable</td>
</tr>
<tr>
<td>Missing description/ metadata</td>
<td>The data may be impossible to interpret correctly.</td>
<td>MD</td>
<td>A serious problem</td>
</tr>
<tr>
<td>Inaccurate description/ metadata</td>
<td>The metadata are verifiably incorrect.</td>
<td>ID</td>
<td>Potential fraud</td>
</tr>
</tbody>
</table>
DETECTING IMAGE MANIPULATION

Image manipulations that violate the rules of good scientific practice are a serious problem in many scientific fields. Especially if erroneous or deliberately falsified images go undetected, there is a risk that research based on them will produce incorrect results.

Detecting image manipulation is by no means trivial, and numerous research teams worldwide are currently working on developing algorithms for the automated detection of manipulations. Comparing and testing the efficiency of such algorithms requires comparable test data. The Image Integrity Database of the HEADT Centre collects images from retracted publications that have encountered problems with images. This comprehensive data pool makes it possible to test algorithms for manipulation detection on images from actual research.

For more information on the image integrity database visit: https://headt.eu/Image-Integrity-Database/
The HEADT Centre Image Integrity Database
Source: Image Integrity Database
(Accessed November 4, 2019, restricted access)
Screenshot: November 4, 2019
1_ THE “GREATEST” FORGER


2_ THE DARK SIDE OF THE MOON


3. SPACE FOR COLOR


4. EXPERTS AGAINST EXPERTS


5. WHAT REMAINS OF THE MASTER’S HAND?

6. DISAPPEARING DATA


7. PLAGIARISM DETECTIVES


8. CENSORING ENVIRONMENTAL INFORMATION


9_ PREDA TORY JOURNALS


NOTE – parts of this text come from:

10_ WHAT COUNTS AS PLAGIARISM?


11_ TEXT RECYCLING VS. SELF-PLAGIARISM


12_ MISIDENTIFIED CELLS


13_ FICTION SCIENCE


14_ ANSWERS BEFORE QUESTIONS


15_ FAKE PARKINSON’S STUDY


16 MONKEY BUSINESS


17 TEN SHADES OF GREY

EXHIBITION

HOW TRUSTWORTHY?
AN EXHIBITION ON NEGLIGENCE, FRAUD, AND MEASURING INTEGRITY

An exhibition of the Berlin School of Library and Information Science at Humboldt-Universität zu Berlin and the HEADT Centre (Humboldt-Elsevier Advanced Data & Text Centre).

The exhibition is the result of a project seminar offered in the joint master's programme on Digital Curation between Humboldt-Universität zu Berlin and King's College London.

Initial venue of the exhibition (autumn 2018):
Jacob- und Wilhelm-Grimm Zentrum, Humboldt-Universität zu Berlin

For further detail and contact please get in touch with ruegenhagen@headt.eu
EXHIBITION TEAM

Student Contributors
Sasha Agins
Tina Christensen
Laura Heikkinen
Ho-Ya Hsu
Christine Therese Kolar
Safete Neza
Jiaming Song
Iulia-Cristina S trafaloea
Catherine Alexa Tershak

Faculty
Dr. Thorsten S. Beck
Ulrike Liebner
Melanie Rügenhagen
Prof. Michael Seadle, PhD.
Katharina Toeppe-Hudy

Exhibition Graphics
Kerstin Kühl

Photographic Collage Illustrations
Dr. Thorsten Beck

Design Exhibition System
Julia Blumenthal, Head of Design Laboratory,
Image Knowledge Gestaltung. An Interdisciplinary Laboratory.
Cluster of Excellence Humboldt-Universität zu Berlin

Production
Radebeuler Machwerk e.K.

Production Print Products
15 Grad

SPECIAL THANKS
Julia Blumenthal, Interaction Designer
Prof. Horst Bredekamp, Art Historian
Nick Brown, Blogger, Nick Brown’s blog
Marcio Cabral, Photographer
Prof. Dr. Andreas Degkwitz, Director, Jacob-und-Wilhelm-Grimm-Zentrum
Amaya Steinhilber, Director, Einstein Centre Digital Future
Deborah Zehnder, Director, Image Knowledge Gestaltung. An Interdisciplinary Laboratory. Cluster of Excellence Humboldt-Universität zu Berlin