Spring 2019

COMPUTER GENERATED PAPERS AS A NEW CHALLENGE TO PEER REVIEW

Malea Dunne

COMPUTER GENERATED PAPERS AS A NEW CHALLENGE TO PEER REVIEW

by

Malea Dunne

A thesis submitted in partial fulfillment of the

requirements for the degree of

Master of Science in Technical Communication

Montana Tech

2019

Abstract

Computer generated papers (CGP) pose a serious problem to academic integrity and publishing. The problem began with SCIgen. Created in 2005 by MIT students, SCIgen is a software program that generates papers with simulated content. In 2014, we learned that more than 120 CGP passed through the peer review process, were published in well-known academic journals, and had to be retracted. I conducted research into the journal editing and peer review process to discover more about this problem and how it might be remedied. I conducted interviews with five journal editors from across the world, coded the information, and performed a thematic analysis. My thesis concludes with recommendations to control the CGP problem, including: increased awareness on the part of journal editors, CGP detection software, improving due diligence on the part of reviewers, and addressing the *publish or perish* paradigm that drives desperate faculty to compromise academic integrity by submitting CGP to journals.

Keywords: Computer Generated Papers (CGP), Peer Review, Publish or Perish, Academic Integrity, Academic Dishonesty, Journal Article Publishing, Journal Editors, Data Sharing

Dedication

For my Mom and Dad

Thank you for pushing me and giving me all your love and support.

Acknowledgements

I would like to thank the many different people who contributed to the work on my master's thesis and their constant advice and support:

Graduate Committee:

Dr. Pat Munday, Professor of Science & Technology Studies, Montana Tech, Graduate Committee Chair

Dr. Glen Southergill, Associate Professor of Technical Communication, Montana Tech, Graduate Committee member

Dr. Dawn Atkinson, Assistant Professor Writing Director, Montana Tech, Graduate Committee member

Dr. Charie Faught, HCI Associate Professor, Montana Tech, Graduate Committee member

To the entire faculty of Professional and Technical Communication

To my family and friends

I would not have successfully completed my Master's Thesis without you.

Table of Contents

Glossary of Terms

Term	Definition
Computer Generated Papers:	Online software program created to produce papers
SciDetect:	Detection software to notice generated papers
SCIgen:	Automated paper generator
Plagiarism:	Stealing another's words or work
Academic Dishonesty (AD):	Any form of cheating occurring in academia
Open-access Journal:	Research published and can be retrieved free of cost
Gibberish:	Nonsense
Gobbledygook:	Nonsense
Integrity:	Quality of being honest
Hoax:	Deception
SOFL:	Structured Object-Oriented Formal Language
Journal Impact Factor:	Measurement to see importance of journals and which are rated high
Academic Integrity:	Moral code in academia
Data Sharing	Data being used for scholarly research

1. Introduction

Computer generated papers (CGPs) pose a serious problem to academic integrity and publishing. The problem began with SCIgen. Created in 2005 by MIT students, SCIgen is a software program that generates papers with simulated content. The MIT students' goal "was to expose the lack of peer review at low-quality conferences that essentially scam researchers with publication and conference fees" (Bohannon, 2015, p.18-19).

I became interested in this problem when having a conversation with Professor Nick Hawthorne in his office one day. I was not sure what I wanted to research for a thesis, or a project and I went to Professor Hawthorne for advice. Professor Hawthorne said he heard about computer generated papers and suggested I look into this and see what I thought. I did some research and instantly noticed that this was an issue in the publishing world. I knew that CGPs were what I wanted to learn more about.

To investigate this problem, I interviewed five academic journal editors working in several countries. The interview questions came out of the key issues I identified in a review of the literature. During the interviews, I found that the participants generally believe the peer review process is the best way to distinguish genuine articles from CGPs. Despite this faith in peer review, CGPs have made it through the peer review process, have been presented at professional meetings, and have even been published in academic journals.

Some authors who have created CGPs and submitted them to journals have been faculty members. Publication boosts an academic's reputation is oftentimes required for promotion in academia. In some cases, individuals will compromise their integrity for publication. The number of CGPs submitted to journals indicates a serious problem with academic integrity at

universities today. Many believe this problem is driven, in part, by the so-called *publish or perish* pressure on junior faculty.

After I conducted and transcribed my interviews, I then coded the information based on the data and conducted a thematic analysis. This methodology proved an effective way to reveal more about the CGP problem. A few especially useful and interesting themes were how the editors became aware of CGPs, if they believed CGPs were an issue, and how they would improve the CGP issue.

With the increasing sophistication of CGP software, CGPs are likely to be a problem long into the future. This thesis is a step in better understanding the problem and concludes with suggestions for dealing with it. Two major recommendations I have for combatting the CGP problem are in-depth training seminars for peer reviewers and incentives for peer review work.

2. Literature Review

2.1. Historical Background of Computer Generated Papers

In 2005, three MIT students, Jeremy Stribling, Dan Aguayo, and Maxwell Krohn, created an online software program that produces computer generated papers. The MIT students' goal "was to expose the lack of peer review at low-quality conferences that essentially scam researchers with publication and conference fees" (Bohannon, 2015, p. 18-19). The students also submitted papers to see if anyone would notice that they were actually gibberish. The online software program SCIgen "throws random, jargon-laden sentences together to produce documents that seem like computer science papers. The program was designed to maximize amusement, rather than coherence" (Wiener-Bronner, 2014). When the program was first developed, its purpose was more for entertainment, like a game. But some authors began to use SCIgen to produce journal manuscripts, and some CGP papers were published.

Lavoie and Krishnamoorthy (2010, p. 1) assert that CGPs are "indistinguishable from papers that humans have written." Unfortunately, CGPs are showing up more in academia and academic journals. Though no sure ways of detecting these academic papers seem to exist yet, some authors believe the peer review process is a viable detection tool. As will be shown in this review, however, the peer review process is not always conducted in ways that detect CGPs. The French scientist/researcher Cyril Labbé submitted his own CGPs under a different name to see if readers/reviewers distinguished them from human- authored papers. Labbé also went one step further:

> Just as the students wrote a quick and dirty program to churn out nonsense papers, so Labbé has written one to spot the papers. He has made it freely available, so

> publishers and conference organizers have no excuse for accepting nonsense work in future" (Sample, 2014).

It is not yet certain, however, that Labbé's program is effective in identifying all CGPs.

While conducting research, Labbé found that more CGPs existed than originally thought. Computer generated papers are not necessarily created only by students, but by some faculty as well. "According to Labbé, the sloppy system can be traced back to high pressure on scientists to publish, which leads directly to too prolific and less meaningful publications'" (Wiener-Bronner, 2014,). When scholars publish academic articles, they are rewarded. Thus, this heavy emphasis on academic publishing alters academic culture, which encourages the creation of CGPs (Edwards & Roy, 2017, p. 1-11).

2.1.1. Academic Integrity and Computer Generated Papers

"Academic integrity is the commitment to and demonstration of honest and moral behavior in an academic setting" (The Writing Center, 2018). Although CGPs are published in journals, they are also a problem in college classes. After MIT students coded the SCIgen program to generate CGPs, the CGP genre changed from a fun game to a scholarly problem. Once news of the genre hit the publication world and editors became concerned about detection, they identified, sometime retroactively, CGPs in journals and conference proceedings. "On the one hand, technological developments have made the writing, publication and dissemination of documents quicker and easier. On the other hand, the 'pressure' of individual evaluation of researchers—publish or perish—is changing the publication process" (Labbé, 2013, p. 379). During the course of two years, Labbé sleuthed out more than 30 conference events that featured CGPs. "The publishers Springer and IEEE are removing more than 120 papers from their subscription services" (Van Noorden, 2014). Once these papers were published in big-name

journals, the effectiveness of the peer review process was called into question. This specific instance of false publication of CGPs was believed to have been done to "boost the authors' profiles" (Devlin, 2014).

Spinak (2014) studied the process of CGPs after Van Noorden (2014) published his article identifying 120 published CGPs. Spinak was initially confused by the thought of CGPs: "this news generated shock effect because these fake papers, automatically created by a computer program, had been accepted by prestigious journal publishers". CGPs are confusing but the inclination to create and submit them is fed by the intense competition and the *"publish or perish"* mandate among academic authors. This competition may thus drive academics to engage in "unethical scientific behaviors" (Amancio, 2015, p. 1764), leading to CGPs and other forms of cheating.

Cheating in college is also sadly becoming a trend. The percentage of students that are cheating their way through degree programs has greatly increased over the years (Josien, Laurent, & Broderick, 2013, p. 93). Academic dishonesty takes many different forms, including computer generated papers. Students who use CGPs as a form of cheating in college gain an unfair advantage, even though they may believe what they are doing is harmless (Josien, Laurent, & Broderick, 2013, p. 93). Although CGPs are more prevalent in academic publications, students may end also end up using the publications for research without knowing they contain invalid information.

Studies have pointed out that "cheating diminishes learning" (Preiss, Klein, Levenburg, & Nohavova, 2013, p. 157). Computer generated papers are ineffective shortcuts for student learning, and CGPs are a problem worldwide. When students participate in the cheating trend by submitting computer generated papers, they may graduate with high honors and publications and

secure sought-after jobs (Preiss, Klein, Levenburg, & Nohavova, 2013, p. 157-167). Using CGPs to get one's name out in the publication world is unethical and violates the honor code of learning (Teixeira & Rocha, 2008).

Lang's (2013) book *Cheating Lessons: Learning from Academic Dishonesty* explores why and how frequently students cheat. Lang's research found that a good portion of students do not have ethical integrity and will cheat. Computer generated papers makes cheating easier for those students. Lang's text is relevant to this literature review because, even though I am primarily interested in the problem of CGPs and journal publications, the student culture of today will likely contribute to the academic culture of tomorrow as students learn these habits and then enter the academic world.

2.1.2. Ethical Concerns with Computer Generated Papers

According to iThenticate (2013, p. 1), "plagiarism and other forms of misconduct are a growing problem in research." With CGPs becoming simpler to generate and submit, this form of cheating is becoming more common in the publication world. Ethical concerns are about a lack of moral or practical thoughts for one's own gain.

> The stakes in peer review are high: the decisions from this process determine who gets published, who gets funded, and who gets promoted. Beyond its impact on personal advancement, peer review maintains, for better or worse, the paradigm of any given scientific discipline insofar as the peer review of funding proposals determines what problems get studied and what solutions are allowed (Souder, 2011, p. 55).

Authors tend to believe that CGPs will not affect others, but that is not the case (DeAndrea, Carpenter, Shulman, & Levine, 2009, p. 944). This form of cheating is an ethical

issue in any aspect of research. The publication of these papers in medical journals, for example, can pose major problems.

> Not only is money and time being wasted trying to replicate questionable research, precious publication space is also wasted on duplicative papers. More importantly, the ethical issues are beginning to increasingly impact the level of trust that the public puts into the medical profession. Even worse, patients sometimes receive ineffective or harmful treatments based on poor or unethical research (iThenticate, 2013, p. 1-8).

To some, CGPs may not seem to be a problem, but they can be especially harmful in medical publications.

Authors who create CGPs also participate in "data sharing" as well (Vines, 2014, p. 44). An ethical issue with data sharing in publication is data is not checked until a paper has been accepted. CGPs are appearing more in high quality journals and this shows that academic peer review is not conducted properly.

CGPs and other articles that are flagged or retracted; removed or taken off of article, cause problems for specific journals. A retracted article can affect the reputation of a journal or publishing house, calling into question the integrity of the peer-review process. "In this environment, journals – particularly top tier ones – are faced with mounting submissions, which taxes the standard journal screening processes. This leads to a rise in 'bad research' allowed through the doors to publication" (iThenticate, 2012, p. 1-10).

A 2018 study conducted by Peled, Eshet, Barcyk, and Grinautski showed that cheating and plagiarism are linked to particular personality traits and cultural backgrounds. This study showed that it is easy to cheat online, and that the cheaters may not understand why it is

unethical to publish such work. Individuals who publish CGPs or other false work do so to boost their achievement goals and future careers.

2.1.3. Plagiarism within Computer Generated Papers

Plagiarism is an increasing problem, and can be seen within a CGP. Digital age plagiarism is changing and has made the process easier (Jones & Sheridan, 2014, p. 712). One issue with CGPs is that some of the information contained in these articles cannot be tracked with plagiarism software because it is nonsense.

> Back in 1999, Austin and Brown suggested that papers could be evaluated for digital plagiarism via general sight evaluation, then suspicious texts could be further investigated using digital technology ranging from searching the sources that might have been used by students through to starting to use the then emerging plagiarism detection software (Jones & Sheridan, 2014, p. 712-714).

The problem is CGPs cannot usually be detected with conventional plagiarism software because the text they contain is fabricated by the software and not lifted verbatim from existing articles.

Academic publishing is important for both faculty and students, but the whole system is undermined by CGPs and other false publications (Pillai, 2015, p. 279-291). Today, the internet makes false publication possible, but not always easy to locate (Batane, 2010, p. 1-12). Some authors may use CGPs to circumvent research requirements. The CGP does not require the author to know the actual information and "this has resulted in the urge to publish more" (Pillai, 2015, p. 279-291).

There are many different forms of plagiarism today, but CGPs have gone largely unnoticed. Authors do not seem to understand or care that they are putting false information out for the world to use. CGPs are an interesting aspect of plagiarism because the phenomenon is

new and not completely understood. Students may submit CGPs to pass their classes, whereas some faculty members use them to improve their careers.

2.1.4. Academic Publishing

Publishing is a thoughtful process, if done appropriately. Academic publishing is crucial because these publications are sources for practical applications and further research. Libraries rely on the integrity of academic publications and pay a price for these publications. “The academic journal publishing industry encompasses the creation, review, packaging and distribution of knowledge and/or information in multiple formats for use mainly by academic and scientific consumers” (McGuigan & Russel, 2008). “Automatically generated papers can be misunderstood as real papers, it becomes of paramount importance to develop means to identify these scientific frauds” (Amancio, 2015, p. 1). Both students and faculty rely on academic publications. Faculty and graduate level publication rely on extensive use of literature. A study was completed in 1997-2009 regarding how much faculty members use academic publications (Tenopir, 2009). This study showed that faculty members rely on academic journals in order to publish themselves. Even with the forms of publication changing, the degree to which faculty use them does not change.

2.2. Peer Review Process

2.2.1. History of Peer Review

Peer review “is the method by which grants are allocated, papers published, academics promoted, and Nobel prizes won” (Smith, 2006, p. 178). The peer review process aims to improve the worth of what is being published and uncover errors and possible plagiarism. “Despite many criticisms about the integrity of peer review, the majority of the research community still believes peer review is the best form of scientific evaluation” (Elsevier, 2018).

The peer review process is the most respected and dependable aspect in journals (Dadkhah, Kahani, & Borchardt, 2017) and should be a trustworthy process. This is because peer reviewers vet articles for publication. They thoroughly read, edit, and study every aspect of manuscripts.

Journals trust the peer review process to evaluate what they publish. Peer review is typically performed by more than one person and, in some cases, is done by a group of people. Peer reviewers are generally experts in the specific field of work covered by the manuscript.

Peer review is imperative for the academic world. Students and faculty members depend on the peer review process when they use scholarly articles – articles that have presumably passed peer scrutiny. However, changes in academic culture can interrupt or manipulate peer review, and "the peer-review system is poorly adapted to recent changes in the discipline and current societal needs" (Ferreira et al., 2016, p. 597).

2.2.2. Contemporary Problems with Peer Review

The peer review process is an extremely trusted and honorable system, but in some instances the peer review process deals with contemporary problems. "On the one hand, it's impressive that computer programs are now good enough to create passable gibberish" (Kakaes, 2014,). As Kakaes points out, the peer review process can be flawed. The process of peer reviewers consists of taking individuals that are "experts" in that specific field and expect them to completely understand what that article is about. "Some reviewers are unqualified and others, because of personal or professional rivalry, are biased" (Kassirer & Campion, 1994, p. 96).

Another issue with the peer review process is fake reviewers with a conflict of interest who will review an article and base their opinions strictly on bias or skip the process altogether (Tancock, 2018, p. 1-5). When this happens, few if any needed revisions may be made to the original manuscript. "Without peer review there is no control in scientific communication"

(Fennel, Corner, & Ash, 2018, p. 1). Some authors care only about getting their name on a publication. When the peer review process fails, this becomes a problem. This is where CGPs can come into play since they pose special problems for and exploit weaknesses in the peer review process.

2.2.3. Peer Review as a Method to Detect Computer Generated Papers

Peer review is crucial process for maintaining the quality of academic publication. "Peer review is widely viewed as an essential step for ensuring scientific quality of a work and is a cornerstone of scholarly publishing" (Bartoli, De Lorenzo, Medvet, & Tarlao, 2016, p. 19). Yet CGPs have been created that passed the peer review process. Peer review should identify information that is not relevant to a manuscript and result in the editor requiring the author to revise or remove it. During the peer review process if an article is "flagged" that means the article needs to be reevaluated (Kalnins, Halm, & Castillo, 2015, p. 1034-1038). If the peer review process is conducted correctly, CGPs should be flagged since the information in a CGP is typically nonsense.

Increasingly, for an article to be published, it has to be vetted through some CGP detection software as a supplement to peer review. SciDetect was created by Labbé specifically to check for CGPs. "SciDetect uses intertextual distance to discover automatically generated text materials" (Griffin, 2015, p. 23). This software works as a tool to identify false information, so that CGPs will not be published.

Researchers in addition to Labbé have worked on this problem. Williams and Giles (2015) used a process called Similarity Search to try and detect CGPs. Similarity Search was used to examine certain text in papers to distinguish if the information was real or fake (computer generated).

Nevertheless, in most subject areas, software cannot be used to detect CGPs, which leads to the reliance on traditional peer review. Peer review is important for screening manuscripts as peer reviewers focus on four aspects: research methods, relevance to readers, writing style/presentation clarity, and significance (Marsh & Ball, 2014, p. 151). If any of these aspects of a paper are suspicious, or are weak or missing, the article should be flagged for revision and perhaps not published at all. Peer reviewers must pay close attention to what they are reading in order to distinguish if the information is real.

2.2.4. Editor's Perspective on Computer Generated Papers

Editors are at the forefront of the CGP problem, but do not always deal with the problem successfully. "Editors have a responsibility to retract seriously flawed articles from their journals. However, there appears to be little consistency in [a] journal's policies or procedures for this" (Williams & Wagner, 2013, p. 1). As an editor, Bohannon noticed a flaw with the peer review and editorial process. Bohannon created a fake editor named Ocorrafoo Cobange, and he "submitted 304 versions of [a] wonder drug paper to open-access journals" (Bohannon, 2013, p. 60) to see if reviewers would notice inconsistencies. Bohannon continued this scheme for over 10 months, and more than half of the flawed papers were accepted by the journals for publication.

"A recent survey of editors of journals from different scientific fields showed that editors reported a 30% prevalence of authorship problems in their journals" (Marusic, 2011, p. 130). Although these editors showed a high prevalence of authorship, they did not believe there was serious harm to the editorial process. "Most editors of science journals seem not very concerned about publication ethics and believe that misconduct occurs only rarely in their journals. Many editors are unfamiliar with available guidelines but would welcome more guidance or training"

(Wagner, Fiack, et al, 2009). In another aspect, scholarly journals have looked at plagiarism within publication as a major issue, “the situation has become so alarming that some editors have publicly complained about the large number of journal submissions with plagiarized materials” (Roig, 2014, p. 557).

2.2.5. Open-Access Journals and New Approaches to Peer Review

Though not directly related to the CGP problem, open-access journals may compromise the peer review process in ways that make it easier to publish CGP. “Open-access publishing gives lawful free access to journal content on the internet and is funded by means other than readers’ subscription” (Schroter, Tite, & Smith, 2005, p. 1-4). Typically, in order for journals to publish there is a payment, whether it is by the author or publishing house. Open-access journal content is then free for the world to use. “The open-access movement, although noble in its intent, has been an unwitting host to… parasitic publishers” (Pisanski, Sorokowski, & Kulczcki, 2017, p. 481).

Some open-access publications are predatory journals that will publish nearly anything for a fee and/or lack peer review (Wicherts, 2016, p. 1-19). “Bogus journals can imitate legitimate ones that also collect fees from authors” (Pisanski, Sorokowski, & Kulczcki, 2017, p. 481). To define which manuscripts submitted to journals are legitimate and which are fake, a strong peer review system is needed. “Peer review is a core mechanism for quality control in scientific publishing, but the quality of peer review itself is often obscured by the fact that it takes places behind closed curtains in most journals” (Wicherts, 2016, p. 2).

3. Methodology

3.1. Qualitative Methodology

I selected qualitative methodology for exploring the CGP issue. Unfortunately, qualitative methodologies have not always been developed as rigorous, well-defined tools. As Attride-Stirling stated in her 2001 article "Thematic Networks. An Analytical Tool for Qualitative Research," "the growth in qualitative research is a well-noted and welcomed fact within the social sciences; however, there is a regrettable lack of tools available for the analysis of qualitative material" (Attride-Stirling, 2001, p. 385). My search, therefore, was for qualitative methods that could be applied in a rigorous way.

We use qualitative methodologies to search for answers to research questions and to gather data. Qualitative methodology is analytical, yet flexible. "One advantage of qualitative methods in exploratory research is that use of open-ended questions and probing gives participants the opportunity to respond in their own words, rather than forcing them to choose from fixed responses, as quantitative methods do" (Family Health International, n.d, p. 4).

Qualitative research investigates how people think and react to given problems or questions and focuses on gathering rich, expansive data sets from a small number of participants. Interviewing is a type of qualitative methodology.

"The discussion in the educational literature concerning qualitative research can be clarified by recognizing that qualitative research comes in many different varieties, which can be more clearly identified and understood by using the notion of research traditions" (Jacob, 1988, p. 16). The qualitative approaches that I found to be best suited for investigating CGPs and their issues were interviews, participant observation, and thematic analysis.

3.2. Interview Methodology

I chose data obtained through interviews as a key to understanding the CGP problem from the point of view of journal editors. “Interviews are one of the most widely used and most fundamental research techniques – and for very good reason. They enable researchers to obtain information they cannot gain by observation alone” (Berger, 2016, p. 191). Interviews can be conducted face-to-face or via phone, email, or interactive video conferencing software such as Zoom. “The expert tries to involve the interviewer in ongoing conflicts in the field and talks about internal matters and intrigues in his or her work field instead of talking about the topic of the interview” (Flick, 2002, p. 165). In the case of my study, face-to-face interviews could not be conducted due to the time and cost of traveling to visit editors in various U.S. states and abroad. “In contrast to biographical interviews, here the interviewee is of less interest as a (whole) person than in his or her capacity of being an expert for a certain field of activity” (Flick, 2002, p. 165).

The procedure of the interview process is very intense: “if the goal of the qualitative researcher is to elicit detailed description and understanding of communication behavior directly from participants through an interview, five different, though somewhat intertwined, steps need to be followed” (Stewart, 2002). These steps are (1) frame a research question, (2) decide the interview format, (3) collect data, (4) transcribe the data, and (5) analyze results. Before interviews are conducted, it is important to develop and test a set of questions. To keep interviews to a reasonable time period and to encourage participation, I developed 10 questions about CGP and the peer review process.

3.3. Participant Observation

As a method of gathering data about communication, participant observation is also an effective approach. “The method of participant observation allows for a consideration of the

empirical fabric of materiality and an application of the analytical concept of presence" (Aagaard & Matthiesen, 2015, p. 40). "In participant observation, researchers become involved in the group, organization, or entity they are studying. Researchers have to balance two roles: that of being participants and that of being observers" (Berger, 2016, p. 253).

> Participant observation, whereby the researcher interacts with people in everyday life while collecting information, is a unique method for investigating the enormously rich, complex, conflictual, problematic, and diverse experiences, thoughts, feelings, and activities of human beings and the meanings of their existence (Jorgensen, 2015).

This form of qualitative methodology takes time and patience because each participant involved may handle the CGP situation differently.

As a way to connect with another individual and understand his or her thoughts, participant observation can be an important complement to interviews. When first developing my research design, I anticipated that this form of qualitative methodology could help me better understand the participants' thoughts and decision processes. As a research tool, participant observation helps transcend the limitations of interviews, which are often thought of as data collected from "naturalistic verbal reports" (Aagaard & Matthiesen, 2015, p. 40). There are many examples of effectively using participant observation to understand informants' thoughts and decision processes, such as the important breakthroughs in understanding science communication achieved by Latour and Woolgar (2013).

Although participant observation is an excellent means of gathering data and understanding a communication process from the point of view of one's informants, and can incorporate informal interviews, I had to reject this methodology since it requires close and

extended observation of a situation, such as what happens when working with or living among one's informants. Participant observation did not work for my particular research due to my international group of interviewees.

3.4. Thematic Analysis

Thematic analysis is a useful approach to qualitative analysis of texts, including interview data. "Thematic analysis (TA) is a method for identifying, analyzing, and interpreting patterns of meaning ('themes') within qualitative data" (Clarke & Braun, 2017, p. 297). To enable thematic analysis, I developed a coding scheme after collecting and reading over the interview data. "Researchers use coding to help identify common themes and topics that may emerge from the interview transcripts; these common themes will help researchers see what is important to informants and what is secondary" (Berger, 2016, p. 203).

In TA, coding depends upon a list of keywords, sentences, and/or concepts that are then paired with similar wording or meaning in the interview data. "Codes are the smallest units of analysis that capture interesting features of the data relevant to the research question" (Clarke & Braun, 2017, p. 297). Although coding and content analysis are often used to build a database for quantitative analysis, this method is also useful for qualitative analysis when considering transcripts of interviews. The idea, especially when paired with thematic analysis, is to discover the themes or ideas that structure meaning.

Using TA, researchers can uncover patterns and views that provide answers to research questions. As a relatively novel topic, some editors may be unaware of the CGP problem, while others might have personally dealt with it. Thematic analysis provides a means to compare the interview data from various editors in a way that establishes the range of how the CGP issue is understood and dealt with.

4. Thematic Analysis

4.1. Introduction

The thematic analysis process began with five interviews with editors from across the world, including several prominent journals and/or publishers. These included: (1) David Grier; editor of *Computer Journa*l in *IEEE*; (2) Tamara Welschot; Springer Nature's Research Integrity Group. "The Springer Nature Research Integrity Group aspires to a positive and proactive approach to preventing publication misconduct and encouraging sound and reliable research practices."; (3) Joaquim Jorges; *Elsevier Computers and Graphics*; and two editors, (4) and (5), who wished to remain anonymous. Once the interviews were conducted and organized, the process of coding the transcripts to identify crucial information surrounding awareness and becoming aware, issues, improvement opportunities, line of defense, rejection rate, COPE, retraction, publication ethics and scanning tools. I read through every transcription and defined common themes that best explained what editors said. These themes stood out in the editors' approach to the CGP problem.

4.2. Awareness

As demonstrated by an article published in the journal *Nature* (Van Noorden, 2014), computer generated papers (CGPs) have become a serious and relatively well-known problem in the publication world. Academic journal editors are a frontline defense against CGPs masquerading as legitimate research. When a CGP is published, that generally means the peer review process somehow failed to identify bad information in that specific article. When I asked five editors if they were aware of CGPs, three answered that they were aware, one said "I have heard this before, but I don't know that much. I don't think it is common in science and engineering publications" (E1, 2019), and David Grier, who has worked personally on the CGP

issue, stated, “In a small tight community of us, everybody knows about it [CGP] and worries about it. …. It’s something of a concern [and has] been a concern on and off for probably twenty years” (DG, 2019).

4.3. Becoming Aware

Some editors routinely come across CGPs. “There was a period, roughly around 2013 when plagiarism and mechanical plagiarism became an issue, at that point I was responsible for the computer journals” (DG, 2019). CGPs have sometimes gone undetected during the peer review process and have been presented at conferences lacking peer review. “I became aware of these through the infamous SCIgen MIT Paper generator that was used to produce a paper accepted at a conference (SCI) without a peer review system” (JJ, 2019).

“We had one [CGP] published, not in a magazine but in a periodical of a conference proceedings. It [CGP] somehow got on the front page and someone levered it against us” (DG, 2019). Editor Two became aware of CGPs when a colleague tricked him into reading one. “It took me three paragraphs to notice it was a fake paper” (E2, 2019).

Once these papers started showing up in journals, they became an embarrassment within academia and academic publishing. Tamara Welschot (2019) became aware of CGPs when she received an email from a scientist warning her about “papers that were apparently very fake.” Thus, from the origin of SCIgen in 2005 to the *Nature* exposé in 2014, ripples of the CGP problem spread through the journal editing community.

4.4. Issue

According to Van Noorden with *Nature* (2014), CGPs have clearly become a problem in academic publishing. Editor One reinforced this judgement: “If it [CGP] happens, it is a very

serious problem" (E1, 2019). However, the other editors that I interviewed disagreed. "The CGPs are not a problem for high quality publications. In fact, it is bitter for low-quality journals and venues that they can be fooled this way. It shows that they are not rigorous" (E2, 2019). "In the engineering community the answer is largely no [CGPs are not a problem], but in the world of the humanities crowd has a different problem" (DG, 2019).

Two editors believed that CGPs were not a serious problem so long as the peer-review system operates as it should. They did, however, admit the possibility that improved algorithms could pose a future problem:

> By looking at the paper and reading it, you should get an idea if this is a hoax paper or a fake paper. I think we should realize that algorithms [for generating CGPs] are becoming much better, and although I am personally thinking that still you can find out about some papers that are machine generated because there is something in the papers that are not matching up (TW, 2019).

> The way I see it, a journal with a decent peer-review system should be able to spot SCIgen generated papers. However, SCIgen may only be the precursor of more sophisticated approaches that may become progressively harder to unearth (JJ, 2019).

4.5. Opportunities for Improvement

With the recognition of the current CGP problem and concerns that the problem could grow, editors recognize opportunities to improve the peer-review process. According to Editor Two (2019), "There are always ways to improve" the CGP issue. David Grier (2019) suggested an incentive for editors and journals to improve CGP screening:

> I mean the strategy on how to get the incentive for this is having one journal or a set of journals where you viewed them as having a greater status than others. And, in particular, if you took a top bit journal and said okay we are only going to have only trained referees and you worked up and built even a small staff and said this is what you'd get, it's part of building a premier product. If you went through that and your article was refereed, we would know the quality of it and the author would get high quality comments back and the work would improve substantially and be able to build a community that this is important, this is good.

Along with providing an incentive for reviewers, Grier also suggested a training seminar on how to review manuscripts. "I think some of the professional societies could do it. …. But one of the challenges again is getting the referees to do it and getting them to feel there is a sense of accomplishment there, that's a tough bit." Grier elaborated:

> At this point, it is spread so widely, they really do no training or very limited training on it. And how you would bring reviewers on and sort of train them and get them to understand the process and think about things and what they should look for would be a good training method. How you get people trained and how you get them to accept them and get common standards, that would be hard.

Better peer-review, or training for peer reviewers, was just one solution envisioned by editors. "Nowadays there are software detectors for similarities to identify plagiarism" (E1, 2019). Similarly, "fake and plagiarized papers need to be fought with software tools" (JJ, 2019).

Although some editors believe there are opportunities for improvement within the peer review process and regarding the CGP problem, others think there is no need for improvement. The latter group maintained the view that CGPs are simply not a problem in journal publishing.

> Well we have various models of peer-review. Essentially it really means that any papers being reviewed by two reviewers, that's the best practice. And of course, you have single lined and double lined and open review and transparent peer-review. But it actually boils down to two experts reading it [the manuscript] or reviewing it and coming back with comments. I think that's still the best way of making sure that what gets reviewed or published has some degree of screening (TW, 2019).

In the end, whether editors expressed the need for improvement or not, the CGP screening issue largely depends on the abilities of reviewers, perhaps combined with the capacity of screening software: "The reviewers need to read the paper carefully, they need to be experts in the field, and they should use some tools, such as iThenticate" (E2, 2019). iThenticate is a plagiarism software detection tool.

4.6. Line of Defense

When a manuscript is submitted for review, the editor often carries out an initial screening, for example to determine if the manuscript is a good fit for the journal. The editor then sends the manuscript to selected peer reviewers who must follow a process and return the manuscript to the editor with an evaluation—often in support of or against publication—along with comments regarding needed revisions. Even if recommended for publication, the author may have to go through one or more rounds of revisions before the editor approves the manuscript for publication.

During this process, the journal's editors have a key responsibility to screen for CGPs: "I think the first line of defense would be the editor-in-chief and then the next step would be the reviewers" (TW, 2019). Peer reviewers and editors have a crucial role in publication, but peer reviewers and editors might be focusing on more than one paper at a time. This can leave gaps

that CGPs can penetrate: “We [IEEE] have a staff of roughly 20 associate editors. All of them are currently working on one to three papers in the peer review process” (DG, 2019). Editors and authors must be patient as a manuscript winds its way through the review process. If the process works, then CGPs should not make it to publication: “As long as the editors and then the reviewers do diligently their work, it should not be a problem” (E2, 2019).

4.7. Rejection Rate

Not all manuscripts submitted to journals are published. Sometimes, rejection occurs in the first step when a manuscript is received by the editor. In part this is a judgement call as to whether the manuscript is a good fit for the journal, and in part it is out of consideration for the peer reviewers. Joaquim Jorge (2019) stated, “I currently reject about 20% of all submitted papers [manuscripts] without sending them to reviewers for various reasons. The main one being out-of-scope papers. Reviewer fatigue is a serious problem, that can lead to bad papers being accepted.” Ideally, due diligence on the part of editors and peer reviewers will detect CGPs:

> The peer review process will definitely filter out these papers. I think in the time of 2014 there was too much trust in having the peer review process handled by others. This has led to this problem, thankfully for us only 18 papers and not 100 papers like IEEE. (TW, 2019)

4.8. Retraction of Papers

In some instances, CGPs have passed the review stage and been published. IEEE has retracted about 100 CGPs whereas Springer has retracted about 18 CGPs. If CGPs that are published later get detected, then the journal must retract the published paper. This is a painful process:

> This [retraction] can possibly be the most embarrassing event to happen to an editor. The text would have to be retracted as soon as detected, identifying the paper as a fake manuscript, with a personal note of apology from the Editor-In-Chief, who would have to have to assume the blame at a personal level (JJ, 2019).

Before a paper is retracted, the editor will approach the author to ask for an explanation, and then advance to the retraction stage. When a paper is retracted from a journal, some journals will post a notice to clarify that retraction has occurred. "There is a banner that says, 'Paper Has Been Removed' and is intellectual property" (DG, 2019). Grier has been involved with approximately "100-200 retraction cases a year" (DG, 2019) over the last decade. Not only are there notices posted to inform readers of the retraction, but the author who submitted the CGP will also be banned from publication with that journal or publisher for a period of time. Depending on the nature of the retraction, the author could be banned from publication for a year or to 10 years.

Editor Two explained their journal's retraction process in great detail:

> If this [retraction] would happen we would start a formal investigation with the Ethics committee. If the result would be positive, we would recommend actions. First, we would inform the authors and ask them for a retraction. If [the authors maintain that] the retraction would be unsatisfactory, the journal would: a) inform the authors about the final decision; b) retract the paper with a mention that it was retracted, with full names of the authors, their affiliation, and the name of the paper; c) depending if this was intent or an error, we would also inform the direct supervisors of the authors about the academic misconduct; and d) the author could also receive a letter that they are not welcome to publish in our journal for certain period of time (usually around 3 years) (E2, 2019).

4.9. Pressure to Publish

The pressure on academics to publish, a situation anecdotally referred to as *publish or perish,* has helped feed the CGP problem. Professors and new faculty are often pressured to publish a certain number of articles in order to be successful in their careers. This pressure has also increased the amount of work for editors and reviewers. “I see CGPs happening as a consequence of the pressure to publish. The submission of large quantities of low-quality manuscripts is motivated by a numbers-focused quantity-over-quality evaluation process for untenured faculty” (JJ, 2019).

The pressure on junior faculty to publish or perish seems to have increased in recent years. Grier explained:

> About a generation ago the faculty that are involved in some colleges had much more limited publications on them then they do now. Some would be expected to write a paper or two to get tenure over a 7-year period. They would be expected to continue to write papers at that kind of pace for the rest of their career so they might end up having 10-12 papers that would be to their name and would mark their career. That is no longer the case. Graduate students coming out used to be expected to have 1 to 2 papers, now it’s about 5 to 6. You are expected to turnout as a new professor in the engineering world or scientific world a paper or two a year. If you are running a research lab, more of a senior professor, you would like to see 8 to 10 come out of your lab a year (DG, 2019).

Publication can make a faculty member’s career that much more successful, but it can also create motivation to submit CGPs or cheat in other ways for publication. “When it becomes your job, in the whole case of plagiarism….you cut corners and in particular all of these groups generally lacked the full skill to write a paper” (DG, 2019).

When faculty members publish articles, they may see a rise in pay and status and be rewarded with promotion or tenure.

> This work is tied up in both prestige and in income and those two things are not the same, but they are often tied up in it. And what people will do to keep their position and to build their sense of identity and authority and prestige on a fault basis. (DG, 2019)

4.10. Publication Ethics

Some editors, at least, believe academia is a long way from solving the CGP problem. Academic integrity, and specifically publication ethics, must be improved before the problem will go away:

> The root causes need to be addressed to solve what is otherwise a very human problem. Until that happens, I could see further and more capable strains of text generators appearing. Indeed, there has been considerable research on computer-generated novels and poetry. Look at the high-profile retractions of manuscripts with fabricated results from *Nature*, *Lancet* and *Science*. Having a program to assist in fabrication is just another way to make the process easier (JJ, 2019).

Publishing a CGP crosses many ethical guidelines. When authors submit CGPs, they undermine their own academic integrity along with the status of academic publications. Researchers rely on the information they find in published articles, so CGPs can affect the results of other people's research. Though not directly connected with the CGP problem, there are many other issues in academic integrity and publication ethics that complicate the work of editors and peer reviewers. One is the seeming repetition that makes it hard to judge whether authors are just repeating what they have said before (self-plagiarism) or actually saying

something new: “Also, there are many papers that have an incremental contribution and it is difficult to distinguish if it actually is self-plagiarism or a new idea. These papers require a huge amount of work to judge and it is very tricky” (E2, 2019).

4.11. COPE

The Committee on Publication Ethics (COPE) is a group of individuals that focus on the ethical guidelines of publication. Most editors in the current study were either aware of COPE or have served on its board. Four out of the five editors I interviewed were explicitly aware of COPE, although one seemed only vaguely aware of the group.

Editor One, Joaquim Jorge, and Tamara Welschot all represented their journals and publications within COPE. In following COPE’s guidelines, Editor One (2019) stated, “we warn and punish certain authors whose papers are identified [as] having plagiarism.” COPE focuses on problems of misconduct, conflicts of interest, and most importantly for my research, the peer review process. Tamara Welschot, a contact person for COPE, explained in great detail what COPE is and her role within the committee:

> So, we register or most publishers register journals with COPE. It’s to make everyone aware that we take research integrity or publication ethics very seriously, so that’s one step. So they take care of making sure these journals are registered with them and are available in their database so everyone can see which journals have been registered and essentially [this] means that all our journals follow the COPE guidelines and as a company our own policies are very much towards the COPE guidelines, or the Code of Practice I should say, and flow charts. I think they have really good discussions on lots of different topics. As publishers we are very much involved and every publisher has particular topics that they work with [in] COPE….I think last year we as a company

> worked with COPE on flow charts on guidelines on [management] on the submission process….I think we cannot go around COPE, I know some of the society's publishers actually have their own guidelines. They're probably a little more strict on their policies or instructions I should say. And that means that it is sometimes said that whenever you have a complaint COPE is not really the one saying, you know, to be a police officer or saying yes or no, they essentially look at whether the processes have been followed. In their [the editors] positions it is probably also the best thing they can do. They [COPE] are essentially not the police, the publisher is like a police officer or a judge. We have to follow due process or diligence to find out if something needs to be retracted or not. I think the whole focus of COPE is really making sure that research literature is, you know, that there is integrity in what is being published (TW, 2019).

4.12. Scanning Tool

Software tools have been developed over the years in the hopes of screening out CGPs. David Grier from IEEE has worked hands-on with such tools:

> What it is for us is several software tools that are combined and merged into managed systems and every paper [manuscript] that comes is scanned for a variety of things and against some other libraries and I believe the current version of it picks off the bibliography and scans against that (David Grier, 2019).

4.13. Conclusion

Computer generated papers may be becoming more popular among faculty researchers who are under pressure to publish, and this is a serious threat to academic integrity and publication ethics. Understanding the peer review process and how it could be modified could

help reduce the publication of CGPs: "It's the kind of thing that [editors] could ban together to do some work on problems such as this" (DG, 2019). When I asked one of the editors if he believed the CGP problem would end, he replied, "Will we be able to solve this? Maybe when we stop having people who have the wrong set of goals." (David Grier, 2019)

5. Conclusion & Recommendations

Understanding the ability of CGPs to pass through the peer review process undetected has been the most interesting part of this research. The peer review process requires that experts in a field review manuscripts submitted for presentations or publication. Journal editors believe and commonsense suggests that peer reviewers, as experts in their field, should spot invalid or nonsensical information when they review a manuscript. Therefore, I found it truly fascinating to see CGPs make it through peer review and be presented at professional conferences or be published in journals. Clearly there are still weaknesses in the peer review process.

It is well known that peer review is inconsistent and subjective, and too often operates as a sort of black box. To correct these deficiencies, various strategies have been suggested, such as training peer reviewers. However, studies of peer reviewer training have not shown it to be effective or significant in improving reviewer performance (Callaham, Wears, & Waeckerle, 1998, p. 318-322). This ineffectiveness could be caused by flaws or weaknesses in the training and does not mean that training could not be effective if well designed and implemented. As a black box, little is known about how journals select peer reviewers, and that is another area that is ripe for future research. Also, little seems to be known about the particular skills that reviewers need and use, and so that is another area meriting future research. Perhaps awareness alone would help solve the issue: for example, if peer reviewers were given a sample reading from a CGP, they might become better at spotting CGPs in the future.

During my investigation of CGPs, I found certain research approaches that worked and others that did not. The original plan was to use participant observation as a qualitative methodology. Through participant observation, I would be able to observe how editors or reviewers dealt with CGPs. After researching participant observation as a qualitative

methodology, however, I came to the conclusion that this method would not work for my study or provide the best outcome for this research, since it would be impractical for me to interact with and observe so many editors in such a short period of time.

I then decided on interview methodology, combined with thematic analysis. Interview methodology and thematic analysis together proved to be a successful approach to my research on CGPs. The interviews were conducted using both email text and through live conferencing (via Zoom). Based on the literature review, I developed 10 questions regarding CGPs and the peer review process. I found the most useful and nuanced information was given through conversations on Zoom. Compared with live face-to-face conferencing using Zoom, I found the information from email text interviews to be relatively brief and often superficial. Zoom proved to be the most successful tool for conducting my interviews because it enabled a true conversation. I was able to get the information I needed, engage in follow up discussion, and also change the minds of the editors from my research. This latter point is especially important: in the process of questioning editors about CGP, I could sometimes observe how the editor's understanding of CGP changed through the discursive practice of the interview, thereby shaping CGP reality.

As discussed in the "Thematic Analysis" section, the peer review process is not the only way to detect CGPs. Good progress has also been made with detection software. Although promising, CGP detection software will likely lead to a sort of arms race with CGP creation software. While detection programs will likely become standard practice for journal editors, the history of computer software shows that researchers cannot become complacent about such technological fixes. As Bohannon (2015, p. 18-19) warns, "I'm willing to bet if someone wanted

to declare an arms race, they could come up with another way to generate papers that would fool [SciDetect] again for a while."

The peer review process is a vital tool within academic research and publication. When researchers use published articles for their own research, they depend on the information to be factual. Typically, peer reviewers are (or should be) chosen based on their knowledge of a specific topic. So, the problem of CGPs getting past the peer review process is a key to understanding the issues in the peer review process, future research on this topic, and ultimately to solutions for this problem. Two editors, David Grier and Tamara Welschot, recommended that a training seminar for beginning peer reviewers would be the best way to improve the peer reviewer process. Clearly this training would need to be well-designed. As a final recommendation, I believe that the peer review process could be improved with in-depth training seminars that included hands-on exercises, the use of software detection tools, and case studies of real problems such as CGPs. My recommendations would probably not end the CGP problem, as new software will likely be developed that generates more sophisticated and harder-to-detect CGPs. However, although it may not permanently solve the CGP problem, we have to start somewhere.

6. References Cited

6.1. Interview References

Editor 1. Personal Interview. January 23, 2019

Editor 2. Personal Interview. January 30, 2019

Grier, David. Personal Interview. February 11, 2019

Jorge, Joaquim. Personal Interview. January 29, 2019

Welschot, Tamara. Personal Interview. February 20, 2019

6.2. Published Articles

"Researchers from University of Sao Paulo Report Recent Findings in General Science (Comparing the topological properties of real and artificially generated scientific manuscripts)." Science Letter, 25 Dec. 2015, p. 1042. Retrieved from http://go.galegroup.com.mtproxy.lib.umt.edu:3048/ps/i.do?&id=GALE|A438070697&v=2.1&u=mtlib_1_1092&it=r&p=AONE&sw=w&authCount=1

Aagaard, J. & N. Matthiesen. 2015. *Methods of Materiality: Participant Observation and Qualitative Research in Psychology.* Vol 13 Issue 1. Retrieved from https://www-tandfonline-com.mtproxy.lib.umt.edu:3443/doi/abs/10.1080/14780887.2015.1090510

Amancio, Diego Raphael. December 2015. "*Comparing the topological properties of real and artificially generated scientific manuscripts."* Springer Link. Vol. 105, issue 3, pp 1763-1779. Retrieved from https://link.springer.com/article/10.1007/s11192-015-1637-z

Anyan, F. 2013. "The Influence of Power Shifts in Data Collection and Analysis Stages: A Focus on Qualitative Research Interview." *The Qualitative Report*. Vol 18, 36, 1-9. Retrieved from https://files.eric.ed.gov/fulltext/EJ1005513.pdf

Artstein R., & M. Poesio. "Inter-Coder Agreement for Computational Linguistics". *Computational Linguistics*. Vol 34, Issue 4. Retrieved from https://www.mitpressjournals.org/doi/abs/10.1162/coli.07-034-R2

Attride-Stirling, J. 2001. "Thematic Networks. An Analytic Tool for Qualitative Research". *Qualitative Research.* Sage Journal. Vol 1. Issue 3. Retrieved from http://journals.sagepub.com/doi/abs/10.1177/146879410100100307

Bartoli, Alberto, Andrea De Lorenzo, Eric Medvet, and Fabiano Tarlao. "Your paper has been accepted, rejected, or whatever: Automatic generation of scientific paper reviews." In International Conference on Availability, Reliability, and Security, pp. 19-28. Springer, Cham, 2016. Retrieved from https://link.springer.com/chapter/10.1007/978-3-319-45507-5_2

Batane, Tshepo. 2010. "Turning to Turnitin to Fight Plagiarism among University Students". *Journal of Educational Technology & Society*. Vol. 13, No. 2 pp. 1-12. Retrieved from https://www.jstor.org/stable/pdf/jeductechsoci.13.2.1.pdf?seq=1#page_scan_tab_contents

Berger, A. 2016. *Media and Communication Research Methods. An Introduction to Qualitative and Quantitative Approaches*. 4th Edition. Sage Publishing.

Bohannon, John. "Who's Afraid of Peer Review?." (2013): 60-65. Retrieved from http://science.sciencemag.org/content/342/6154/60

Bohannon, John. April 2015. "Hoax-detecting software spots fake papers." *Science*. Vol 348, Issue 6230. Retrieved from http://science.sciencemag.org/content/348/6230/18

Burdick, Alan. 2017. "Paging Dr. Fraud: The Fake Publishers That are Ruining Science*". The New Yorker*. Retrieved from https://www.newyorker.com/tech/elements/paging-dr-fraud-the-fake-publishers-that-are-ruining-science

Butler, Declan. "Journals step up plagiarism policing." *Nature* (2010): 167-167. Retrieved from https://www.nature.com/news/2010/100705/full/466167a.html

Callaham, Michael L., and John Tercier. "The relationship of previous training and experience of journal peer reviewers to subsequent review quality." *PLoS medicine* 4, no. 1 (2007): e40

Callaham, Michael L., Robert L. Wears, and Joseph F. Waeckerle. "Effect of attendance at a training session on peer reviewer quality and performance." *Annals of emergency medicine* 32, no. 3 (1998): 318-322

Canagarajah A. 1996. "Nondiscursive Requirements in Academic Publishing, Material Resources of Periphery Scholars, and the Politics of Knowledge Production*." Written Communication*. Sage Journals. Vol 13. Issue 4. Retrieved from http://journals.sagepub.com/doi/abs/10.1177/0741088396013004001

Clarke, V., & V. Braun. 2017. “Thematic Analysis.” *The Journal of Positive Psychology*. 12:3, 297-298. Retrieved from https://www-tandfonline-com.mtproxy.lib.umt.edu:3443/doi/pdf/10.1080/17439760.2016.1262613?needAccess=true

Dadkhah, M., M. Kahani, & G. Borchardt. 2017. “A Method for Improving the Integrity of Peer Review”. *Science and Engineering Ethics.* SpringerLink. Vol. 24. Issue 5. Retrieved from https://link-springer-com.mtproxy.lib.umt.edu:3443/article/10.1007/s11948-017-9960-9

Dahl, Stephan. 2007. “The student perspective on using plagiarism detection software.*”* Turnitin. Retrieved from http://journals.sagepub.com/doi/abs/10.1177/1469787407074110

DeAndrea, David, Christopher Carpenter, Hillary Shulman, & Timothy Levine. 2009. “The Relationship Between Cheating Behavior and Sensation-Seeking.*” Personality and Individual Differences*. Elsevier. Vol 47 Issue 8. Retrieved from https://www-sciencedirect-com.mtproxy.lib.umt.edu:3443/science/article/pii/S0191886909003390

Devlin, Hannah. *“*Why some research is real rubbish*.”* The Times [London, England], *Business Insights: Essentials*. February 2014. Retrieved from http://bi.galegroup.com/essentials/article/GALE%7CA359884027/7b10700132e573b0982fb51c83669c2a?u=mtlib_1_1092

Dr. Hany Farid. November 2006. “Digital doctoring: How to tell the real from the fake.” *Significance.* The Royal Statistical Society. Vol. 3, Issue 4. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2006.00197.x/full

EBook Central. Retrieved from calendar.google.com

Edwards, Marc A., and Siddhartha Roy. "Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition." *Environmental Engineering Science.* 34, no. 1 (2017): 51-61. https://www.liebertpub.com/doi/abs/10.1089/ees.2016.0223

Elsevier. 2018. “What is Peer Review?” Retrieved from file:///F:/Thesis%20Project/What%20is%20Peer%20Review%20-%20Elsevier.pdf

Family Health International. N.D. “Qualitative Research Methods Overview.” *Qualitative Research Methods: A Data Collector’s Field Guide*. Retrieved from https://course.ccs.neu.edu/is4800sp12/resources/qualmethods.pdf

Fennell, Catriona, Ann Corner, & Elizabeth Ash. 2018.*"* Transparency – The Key to Trust in Peer Review.*"* Elsevier. Retrieved from file:///F:/Thesis%20Project/elsevier.com-Transparency%20%20the%20key%20to%20trust%20in%20peer%20review.pdf

Ferreira, Catarina, Guillaume Bastille-Rousseau, Amanda M. Bennett, E. Hance Ellington, Christine Terwissen, Cayla Austin, Adrian Borlestean et al. "The evolution of peer review as a basis for scientific publication: directional selection towards a robust discipline?." *Biological Reviews* 91, no. 3 (2016): 597-610. https://onlinelibrary.wiley.com/doi/abs/10.1111/brv.12185

Flick, U. 2002. "An Introduction to Qualitative Research." *The Expert Interview*. 2nd Edition. SAGE Publications.

Gelbukh A. 2007. "Computational Linguistics and Intelligent Text Processing.*"* *Computer Science*, Springer. Vol 4394. Retrieved from https://link-springer-com.mtproxy.lib.umt.edu:3443/book/10.1007%2F978-3-540-70939-8

Ginsparg, Paul. April 2014. "Automated screening: ArXiv screens spot fake papers." *Nature International Weekly Journal of Science*. Vol. 508, Issue 7494. Retrieved from https://www.nature.com/nature/journal/v508/n7494/full/508044a.html

Godwin-Jones, R. "Emerging Technologies. Messaging, Gaming, Peer-to-peer Sharing: Language Learning Strategies & Tools for the Millennial Generation". *Language Learning & Technology*. Vol 9. Retrieved from https://scholarspace.manoa.hawaii.edu/bitstream/10125/44003/1/09_01_emerging.pdf

Griffin, Donovan. 2015. "Weeding Out Fake Science". *Information Today.* Vol 32 Issue 4. Retrieved from http://web.a.ebscohost.com.mtproxy.lib.umt.edu:3048/ehost/detail/detail?vid=0&sid=74a7fc42-20e3-4930-afd8-fa447a2192a9%40sessionmgr4007&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d#AN=102934222&db=a9h

Grimes, David Robert, Chris T. Bauch, and John PA Ioannidis. "Modelling science trustworthiness under publish or perish pressure.*"* *Royal Society Open Science* 5, no. 1 (2018): 171511

Haider, Jutta, and Fredrik Åström. "Dimensions of trust in scholarly communication: Problematizing peer review in the aftermath of John Bohannon's “Sting” in science." *Journal of the Association for Information Science and Technology* 68, no. 2 (2017): 450-467. https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.23669

iThenticate Paper. 2013. “Rising Tide of Plagiarism and Misconduct in Medical Research”. iThenticate Professional Plagiarism Prevention.

iThenticate Report. 2012. “True Costs of Research Misconduct”. iThenticate Professional Plagiarism Prevention.

Jacob E. 1988. “Clarifying Qualitative Research: A Focus on Traditions.” Vol 17 Issue 1. Retrieved from http://journals.sagepub.com.mtproxy.lib.umt.edu:3048/doi/abs/10.3102/0013189X017001016

Jones, Michael & Lynnaire Sheridan. August 2014. *“*Back Translation: An Emerging Sophisticated Cyber Strategy to Subvert Advances in ‘Digital Age’ Plagiarism Detection and Prevention”. *Assessment & Evaluation in Higher Education*. Doi.org/10.1080/02602938.2014.950553

Jorgensen, D. 2015. *Participant Observation. Methods of Research.* Retrieved from https://doi.org/10.1002/9781118900772.etrds0247

Josien, Laurent, and Britton Broderick. 2013. "Cheating in higher education: the case of multi-methods cheaters." *Academy of Educational Leadership Journal* 17.2 Business Insights: Essentials. Retrieved from http://bi.galegroup.com.mtproxy.lib.umt.edu:3048/essentials/article/GALE%7CA332788076/ddef71d5157d836d4e4fff5ea60cdebc?u=mtlib_1_1092

Kakaes, Konstantin. February 2014. “How Gobbledygood Ended up in Respected Scientific Journals*.” Slate*. Retrieved from http://www.slate.com/blogs/future_tense/2014/02/27/how_nonsense_papers_ended_up_in_respected_scientific_journals.html

Kalnins, K. Halm & M. Castillo. 2015. “Screening for Self-Plagiarism in a Subspecialty-versus-General Imaging Journal Using iThenticate”. *AJNR American Journal of Neuroradiology*. Vol 36 Issue 6. DOI: https://doi.org/10.3174/ajnr.A4234 . Retrieved from http://www.ajnr.org/content/36/6/1034

Kassirer, Jerome & Edward Campion. 1994. “Peer review: Crude & Understudied, but Indispensable.” Vol 272, No. 2. Retrieved from doi:10.1001/jama.1994.03520020022005

Kerner, Yaakov HaCohen, Aharon Tayeb, Natan Ben Dror. 2010. Detection of Simple Plagiarism in Computer Science Papers. *Internation Conference on Computational*

Labbé, Cyril & Dominique Labbé. “Duplicate and fake publications in the scientific literature: how many SCIgen papers in computer science?” *Scientometrics.* Springer Link. January 2013. Vol. 94, No. 1, pp 379-396. Retrieved from https://link.springer.com/article/10.1007/s11192-012-0781-y

Labbé, Cyril. "Trends in gaming indicators: On failed attempts at deception and their computerised detection." In 7th International Workshop on Bibliometric-enhanced Information Retrieval (BIR 2018) co-located with the 40th European Conference on Information Retrieval (ECIR 2018), Grenoble.2018

Lang, James. 2013. “Cheating Lessons: Learning from Academic Dishonesty.*”*

Latour, Bruno, and Steve Woolgar. 2013. “Laboratory life: The construction of scientific facts.” Princeton University Press.

Lavoie, Allen & Krishnamoorthy, Mukkai. 2010. *Algorithmic Detection of Computer Generated Text*. Cornell University Library. Retrieved from https://arxiv.org/abs/1008.0706

Lee, Samuel. 2008. *Cheating in Scientific Publishing: The Scourge Must Stop!* Liver International. Vol 28 issue 5. Retrieved from https://onlinelibrary-wiley-com.mtproxy.lib.umt.edu:3443/doi/full/10.1111/j.1478-3231.2008.01753.x

Linguistics. Retrieved from http://www.aclweb.org/anthology/C10-1048

Liu, Shaoying. 2010. “An Approach to Applying SOFL for Agile Process and Its Application in Developing a Test Support Tool.” *Innovations in Systems and Software Engineering.* Springer. Vol 6. Issue 1-2. Retrieved from https://link-springer-com.mtproxy.lib.umt.edu:3443/article/10.1007/s11334-009-0114-3

Marsh, Herbert & Samuel Ball. 2014. “The Peer Review Process Used to Evaluate Manuscripts Submitted to Academic Journals.” Interjudgmental Reliability. *The Journal of Experimental Education.* Retrieved from https://doi.org/10.1080/00220973.1989.10806503

Marušić, A. "Problems of editors with authorship in small medical journals." *The International Journal of Occupational and Environmental Medicine* 2, no. 3 July (2011). Retrieved from http://www.theijoem.com/ijoem/index.php/ijoem/article/view/97/190

Masse, Bryson. 2017. “There Are Now 8,000 Fake Science ‘Journals’ Worldwide, Researchers Say. Researchers from High-Profile Institutions are Falling for These Scams.” *MotherBoard*. Retrieved from https://motherboard.vice.com/en_us/article/xwwb8n/there-are-now-8000-fake-science-journals-worldwide-researchers-say

McArthur, Tom, Lam-McArthur, Jacqueline & Lise Fontaine. “Computational Linguistics. The Oxford Companion to the English Language.” Retrieved from http://www.oxfordreference.com.mtproxy.lib.umt.edu:3048/view/10.1093/acref/9780199661282.001.0001/acref-9780199661282-e-1361

McGuigan G., & R Russel. 2008. “The Business of Academic Publishing: A Strategic Analysis of the Academic Journal Publishing Industry and its Impact on the Future of Scholarly Publishing*.” Electronic Journal of Academic and Special Librarianship.* Vol 9 No. 3. Retrieved from http://southernlibrarianship.icaap.org/content/v09n03/mcguigan_g01.html

McNutt, Robert A., Arthur T. Evans, Robert H. Fletcher, et al. 1990. *“*The Effects of Blinding on the Quality of Peer Review: A Randomized Trial.” Vol 263, pg 1371-1376. Retrieved from https://jamanetwork.com/journals/jama/article-abstract/380957

Mehmani, Bahar. 2016. “Is Open Peer Review the Way Forward?” Elsevier. Retrieved from file:///F:/Thesis%20Project/elsevier.com-Is%20open%20peer%20review%20the%20way%20forward.pdf

Nagoya, F., S Liu. N.d. “Development of a web-based conference management system using SOFL*.”* ACM DL Digital Library. Doi: 10.1145/2811411.2811502 Retrieved from https://dl-acm-org.mtproxy.lib.umt.edu:3443/citation.cfm?doid=2811411.2811502

Pei-Yu Lin, Ran-Zan Wang, Yu-Jie Chang & Wen-Pinn Fang. 2015. “Prevention of Cheating in Visual Cryptography by Using Coherent Patterns.” *Information Sciences.* Volume 301. Retrieved from https://www-sciencedirect-com.mtproxy.lib.umt.edu:3443/science/article/pii/S0020025514011980

Pickover, Clifford. 1998. “A Computer Graphical Journey: Ten Year Compilation of Advanced Research.” *Chaos and Fractals*. Elsevier Science & Technology. Retrieved from https://ebookcentral-proquest-com.mtproxy.lib.umt.edu:3443/lib/umontana/detail.action?docID=29905

Pillai, Mini G. 2015. “Experiences of University Libraries in Kerala with anti-plagiarism softwares iThenticate and Turnitin. Blended libraries and information centres: a blueprint for the development of information profession in India.” *Southern Book Star*, Thiruvananthapuram, pp. 279-291. Retrieved from http://eprints.rclis.org/25209/

Pisanski, K., P. Sorokowski, & E. Kulczycki. 2017. “Predatory Journals Recruit Fake Editor.” Nature. https://www.nature.com/polopoly_fs/1.21662.1490714371!/menu/main/topColumns/topLeftColumn/pdf/543481a.pdf?origin=ppub

Preiss, M., H. Klein, N. Levenburg, A. Nohavova. 2013. “A Cross Country Evaluation of Cheating in Academia – A Comparison of Data from the US and the Czech Republic.” *Journal of Academic Ethics*. Vol 11 Issue 2. Retrieved from https://link-springer-com.mtproxy.lib.umt.edu:3443/article/10.1007/s10805-013-9179-6

Randell-Moon, Holly et al. 2011. “Journal editing and ethical research practice: perspectives of journal editors.” *Ethics and Education*. Vol 6 Issue 3. Retrieved from https://www-tandfonline-com.mtproxy.lib.umt.edu:3443/doi/abs/10.1080/17449642.2011.632717

Roig, Miguel. "Journal editorials on plagiarism: what is the message?." Eur Sci Ed 40 (2014): 58-59. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5334150/

Sample, Ian. “How computer-generate fake papers are flooding academia.” *The Guardian.* February 2014. Retrieved from https://www.theguardian.com/technology/shortcuts/2014/feb/26/how-computer-generated-fake-papers-flooding-academia

Schroter, S., N. Black, S. Evans, F. Godlee, L. Osorio, and R. Smith. 2008. *What errors do peer reviewers detect, and does training improve their ability to detect them*?. Journal of the Royal Society of Medicine, 101(10), pp.507-514

Schroter, Sara, Leanne Tite, and Richard Smith. *"Perceptions of open-access publishing: interviews with journal authors.*" Bmj 330, no. 7494 (2005): 756. https://www.bmj.com/content/330/7494/756.short

Schroter, Sara, Nick Black, Stephen Evans, James Carpenter, Fiona Godlee, and Richard Smith. "*Effects of training on quality of peer review: randomised controlled trial.*" Bmj 328, no. 7441 (2004): 673

Shyam, Ashok. 2015. “Predatory Journals: What are they?” *Journal of Orthopaedic Case Reports*. Doi: 10.13107/jocr.2250-0685.330. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4845443/

Smith, Richard. 2006. “Peer review: A flawed Process at the Heart of Science and Journals.” *Journal of the Royal Society of Medicine*. Volume 99, pg 178-182. Retrieved from http://journals.sagepub.com/doi/pdf/10.1177/014107680609900414

Souder, Lawrence. "The ethics of scholarly peer review: a review of the literature." *Learned Publishing* 24, no. 1 (2011): 55-72. Retrieved from https://onlinelibrary.wiley.com/doi/pdf/10.1087/20110109

Spier, Ray. 2002. “The History of the Peer-Review Process.” *Trends in Biotechnology*. Vol 20, Issue 8, Pg 357-358. Retrieved from https://doi.org/10.1016/S0167-7799(02)01985-6

Spinak, Ernesto. March 2014. “In the beginning it was just plagiarism-now its computer generated fake papers as well.” *SciELO in Perspetive*. Retrieved from http://blog.scielo.org/en/2014/03/31/in-the-beginning-it-was-just-plagiarism-now-its-computer-generated-fake-papers-as-well/#.WdPuDhNSx0t

Stewart, T. 2002. “Principles of Research in Communication. Interviewing Procedures.” Allyn and Bacon.

Tancock, C. 2018. “When reviewing goes wrong: the ugly side of peer review.” Elsevier. Retrieved from https://www.elsevier.com/connect/editors-update/when-reviewing-goes-wrong-the-ugly-side-of-peer-review

Teixeira, Auorora. 2008. “Academic Cheating in Spain and Portugal: An Empirical Explanation.” *International Journal of Iberian Studies*. Vol 21 Issue 1. Retrieved from http://web.a.ebscohost.com.mtproxy.lib.umt.edu:3048/ehost/detail/detail?vid=0&sid=a77f8647-05e6-4193-95c5-963377bb3e38%40sessionmgr4007&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d#AN=32549093&db=a9h

Tenopir, Carol, et al. 2009. “Electronic Journals and Changes in Scholarly Article Seeking and Reading Patterns.” Aslib Proceedings. Vol 61 Issue 1. Retrieved from https://www.emeraldinsight.com/doi/abs/10.1108/00012530910932267

The Writing Center. 2018. *Academic Integrity*. Retrieved from https://writingcenter.unc.edu/esl/resources/academic-integrity/

Van Noorden, Richard. “Publishers withdraw more than 120 gibberish papers.” Nature, *International Weekly Journal of Science.* February 2014. doi:10.1038/nature.2014.14763 https://www.nature.com/news/publishers-withdraw-more-than-120-gibberish-papers-1.14763

Vines, Timothy. April 2014. ”Scientific Community: Journals must Boost Data Sharing.” *International Weekly Journal of Science*. Vol. 508, Issue 7494. Retrieved from https://www.nature.com/nature/journal/v508/n7494/full/508044c.html

Wagner, Elizabeth, Suzan Fiack, Chris Graf, Andrew Robinson, and Ian Rowlands. "Science journal editors’ views on publication ethics: results of an international survey." *Journal of medical ethics* 35, no. 6 (2009): 348-353. Retrieved from https://jme.bmj.com/content/35/6/348

Wicherts, Jelte M. "Peer review quality and transparency of the peer-review process in open-access and subscription journals." PloS one 11, no. 1 (2016): e0147913. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0147913

Wiener-Bronner, Danielle. March 2014. “On the Exploration of Virtual Machines- More Computer-Generated Nonsense Papers Pulled from Science Journals”. *The Atlantic*. Retrieved from https://www.theatlantic.com/technology/archive/2014/03/more-computer-generated-nonsense-papers-pulled-science-journals/358735/

Williams, Kyle, and C. Lee Giles. "On the use of similarity search to detect fake scientific papers." *In International Conference on Similarity Search and Applications*, pp. 332-338. Springer, Cham, 2015. Retrieved from https://link.springer.com/chapter/10.1007/978-3-319-25087-8_32

Williams, Peter, and Elizabeth Wager. "Exploring why and how journal editors retract articles: findings from a qualitative study." *Science and engineering ethics* 19, no. 1 (2013): 1-11. Retrieved from https://www.researchgate.net/publication/51494585_Exploring_Why_and_How_Journal_Editors_Retract_Articles_Findings_From_a_Qualitative_Study

Xin-xin Zhang, Zhao-lin Huo & Yue-hong Zhang. April 2013. “Detecting and (not) Dealing with Plagiarism in an Engineering Paper: Beyond CrossCheck – A Case Study”. Science & Engineering Ethics – Springer Science and Media B.V. DOI 10.1007/s11948-013-9460-5

Yehuda Peled, et al. 2018. “Predictors of Academic Dishonesty Among Undergraduate Students in Online and Face-to-face Courses.” *Computers & Education*. Retrieved from https://www-sciencedirect-com.mtproxy.lib.umt.edu:3443/science/article/pii/S036013151830112X

Younghwa Lee. 2011. “Understanding anti-plagiarism software adoption: An extended protection motivation theory perspective.” *Decision Support Systems*. Elsevier. Vol 50 Issue 2. doi.org/10.1016/j.dss.2010.07.009. Retrieved from https://www-sciencedirect-com.mtproxy.lib.umt.edu:3443/science/article/pii/S0167923610001156

Zhang Yuehong & Jia Xiaoyan. 2012. “A Survey on the Use of CrossCheck for Detecting Plagiarism in Journal Articles.” *Learned Publishing.* Vol 25 Issue 4. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1087/20120408

7. Appendices

7.1. Appendix A: IRB Approval Form

Form RA-108
(Rev. 09/16)

THE UNIVERSITY OF MONTANA-MISSOULA
Institutional Review Board (IRB)
for the Protection of Human Subjects in Research
APPLICATION FOR IRB REVIEW

IRB Protocol No.: 6-19

At the University of Montana (UM), the Institutional Review Board (IRB) is the institutional review body responsible for oversight of all research activities involving human subjects as outlined in the U.S. Department of Health and Human Services' Office of Human Research Protection and the National Institutes of Health, Inclusion of Children Policy Implementation.

Instructions: A separate application must be submitted for each project. IRB proposals are approved for no longer than one year and must be continued annually (unless Exempt). Faculty and students may email the completed form as a Word document to IRB@umontana.edu, or submit a hardcopy (no staples) to the IRB office in the Interdisciplinary Sciences Building, room 104. Student applications must be accompanied by email authorization by the supervising faculty member or a signed hard copy. *All fields must be completed. If an item does not apply to this project, write in: N/A.* Questions? Call the IRB office at 243-6672.

1. **Administrative Information**

Project Title: Computer Generated Papers as a New Plagiarism Challenge in Communication	
Principal Investigator: Malea Dunne	UM Position: Graduate Student
Department: Technical Communication	Office location: Eng 103B (MTech)
Work Phone: (406) 496-4461	Cell Phone:

2. **Human Subjects Protection Training** *(All researchers, including faculty supervisors for student projects, must have completed a self-study course on protection of human research subjects within the last three years and be able to supply the "Certificate(s) of Completion" upon request. If you need to add rows for more people, use the Additional Researchers Addendum.)*

All Research Team Members (List yourself first)	PI	CO-PI	Faculty Supervisor	Research Assistant	DATE COMPLETED (IRB-approved Course mm/dd/yyyy)
Name: Malea Dunne Email: mdunne1@mtech.edu	☒	☐	☐	☐	11/09/2017
Name: Pat Munday Email: pmunday@mtech.edu	☐	☐	☒	☐	10/11/2016
Name: Email:	☐	☐	☐	☐	
Name: Email:	☐	☐	☐	☐	

3. **Project Funding** *(If federally funded, you must submit a copy of the abstract or Statement of Work.)*

Is grant application currently under review at a grant funding agency? ☐Yes *(If yes, cite sponsor on ICF if applicable)* ☒No			Has grant proposal received approval and funding? ☐Yes *(If yes, cite sponsor on ICF if applicable)* ☒No		
Agency	Grant No.	e-Prop #	Start Date	End Date	PI on grant

For UM IRB Use Only

IRB Determination:

___ Not Human Subjects Research
✓ Approved by Exempt Review, Category # 6, 2 *(see memo)*
___ Approved by Expedited Review, Category # ____ *(see Note to PI)*
Full IRB Determination
___ Approved *(see Note to PI)*
___ Conditional Approval *(see memo)* - IRB Chair Signature/Date: _____
___ Conditions Met *(see Note to PI)*
___ Resubmit Proposal *(see memo)*
___ Disapproved *(see memo)*

Risk Level: Minimal

Note to PI: Non-exempt studies are approved for one year only. Use any attached IRB-approved forms (signed/dated) as "masters" when preparing copies. If continuing beyond the expiration date, a continuation report must be submitted. Notify the IRB if any significant changes or unanticipated events occur. When the study is completed, a closure report must be submitted. Failure to follow these directions constitutes non-compliance with UM policy.

Final Approval by IRB Chair/Manager: _____ Date: 1/14/19 Expires: NA

UNIVERSITY OF MONTANA

INSTITUTIONAL REVIEW BOARD
for the Protection of Human Subjects in Research
FWA 00000078
Research & Creative Scholarship
University Hall 116
University of Montana
Missoula, MT 59812
Phone 406-243-6672 | Fax 406-243-6330

Date: January 14, 2019

To: Malea Dunne, Technical Communication, Montana Tech

From: Scott Risser, Montana Tech IRB Representative

RE: IRB #6-19: "Computer Generated Papers as a New Plagiarism Challenge in Communication"

Your IRB proposal cited above has been **APPROVED** under the **Exempt** category of review by the Institutional Review Board in accordance with the Code of Federal Regulations, Part 46, section 101. The specific paragraph which applies to your research is:

__X__ (b)(2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (i) Information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.

University of Montana IRB policy does not require you to file an annual Continuation Report for exempt studies, as there is no expiration date on the approval. However, you are required to notify the IRB of the following:

Amendments: Any changes to the originally-approved protocol must be reviewed and approved by the IRB ***before*** being made (unless extremely minor). Requests must be submitted using Form RA-110.

Unanticipated or Adverse Events: You are required to timely notify the IRB if any unanticipated or adverse events occur during the study, if you experience an increased risk to the participants, or if you have participants withdraw from the study or register complaints about the study. Use Form RA-111.

Please contact the IRB office with any questions at (406) 243-6672 or email irb@umontana.edu.

7.2. Appendix B: Interview Information: Participants, Dates, Responses

Graduate Thesis Interviews - CGP Journal Interview									
JOURNAL NAME	EDITOR	EDITOR EMAIL	EMAIL DATE	FOLLOW UP DATE	RESPONSE	INTERVIEW DATE	EMAILED RESPONSES OR INTERVIEW	ANOMYNITY NUMBER	NOTES
Nature	Rebecca (Bex) Walton	r.walton@nature.com	1/9/19	1/23/19	XXX	XXX	XXX	XXX	Suggested Tamara Welschot from Springer Nature
Science	Jeremy Berg	jberg@aaas.org	1/9/19	XXX	Received failed delivery 1/11/2019	XXX	XXX	XXX	XXX
SoftwareX (Elsiever)	Dr. Kate Keahey	keahey@anl.gov	1/14/19	1/23/19	NONE	XXX	XXX	XXX	XXX
The Computer Journal (Oxford University Press)	Stephen Furber	compj@bq.bcs.org.uk	1/9/19	1/23/19	NONE	XXX	XXX	XXX	XXX
Network Science (Cambridge University Press	Ronald Breiger	breiger@email.Arizona.edu	1/9/19	1/23/19	NONE	XXX	XXX	XXX	XXX
Computer (IEEE)	David Grier	grier@email.gwu.edu	1/14/19	XXX	AGREED	2/11/19	INTERVIEW	XXX	Also suggested Bill Hagen. Zoom interview & transcription completed
Automated Software Engineering (Springer)	Robert Hall	bob.ASEJ@gmail.com	1/14/19	1/23/19	NONE	XXX	XXX	XXX	XXX
			1/23/19	XXX	AGREED	XXX	EMAILED	E1	Provided answers to questions. Wants to remain anonymous
Computer Architecture Letters (IEEE)	Daniel Sorin	sorin@ee.duke.edu	1/23/19	2/5/19	NONE	XXX	XXX	XXX	XXX
Technology & Society Magazine (IEEE)	Jeremy Pitt	j.pitt@imperial.ac.uk	1/23/19	2/5/19	NONE	XXX	XXX	XXX	XXX
Computer Graphics & Applications (IEEE)	Torsten Moller	torsten.moeller@univie.ac.at	1/23/19	XXX	DECLINED	XXX	XXX	XXX	Suggested Werner Purgathofer
Professor - Institute of Visual Computing & Human-Centered Technology	Werner Purgathofer	wp@cg.tuwien.ac.at	1/23/19	XXX	DECLINED	XXX	XXX	XXX	Suggested Eduard Groller, Nadia Magnenat-Thalmann, Joaquim Jorge, Min Chen
Professor - Institute of Visual Computing & Human-Centered Technology	Eduard Groller	groeller@cg.tuwien.ac.at	1/28/19	XXX	DECLINED	XXX	XXX	XXX	Suggested finding other editors; Bedrich Benes
Professor at MIRALab	Nadia Magnenat-Thalmann	visualcomputer@miralab.ch	1/28/19	2/5/19	NONE	XXX	XXX	XXX	XXX
Computers and Graphics Journal (Elsevier)	Joaquim Jorge	computers-and-graphics@inesc-id.pt	1/28/19	XXX	AGREED	1/29/19	EMAILED	XXX	Emailed interview responses. Can use name and Journal
Professor at Uni of Oxford	Min Chen (CGF)	minchencgf@gmail.com	1/28/19	2/5/19	NONE	XXX	XXX	XXX	XXX
			1/30/19	XXX	AGREED	1/30/19	EMAILED	E2	Emailed interview responses. Wishes to remain ANAONYMOUS.
Intellectual Property Rights Manager of IEEE	Bill Hagen	w.hagen@ieee.org	1/30/19	2/7/19	NONE	XXX	XXX	XXX	XXX
Springer Nature's Research Integrity Group	Tamara Welschot	tamara.welschot@springernature.com	1/31/19	XXX	AGREED	2/20/19	INTERVIEW		Zoom interview completed.

7.3. Appendix C: Interview Questions

Interview Questions:

1. Are you aware of Computer Generated Papers (CGP)?
2. If so, how did you become aware of CGP?
3. Do you see an issue with CGP and journal publications?
4. If yes, what are the specific issues with CGPs and academic journals?
5. Are there opportunities for improvement in the peer review process? (As a connected follow-up, should they not mention CGPs, ask if they have changed the peer review process to better identify CGPs.)
6. If yes, what strategies do you recommend for improving the peer review process?
7. As an editor, do you work at all with Committee on Publication Ethics (COPE)?
8. If you were to find out that your journal published a CGP, what would the retraction process be?
9. Is there anything else you wish to add regarding the issue of CGPs, peer review, and publication ethics?

7.4. Appendix D: Interview Request

Dear *Editor Name,*

My name is Malea Dunne and I am a Graduate Student in the field of Technical Communication at Montana Tech in Butte Montana.

My graduate thesis focuses on Computer Generated Papers (CGPs) as a problem in academic journal publication. Specifically, I am looking into the peer review process, where it might be flawed, and what improvements could be made when it comes to Computer Generated Papers and academic publications. For my thesis research, I am conducting interviews with journal editors.

Would you speak with me via telephone or Skype about CGPs and the peer review process? I have a list of questions I am more than happy to share with you before the interview. If you cannot speak with me, would you be willing to fill out my questions via email?

If you are willing to proceed, may I use your name and your journal's name in my thesis? If not, may I quote you as anonymous?

Proceeding with the interview process implies your consent for me to use your information. As stated, I will protect your identity, and the identity of your journal, at your request.

I look forward to hearing back from you, and I plan to begin the interviews January 21, 2019.

Thank you,

Malea Dunne
MSTC Student – Montana Tech of the University of Montana
4065603736
Mdunne1@mtech.edu
Malea.dunne@gmail.com

7.5. Appendix E.1: Interview Transcription (Email) Joaquim Jorge

Interview (Emailed) with Joaquim Jorge – Computers and Graphics Journal Elsevier
1/29/2019

MD: Malea Dunne (Researcher)
J.J: Joaquim Jorge (Editor)

M.D: **Are you aware of Computer Generated Papers (CGP)?**

J.J: "Yes"

M.D: **If so, how did you become aware of CGP?**

J.J: "I became aware of these through the infamous SCIGEN MIT Paper Generator that was used to produce a paper accepted at a conference (SCI) without a peer-review system. There is a nice article on this, that you may be aware of, published in 2015 http://news.mit.edu/2015/how-three-mit-students-fooled-scientific-journals-0414

M.D: **Do you see an issue with CGP and journal publications?**

J.J: "The way I see it, a Journal with a decent peer-review system should be able to spot SCIGEN generated papers. However, SCIGEN may only be the precursor of more sophisticated approaches, that may become progressively harder to unearth"

M.D: **If yes, what are the specific issues with CGP's and academic journals?**

J.J: "In recent years, SCIGEN and other (semi)automatic approaches to generating nonsensical papers that would not resist anything deeper than a cursory examination have been used to identify predatory publishers. Unfortunately, the business of predatory publishing (printing anything that remotely looks like a scholarly paper for a fee) has been flourishing due to the pressure to publish for academic promotion and tenure. I have been approached by paper mills (that is whole enterprises in the business of publish-for-hire) who offered to pay me to publish whatever papers they would produce. SCIGEN has shown the way to produce more and more sophisticated approaches to fabricate realistic-looking academic papers. There have been pre-SCIGEN attempts to discredit Humanities and Arts Scholarly Journals by sending human-generated gibberish papers. There is an infamous fake paper published by physicist Alan Sokal of New York University in the journal Social Text in 1996. Other human-generated fake papers include

Interview (Emailed) with Joaquim Jorge – Computers and Graphics Journal Elsevier
1/29/2019

MD: Malea Dunne (Researcher)
J.J: Joaquim Jorge (Editor)

the VIDEA incident in 1995 where Prof Purgathofer's group sent half a dozen nonsensical abstracts that were accepted without reviewing. Since you mention his name, I am sure you have all the relevant details about this.
There is also this interesting piece published by Nature News a few years ago:
https://www.nature.com/news/publishers-withdraw-more-than-120-gibberish-papers-1.14763"

M.D: **Are there opportunities of improvement in the peer review process? (As a connected follow-up, should they not mention CGPs, ask if they have changed the peer review process to better identify CGPs)**

J.J: "Yes, we are using more and more automated tools to detect plagiarism. Fake and plagiarized papers need to be fought with software tools. Human Reviewers are becoming more and more overworked due to the proliferation of Papers, Conferences and Journals. It is important to use automation to separate the wheat from the shaff before asking human reviewers to assess legit submissions. I believe that with the current state of the art, competent human reviewers should be able to distinguish legit from CGP submissions, at least in Scientific Fields."

M.D: **If yes, what strategies do you recommend for improving the peer review process?**

J.J: "See above. I currently reject about 20% of all submitted papers without sending them to reviewers for various reasons. The main one being out-of-scope papers. Reviewer fatigue is a serious problem, that can lead to bad papers being accepted."

M.D: **As an editor, do you work at all with Committee on Publication Ethics (COPE)?**

J.J: "Yes. I represent Elsevier multimedia publications in COPE

Interview (Emailed) with Joaquim Jorge – Computers and Graphics Journal Elsevier
1/29/2019

MD: Malea Dunne (Researcher)
J.J: Joaquim Jorge (Editor)

M.D: **If you were to find out that your journal published a CGP, what would the retraction process be?**

J.J: "This can possibly be the most embarrassing event to happen to an Editor. The text would have to be retracted As Soon As Detected, identifying the paper as a Fake Manuscript, with a Personal Note of Apology from the Editor-in-Chief. Who would have to assume the blame at a personal level."

M.D: **Is there anything else you wish to add regarding the issue of CGPs, peer review, and publication ethics?**

J.J: "I see CGPs happening as a consequence of the pressure to publish. The submission of large quantities of low-quality manuscripts is motivated by a numbers-focused quantity-over-quality evaluation process for untenured faculty. I see this in the context of a plethora of shady practices that include paper mills, predatory publishers and revenue-oriented conferences. The root causes need to be addressed to solve what is otherwise a very human problem. Until that happens I could see further and more capable strains of text generators appearing. Indeed there has been considerable research on computer-generated novels and poetry. Look at the high-profile retractions of manuscripts with fabricated results from Nature, Lancet and Science. Having a program to assist in Fabrication is just another way to make the process easier."

J.J: "21 Years ago, a paper with fabricated results was published in the Lancet, a very prestigious Medical Journal:
https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(10)60175-4/fulltext
This is the infamous paper that the VAXXER movement uses to claim that Vaccines cause autism. The results, method and even ethics approval were all fabricated.
The paper has since been retracted and its author lost all credentials. yet, many thousands if not millions of people have died because of credulous parents not vaccinating their children.
Fabricated papers detract from the virtues of science and should be identified and retracted as bad science can kill millions of people."

7.6. Appendix E.2: Anonymous Interview Transcription (Email) E1

ANONYMOUS (Email) Interview 1

MD: Malea Dunne (Researcher)
E1: Anonymous Interview 1

1	**M.D:**	**Are you aware of Computer Generated Papers (CGP)?**
2		
3	E1:	"I don't know that much"
4		
5	**M.D:**	**If so, how did you become aware of CGP?**
6		
7	E1:	"I have heard this before"
8		
9	**M.D:**	**Do you see an issue with CGP and journal publications?**
10		
11	E1:	"I don't think it is common in science and engineering publications"
12		
13	**M.D:**	**If yes, what are the specific issues with CGP's and academic**
14		**journals?**
15		
16	E1:	"If it happens, it is a very serious problem"
17		
18	**M.D:**	**Are there opportunities of improvement in the peer review process?**
19		**(As a connected follow-up, should they not mention CGPs, ask if**
20		**they have changed the peer review process to better identify CGPs)**
21		
22	E1:	"Nowadays there are software detectors for similarities to identify
23		plagiarism. This process is improving"
24		
25	**M.D:**	**If yes, what strategies do you recommend for improving the peer**
26		**review process?**
27		
28	E1:	"Not very specific"
29		

ANONYMOUS (Email) Interview 1

MD: Malea Dunne (Researcher)
E1: Anonymous Interview 1

M.D: **As an editor, do you work at all with Committee on Publication Ethics (COPE)?**

E1: "Yes, we warn and punish certain authors whose papers are identified having plagiarism"

M.D: **If you were to find out that your journal published a CGP, what would the retraction process be?**

E1: "We haven't done that in the past, but I think we should publish the findings"

M.D: **Is there anything else you wish to add regarding the issue of CGPs, peer review, and publication ethics?**

E1: "Not very specific"

7.7. Appendix E.3: Anonymous Interview Transcription (Email) E2

ANONYMOUS (Email) Interview 2

MD: Malea Dunne (Researcher)
E2: Anonymous Interview 2

M.D: **Are you aware of Computer Generated Papers (CGP)?**

E2: "Yes"

M.D: **If so, how did you become aware of CGP?**

E2: "A colleague told me about it many years ago. He actually tricked me to read one. It took me three paragraphs to notice it was a fake paper."

M.D: **Do you see an issue with CGP and journal publications?**

E2: "Not really for high quality publications. In fact, it is bitter for low-quality journals and venues that they can be fooled this way. It shows that they are not rigorous."

M.D: **If yes, what are the specific issues with CGP's and academic journals?**

E2: "As long as the editors and the reviewers do diligently their work, it should not be a problem."

M.D: **Are there opportunities of improvement in the peer review process? (As a connected follow-up, should they not mention CGPs, ask if they have changed the peer review process to better identify CGPs)**

E2: "There are always ways to improve. The reviewers need to read the paper carefully, they need to be experts in the field, and they should use some tools, such as iThenticate. I can also imagine some AI tool to quickly discover if the paper has been automatically generated."

M.D: **If yes, what strategies do you recommend for improving the peer review process?**

E2: "See Number 5 (Above)"

M.D: **As an editor, do you work at all with Committee on Publication Ethics (COPE)?**

E2: "Yes I do"

ANONYMOUS (Email) Interview 2

MD: Malea Dunne (Researcher)
E2: Anonymous Interview 2

M.D: **If you were to find out that your journal published a CGP, what would the retraction process be?**

E2: "I think for our journal it is almost impossible. We really read each paper many times. However, if this would happen, we would start a formal investigation with the Ethic committee. If the result would be positive, we would recommend actions. First, we would inform the authors and ask them for a reaction. If the reaction would be unsatisfactory, the journal would a) inform the authors about the final decision b) retract the paper with a mention that it was retracted, with full names of the authors, their affiliation, and the name of the paper c) depending if this was intent or an error, we would also inform the direct supervisors of the authors about the academic misconduct. d) The author could also receive a letter that they are not welcome to publish in our journal for certain period of time (usually 3 years)."

M.D: **Is there anything else you wish to add regarding the issue of CGPs, peer review, and publication ethics?**

E2: "The CGP papers are not a problem. We have more serious problems with double submissions where people submit to journals that do not know about each other (their reviewers maybe different) and then wait for the response. We had to follow the process mentioned in 8 several times recently for such papers. Also, there are many papers that have an incremental contribution and it is difficult to distinguish if it actually is self-plagiarism or a new idea. These papers require a huge amount of work to judge and it is very tricky."

7.8. Appendix E.4: Interview Transcription David Grier

Interview with David Grier – Editor with IEEE
2/11/2019

MD: Malea Dunne (Researcher)
DG: David Grier (Editor)

M.D: I stumbled onto the topic of CGP and became interested because no one really knows much about it.

D.G: In a small tight community of us everybody knows about it and worries about it. But no one really talks to us about it. It's something of a concern it's been a concern on and off for probably twenty years in the interests you have in the topic. You stumbled into me and that means you stumbled into a person who has a long history in technical publication and knows a bunch. So, I'm not your typical editor and Computer is not your typical journal. So, with that, where would you like to begin?

M.D: So, you already said you are aware of CGP, so I'd like to ask more of, how did you become aware of these papers?

D.G: There was a period, roughly around 2007 through 2011, 2012, 2013 something like that, when plagiarism and mechanical plagiarism became an issue that we were aware of in particular two things, the WE, being IEEE, we publish about 200 scholarly journals and I was at that point responsible for the computer journals.

D.G: We had one [CGP] published, not in a magazine but in a periodical of a conference proceedings. It was highly embarrassing for us. And I don't remember the exact details of this, but it would usually be a conference that would consider itself to have a fairly strong reputation that was public. It [CGP] somehow got on the front page and someone levered it against us. And that was marked at the same time and these two things in our minds cannot be separated. Between 2007 and 2013 the number of plagiarism cases going completion, final judgement, it was roughly a factor of 10. So, lets say 10 to 250.

M.D: Wow, okay.

D.G: Yeah. And several things were driving it. The first was first primarily China and India. China rapidly expanding its education and India's educational industry putting more pressure on Indian professors to improve the quality of their work and similar pressure being brought to bear on mid-tier and lower tier engineering schools in the United States. Montana State is a morel school and among the morel schools its ranked like number four. And amongst engineering schools as a whole its ranked top 50 or so. But those schools are fine and faculty are fine. It's the ones at smaller schools that often don't have the background to do it and yet they are having pressure put on them to follow those standards.

Interview with David Grier – Editor with IEEE
2/11/2019

MD: Malea Dunne (Researcher)
DG: David Grier (Editor)

M.D: Okay.

D.G: When it becomes your job, in the whole case of plagiarism, um, that's what you do. You cut corners and in particular all of these groups generally lacked the full skill to write a paper, as I saw it at the time. It's a huge problem in China and Chinese education. They simply don't know how to do it. India is a similar problem because no matter what you argue, English is the language unifies India. It's not necessarily the first language in India and how you structure an English language paper in India, I mean they don't always understand. It's the same thing as dealing with less educated engineers at marginal schools.

M.D: Okay.

D.G: The mechanical faked papers came out of a program, this is how I understand it, but out of the artificial intelligence lab at MIT.

M.D: Yes, that is correct.

D.G: A student did this as a project I believe, and you would feed it three papers and it would make up stuff out of it. And it was largely done as a prank. Computer science has a long long, engineering does and so does MIT, of projects and pranks all wrapped up. They are technically challenging and they engage students because of the nature and it attracts them and they run off and just do it and once you take it out, it's tough to put it back in. And people saw the opportunity to mock and make fun of people. I'm sure there is still a program of it out there you could find.

M.D: I think I have found somewhat of what the program is.

D.G: Where did you find it? I am just curious.

M.D: Ohh, I am not sure. I will have to look back at my notes and then let you know.

D.G: The first place I would look, are you familiar with Gethub?

M.D: I feel like I have come across the site. If not, I will look into it.

D.G: Okay. Gethub is the big repository of people software projects. Its huge and has millions of programs you can download or modify. And a lot of them are big programs. A lot are just things that people want to show off. And I will give you a quote right now that I'm not sure if I want my name

Interview with David Grier – Editor with IEEE
2/11/2019

MD: Malea Dunne (Researcher)
DG: David Grier (Editor)

associated with but in engineering education there's a lot of it that re enforces habits and ideas that we associate with adolescents, um, there is a certain kind of superficial bonding over technology and team building nature because it is in its own kind of bubble in the outside world not understanding what it is. It emphasizes play and limits this sort of sense of learning responsibility for a bigger group of people and I think that aspect of engineering education which I'm not the only person to ensure identify it is one of the challenges with problems like this that the technical challenges could appeals to them, you could see where it would drive technical skill forward, but there is a lack of clear thought about what the ramifications might be and what you are doing.

M.D: Okay.

D.G: And I mean scientific research as a whole has this although a lot of scientific research is so distant from things that impact you and i. um, you don't see it. Engineering particularly computer engineering can and does. Are you familiar with um, Gullavers Travels?

M.D: Um, I believe so.

D.G: Um, chapter 3, the book on the research community. If you haven't read it, you should. They describe a good research community. The research is old but explains it well. The research remarks it as having no responsibility.

M.D: Okay, I will look into that research community.

D.G: The tools are so interesting for exploring technology that they forget that other people wont understand. I think those are the two very different aspects of plagiarism that when people try to preserve their job and when people are exploring speech synthesis, which is a challenging problem and happens a lot in computer science. To the extent, this is an overstatement, but a whole series of basic ideas of linguistics and language works comes from the early 1950s and were financed from being in part of computer systems and at some level there's an aspect of linguists and the study of language that is completely dominated because that's what is the audience. And its problems that graduates need to know and need to understand how language works and it's a growing problem.

M.D: Before I forget, earlier you mentioned about a specific quote being anonymous. Would you like me to take your name off that quote?

Interview with David Grier – Editor with IEEE
2/11/2019

MD: Malea Dunne (Researcher)
DG: David Grier (Editor)

D.G: That well you know, a whole bunch of people have said that with the research, and I don't want to come across as accusing anyone. I want to say that yes, this is an issue that many people have review and commented upon and that there is a challenge in the technical community with large, especially with graduates students community and people who are young of learning out of adolescents to full adult hood even though they are like 28 years old. It's a step of coming of age that is going to technical communication. I am not the only one who has said this. I guess what I am saying is I don't want to come across as the only guy who has said this. It is a common challenge of some of the work I do right now, its most interesting, how do you take 20 somethings in a technical world and turn them into leaders. And sense technology changes so fast by the time they are 40 they are going to become a technical wizard or they won't pass the bar. And for most people it's the latter.

M.D: Okay, perfect. I just wanted to make sure, so I didn't lose track of anything.

D.G: Thank you for asking.

M.D: So you kind of touched on it, but the next question do you see an issue with the CGP typing into publication?

D.G: Um, in the engineering community the answer is largely no. but in the world of the humanities crowd has a different problem. I suggest finding an individual with humanities to ask about their problems. Um, in engineering and the scientific world as well, the biggest issue is ya know, the ability to replicate results. And the reward structure of scientific research doesn't really reward you for saying "I took someone else's experiment and duplicated it exactly and I was right they were wrong". But, um, people do generally all the time take someone else's paper and say, "ah I can use this idea for this, and I can adapt it for my needs." And for many people particularly young people building their career that's a very common theme. You take someone else's experimental setting and their tools, and you say let me apply this here. And if those tools don't work or don't make sense um then there's an obvious short force to check from and that checking happens a lot. When I was managing Computer Journal there would people flagging stuff saying, "I don't think this is real". The first step you take is usually to go to the author and say "what's going on?" and then if that didn't work then you'd say that this person seems too advanced. And I believe that happened two or three times. Compared to the other um, forms of plagiarism it was a drop in the bucket. In particular there was one paper that I got that I had to deal with that was a little more theoretical

when reading it you could notice quickly identify the papers that the program had used to build it and they had been chosen to be so disparate that this isn't real work and you rejected it and you then heard nothing back. From the point of view from the IEEE the big problem is the waste of time.

M.D: Okay.

D.G: Computer right now is a fairly big journal right now. We have a staff of roughly 20 associate editors all of them are currently 1-3 papers in the review process. And if something proves to be completely false then you are wasting their time and the reviewers get a little pissed. The associate editors don't like that, the thought they have wasted their time. That's the real drawback, sort of keeping your editorial staff focused on what they are doing.

M.D: Okay, okay. You said no for that answer, but do you see any specific issues that tie in with academic journals?

D.G: In academic journals? What do you mean exactly?

M.D: Um, some of the papers that I have found and articles that I have read kind of tied in where they have found CGP in specific academic journals, like Nature and Springer. I guess doesn't have to be academic, but do you see an issue?

D.G: In IEEE are about 50 50 academic publication. I mean academic is a little more different in that the practitioner when there isn't the kind of check in building that an idea that the engineering community has. At some level the humanities and social sciences that they have been building intellectual frameworks that involve some fairly rare ideas. And sometimes those ideas are really able to get at the heart of the problem to understand it and strip away the problems that make it difficult to grasp and help this advanced knowledge. And when you have got CGP in effect are, and its very difficult to sort out sometimes who is showing off and just using big concepts and not really getting anywhere. And who is getting that really fundamental idea? And in particular because new professors and young processors in their first seven years of work are often trying to write papers to understand the tools and be able to apply them and when you have got people at work think there are papers that are meaningless and they are there solely to mock you are disrupting the process and you are making people question work that is in fact honestly done, but may not be particularly comprehensible.

Interview with David Grier – Editor with IEEE
2/11/2019

MD: Malea Dunne (Researcher)
DG: David Grier (Editor)

M.D: Okay.

D.G: Ya know when you are mastering sort of a difficult more philosophically based, like are you familiar with the writings from Eruko?

M.D: I don't believe that I am.

D.G: diserplented birth of the clinic. (CONFUSED WITH THE MIX UP). He is a philosopher, hes been dead for about 20 years, shoot I mean Bruno Latour, he wrote Opening Pandora's Box, and it was a book that claimed that the model of how science is done is kind of like a ladder. You go up the ladder and the ladder represents steps of abstraction and at each stage of the ladder you can work within the framework you understand but then you kind of take a step, a little bit of a leap of faith apparently it's a map of taking ideas and compressing them into a fixed set of categories and he gives an example of suppose you are studying soil dynamics and you go out and collect a bunch of soil and you have red soil, black soil or soil with bugs in it, and that's fine. But then you invent a set of categories that say these are different classes of soil and you go at it and say this is type a, type b, type c. when you have categorized it you are in fact abstracting ideas and in effect making a leap of faith that you have in fact captured all the stuff you have needed to captured.

M.D: Okay.

D.G: You go there and its step by step by step and at each step there is a jump. The humanist have gromed on this for looking at both history and at certain kinds of language in novels and these ideas are not easy to understand and some of the applications in literature are tough to grasp what they are getting at and there are some valid insights it can give you, but you have to work out some problems to understand how the tools work and how the tools operate and how you can apply them. And if you got a paper generator and you have a Latour paper as one of its bases you are disrupting people who are trying to be honest scholars.

M.D: Okay.

D.G: And so, I would do less academic, most of the stuff we publish in IEEE is engineering professors trying to do good work. But in places were replication is possible, is engineering and science, or near replication that you do something close enough. You do experiment A and I do experiment B, but B isn't working, then you gotta go back and look at

experiment A and say am I doing it wrong or did you fake it? Whereas in humanities and social sciences that's not necessarily the case. You can critique someone's work, but you can't quite duplicate it and you aren't reaching places where you ask did you do it? And that's where the problem really lies and where the real damage is done.

M.D: Okay.

D.G: And you can speak with humanists and scientists to critique me. As I said, the raw plagiarism is a much bigger deal.

M.D: Okay, perfect. Do you think there are opportunities to improve the peer review process? In your opinion.

D.G: Oh yeah, there is. It kind of goes from subject to subject. Ya know peer review used to be done at some level with a group of people sitting around a table passing notes and talking about what their students have done. And there was nothing. At this point it is spread so widely, they really do no training or very limited training on it. And how you would bring reviewers on and sort of train them and get them to understand the process and think about things and what they should look for would be a good training method. How you get people trained and how you get them to accept them and get common standards, that would be hard. I think some of the professional societies could do it. I think IEEE and Computing ICM could do it. But one of the challenges again getting the referees to do it and getting them to feel there is a sense of accomplishments there, that's the tough bit.

M.D: Okay, okay.

D.G: I could easily see a seminar that would teach at conferences to be a referee that this is what you do. That right now this is what you do but they have the professional societies like IEEE and ICM and things like that doing it. But you have private publishers too, the big one is Elsevier, the kind of people that work for them are the same sorts that work for the professional societies. Don't always completely overlap, the overlap is substantial but is far over 50% and far less. There incentives are making money on publications, where professionals that inventive fell off the table years ago. Um, and it's not clear how much they would buy into supporting that. Um, the groups that could really force that, like the National Science Foundation, right now they have really no interest in doing that. It's not the kind of thing they feel is in their purview.

Interview with David Grier – Editor with IEEE
2/11/2019

MD: Malea Dunne (Researcher)
DG: David Grier (Editor)

M.D: Okay.

D.G: A different administration might change that, but we will see.

M.D: Okay, and you kind of hit on it, but would you suggest any strategies on the improvements?

D.G: Um, I mean the strategy on how to get the incentive for this is having one journal or a set of journals where you viewed them as having a greater status than others and in particular if you took a top bit journal and said okay we are only going to have only trained referees and you worked up and built even a small staff and said this is what you'd get its part of building that of a premier product that if you went through that and your article was refereed we would know the quality of it and the author would get high quality of that comments back and the work would improve substantially and be able to build a community that this is important, this is good. And get people to start flowing into that and duplicating it. Much of the direction has been kind of the other way. I don't wish to do a complete dump on it, but open access. The initial ____ (CANT UNDERSTAND) is quite interesting. 31:29 but kind of after open access got started and it settled into its current model um, its not clear that people take the work there as seriously from a professional society or privately published journals or that its as well done. The refereeing process is done now their websites claim otherwise.my experience is that they are looking at things optimistically and there are certainly some of the smaller regional open access journals that are barely refereeing anything.

M.D: Okay.

D.G: The big ones are okayish the smaller ones are not. So, I don't see a push, I wish I did but I don't see it right now.

M.D: So you would kind of suggest the incentive route then for these editors?

D.G: Yes. I think there's gotta be some sort of incentive. You can't require it, um, ya know and, one of the issues we are facing is in the united states there are about 3000 colleges and universities. Um, and it used to be, about a generation ago that the faculty that are involved in these colleges had much more limited publications on them then they do now, than someone does at Montana Tech or Montana State would be expected to write a paper or two to get tenure over a 7 year period and they would be expected to continue to write papers at that kind of pace for the rest of their career so they might end up having 10-12 papers that would be to

their name and would mark their career. That is no longer the case. Graduate students coming out are expected to have used to be 1-2 papers now it's about 5-6. You are expected to turnout as a new professor in the engineering world or scientific world a paper or two a year. If you are running a research lab, ya know more of a senior professor, you would like to see 8-10 come out of your lab a year. And that is obscene in the number of pages in journals that have expanded. When I first got involved in the IEEE, we had maybe 22 journals and when I became Vice President of publications 15 years later we have gotten up to 34 that is a substantial increase in the amount of pages available in publication and then you get the minor journals and the other things out there, there is a lot more space out there. And there is a lot more demand to fill it.

M.D: Okay.

D.G: Ya know the median number of downloads in IEEE journals last I checked, 50% above and 50% below its like .9 that most of the journal articles were never downloaded. Now that's not the same thing as cited or aware of them. But that what you looked up for like the 90th percentile was like 4 or 8 or something small. That as a worker who read literature these papers are not important. To put that positively it becomes part of effective training and validating that professors are expected to engage their field every year, expected to solve problems in their field every year and they are expected to have those problems verified by their peers. And that is a good thing, that is fine. And that is not the same thing as producing papers that should be widely disseminated and educated the field at large. And we may at some point want to distinguish between those two activities. And have one mechanism set up for just trying to keep their careers validated and another when you are with the field at large. Because I think right now one of the things that is true about the papers I deal with, most of the people that are important, who can use your research, can use your paper before the refereeing process is over. There are people to send it to plus the referee and then you are done.

M.D: Okay. Have you dealt at all with COPE? Committee on Publication Ethics?

D.G: I don't think so. *Looked up COPE during interview* The NSF had a committee and let me see how this is related to it. I have met some of the individuals on the team. The answer is no. I think I have been more aware of it than I apparently knew. This is the kind of organization you see in professional societies or a group of them banned together. It's the kind of thing that could ban together to do some work on problems such as this.

MD: Malea Dunne (Researcher)
DG: David Grier (Editor)

while researching in interview Interesting, they have been running plagiarism seminars in Beijing.

M.D: Okay. At IEEE, is there a retraction process and what is the retraction process?

D.G: Yes. There is basically a plagiarism committee and there is a process that I was involved in refining when it moved from 25-250 its this incredibly messy spreadsheet, because engineers love spread sheets. And it more or less explains it but you have got to remember the audience and if you go to the IEEE you will find regular papers that have been removed and there is a banner that says "Paper has been removed" and is intellectual property. And that's used a lot more than we would have liked. When I was involved we were dealing with 250 cases a year. My understanding is at some level the growth has stopped, but its still a really large number of cases.

M.D: Okay.

D.G: and I have done a certain amount of training with people that basically say remember you can be yelled at for having too much quoted text and having too many footnotes but you cant be banned from publication. Whereas if you don't have enough footnotes and you copy footnotes you could be banned from publication so footnote everything! And there is a truth in the matter if people who are struggling to write papers and they copy text and say "im doing an experiment like this" and copy paragraph after paragraph all in quotes and with footnotes people will say "you shouldn't do that" but you wont get banned.

M.D: Okay.

D.G: You are not being illegal you are just not being a very good writer.

M.D: So does anything specifically happen to that editor or that writer if a paper is retracted?

D.G: Yes. You are banned from publication. You could be banned for a year or you could be banned for up to 10 years. Sometimes universities will go after the professor for intellectual misconduct. Where the dean called me and we sat down and went through the papers and it was clearly a disciplinary hearing about the faculty member and I did not probe too deeply because it was not my business but I got enough to know more commonly and let him know the contract will expire and told him "you will

never get this contract back". I went through this once when I was the dean, there was one person who claimed they had done a certain amount of intellectual work. Those claims proved to be false and we talked it through and we said "you've got 18 months and you can finish your contract and that is it".

M.D: Okay, interesting

D.G: And when you are dealing with large human organizations that often the safest thing to do that it reduces the least shock that everybody realizes what is happening. Because once you go on a removal, once you are on their fire for cause, lawyers can get involved quickly and can consume a lot of money and a lot of time.

M.D: Okay.

D.G: And rather than say "we don't have ground to renew" that's often much easier. And its much easier on the whole organization. Because when you are firing someone faculty takes sides, there are often depositions, everyone knows that often there is a leader that is making that decision and their judgement is questioned and have to pull out of the decision making process temporarily to make sort of the organization going. It can be incredibly disruptive.

M.D: Okay. And the last question, is there anything that you would like to add to this interview regarding CGP, peer review or publication ethics? Anything you find important or that you think would help me out?

D.G: Well, I think I've given you the key inside that I claim to have, which is there is a difference which sort of sums up the difference with computer science and engineering on one side and humanities and social sciences on one side. Um, the other bit that's involved in it that goes back to an earlier set of comments, is how deeply this work is tied up in both prestige and in income and those two things are not the same, but they are often tied up in it. And what people will do to keep their position and to build their sense of identity and authority and prestige on a fault basis. Im not the biggest fan of academic novels, but there is an author I liked a lot, but I think he is dead and gone, named David Lodge in the 90s and he wrote a bunch of things about an English professor who's goal it was to be the highest paid English professor in the UK. And it went through and really sort of honed a way at people for who being perceived as important was more important than doing good work. And that's a problem and its not really a problem of anything other than mortal existence that there are

MD: Malea Dunne (Researcher)
DG: David Grier (Editor)

people who's ideas about themselves and how they should be able to demand attention is misguided.

M.D: Okay.

D.G: Will we be able to solve this? Maybe when we stop having people who have the wrong set of goals.

M.D: The last thing I would like to bring up. In one of your emails you mentioned a scanning tool to detect these papers?

D.G: Yeah, we now have a scanning tool that goes on to find these. You will need to contact Bill Hagen who will be able to give you more information on it. What it is for us is several software tools that are combined and merged into managed systems and every paper that comes is scanned for a variety of things and against some other libraries and I believe the current version of it picks off the bibliography and scans against that.

M.D: Okay, perfect. Thank you very much for speaking with me and giving me this information!

7.9. Appendix E.5: Interview Transcription Tamara Welschot

Interview with Tamara Welschot – Springer Nature's Research Integrity Group
2/20/2019

MD: Malea Dunne (Researcher)
TW: Tamara Welschot (Editor)

M.D: How would you like to proceed with the interview? I can just ask you each of the questions if that is easier to flow with the interview?

T.W: That would be perfect Malea.

M.D: I can just read off each of the questions if that makes it easier for you?

T.W: That would be the easiest, Malea. Go ahead.

M.D: So the first question, which you basically answered in your previous email, but are you aware of CGPs?

T.W: Yes, I am aware of computer generated papers.

M.D: More, how did you become aware of these papers?

T.W: Ah okay maybe I should explain a little more of my role was and is at the moment. So I work for Springer and Nature and before the merger with Nature we were just Springer Science and Media. And uh for that company I was working as a director of research integrity and publishing services, meaning that I was heading up a small team doing various things and one of them was um handling complaints and issues in regard to at the time it was called Publication Ethics it wasn't necessarily called Research Integrity, just publication ethics. Um, so because I handle the complaints and questions from colleagues about policies like how to deal with plagiarism and these kinds of things. We had our public email where people could submit questions and complaints and indeed, I got a message from a scientist saying ya know "I would like to warn you because there was a couple of papers that were apparently very fake." And so I thanked him and then I took some time to look into that and when I looked into those papers, which were actually published by the company I saw that they were rubbish, not real scientific papers and that has triggered the whole process into an investigation. And this was all in the computer science discipline. So I worked with the employees in the computer sciences and um, yeah Scigen, computer science papers. And that is how I found out about it when I started dealing with these papers.

M.D: Okay, interesting. Off of the top of your head, how many of these papers have you come across? If you had to guess a number?

T.W: Um, well this was already a couple years ago. I can't recall how many paper we actually did retract. Because it was another publishing company,

IEEE, they had to retract quite a few papers I think we had about if I'm not mistaken 20 papers. Let me look quick.

M.D: Okay.

T.W: It was 18 papers. And IEEE had more than 100 papers.

M.D: Yeah, that is crazy.

T.W: Yes, yes. And if you wonder how that happened, um we are talking 2014 if I remember correctly. Ya know it was the time that people weren't really aware of these kinds of things maybe a few were but not within the publishing company and a lot of um, signs based on trust and papers are being submitted for publication and of course peer review and etc. um, and in this case peer review was not handled by us as a publisher but we trusted to conference organizations who take care of peer review and I think this is too much trust put in others and it also has to do with a lot of pressure amongst office and while researchers and scientists publish papers and so we saw that a lot of people from particular parts of the world use this paper generator to have papers published so they have records for themselves.

M.D: Okay. So do you see kind of an issue of how CGPs are being published or an issue with CGP being tied in with journal publication?

T.W: Um, well it's both proceedings and journal articles. I mean these are proceedings. Yeah, computer science proceedings after that what happened, we really started looking into conferences, which conferences are presenting with these papers because you may notice there are a lot of predatory conferences out there and a lot of predatory journals out there and it's their role to lure authors into publishing with them authors have to pay for it uh, and office mistakenly submit papers to these type of conference organizations etc etc. We have learned a lot from that and colleagues in computer science have really very much strengthened the process around publication of these proceedings' papers. We also develop kind of an algorithm that is kind of still being used to see if we can spot these papers. I think this was a one time off to be honest. I realize there are much more paper generators like Mathgen and PhisGen there cannot be any guarantees on cases specifically for journals where you have or should have a peer review process in place. Ya know by looking at the paper and reading it you should get an idea if this is a hoax paper or a fake paper etc etc. Um, I think we should realize that algorithms are becoming much better and although I am personally thinking that still you

MD: Malea Dunne (Researcher)
TW: Tamara Welschot (Editor)

can find out about some papers that are machine generated because there is something in the papers that are not matching up. But when you read them you actually see or notice that there is something wrong with. But again, there is a lot of trust in humans to spot these kinds of things and with I don't know how many submissions we get on a daily basis for a wide variety of journals it all rests on the shoulders of our editors of these journals and the reviewers. That was one of your intentions right, to talk to and find out how you deal with it and how do you find out about it?

M.D: Yes.

T.W: Quite recently I was informed or alerted I should say, to a paper that was rejected by an editor in chief for being completely globlygook and I looked into the paper and indeed it was, it looked very much like it was a machine generated article and I put it in ithenticate, I think you probably know about Ithenticate or Turnitin and you know there are lots of snippets of text from text from Wikipedia. Again, this is from a different part of the world where pressure is high, the question is of course why do they do it. I think on our end by using ithenticate *unclear what is being said after* or editors being more vigilant and aware of these things it will, I don't think it can ever uh get rid of it. Because like I said, there's still quite a bit of trust with editors into the whole process of papers being submitted by ammos authors. My role now a days is to provide training to our own colleagues in the company but also external to make people aware of what can actually happen. And every time I learn that people are surprised to be aware of these things, not everyone, there are lots of editors in chief that are actually aware of it because they see other things, they see what researchers can do to delay process or manipulate in whatever way to get a paper published. But if you realize there are particular disciplines where none of this has happened before you can imagine these editors have no clue what to look for. They might get a paper they might only look at the title and think "oh this is interesting" and send it off to peer review and peer review might say "ya know this just isn't a good paper" or "this is complete nonsense. So far I must admit like I said we only have a very limited instances in these retractions we did in 2014 and yeah.

M.D: Okay. I like the wording you use with calling computer generated papers "rubbish" and "globlygook". Some of the articles I have read call computer generated papers that, but I haven't heard someone actually call them that. So that is very interesting to hear. I like that.

T.W: Well I have dealt with that first hand. And they really don't make any sence. But yeah, it has happened and we have chosen from what

MD: Malea Dunne (Researcher)
TW: Tamara Welschot (Editor)

happened in 2014 and we have to learn from that and have to make others aware that this could happen.

M.D: Okay. With the peer review process do you think there are opportunities for improvement with the peer review process or what kind of peer review process are you aware of or use?

T.W: Well we have various models of peer review. Accentually it really means that any papers being reviewed by 2 reviewers that's the best practices and of course you have single lined and double lined and opened review and transparent peer review. But it actually boils down to two experts reading it or reviewing it and coming back with comments. Um, I think that's still the best way of making sure that what gets reviewed or published has some degree of screening. In my view when such a paper does, when an editor in chief has not seen a nonsense paper and still sends it off to peer review it would assume that peer reviewers would definitely alert the editor in chief to the fact that this is not in fact a good enough paper. That has not come to my attention, it could be that this has happened, but we are a large company and of course editors in chief have their own responsibilities. They can use their own editorial discretions so if they essentially reject a paper because its nonsense it doesn't mean that we are always aware of that. So its part of the editorial responsibilities so to say. It could be that an editor in chief might flag it to their person who might flag it to me, ya know we have found about this that this is a nonsense paper and make aware of that, because within the company we have particular ways of ya know, various databases where we can check if an offer has been more active on that level or not. So far we have not come across that, so I think that the first line of defense would be the editor in chief and then the next step would be the reviewers. So they have an important role in essentially really reviewing the paper and has scientific merits to get published and accept it or revise before getting accepted.

M.D: Do you think there could be room for improvement in this process?

T.W: No, essentially I think its good enough. There are other ways but I think that this line of defense maybe something else that could be, but its something that we are already doing is that papers are being screened for plagiarism and all kinds of other things via iThenticate being done for a lot of other journals. But also there we have not found any instances of machine generated papers or nonsense papers.

MD: Malea Dunne (Researcher)
TW: Tamara Welschot (Editor)

M.D: Okay. A little change in the track, but have you worked at all with COPE? The Committee on Publication Ethics?

T.W: Absolutely. Actually I am one of the contact persons for the company.

M.D: Oh great! Could you tell me a little more about COPE and working with COPE?

T.W: Yes, so we register or most publishers register journals with COPE its to make everyone aware that we take research integrity or publication ethics very seriously, so that's one step. So they take care of making sure these journals are registered with them and are available in their database so everyone can see which journals have been registered and essentially means that all our journals follow the COPE guidelines and as a company our own policies are very much towards the COPE guidelines, or the Code of Practice I should say and flow charts. I think they have really good discussions on lots of different topics as publishers we are very much involved and every publisher has particular topics that they work with COPE and so in the past the colleagues by the * cant understand * imprint they have worked with COPE on Techs recycling guidelines I think last year we as a company worked with COPE on flow charts on guidelines on manipulation on the submission process. There's currently a survey out that COPE is doing to get the * cant understand* so every publisher more or less works with COPE on particular topics. I think we cannot go around COPE, I know some of the society publishes actually have their own guidelines. They're probably a little more stricter on their policies or instructions I should say. And that means that it is sometimes said that whenever you have a complaint COPE is not really the one saying ya know to be a police officer or saying yes or no, they essentially look at whether the processes have been followed. In their positions it is probably also the best thing they can do. They are essentially not the police and the publisher is like a police officer or a judge. We have to follow due process or diligence to find out if something needs to be retracted or not. I think the whole focus of COPE is really making sure that research literature is ya know that there is integrity into what is being published.

M.D: Okay. That's an interesting process.

T.W: Yeah. I think in the US COPE is not, I think everybody is probably aware of COPE but I am also aware of people wanting to start their own COPE in the US. Yes.

MD: Malea Dunne (Researcher)
TW: Tamara Welschot (Editor)

M.D: Okay. Um, you kind of talked a little earlier about the retraction process, but could you talk a little more about what exactly the retraction process is or what it was with finding a CGP?

T.W: Yeah. So once, I mean it is quite easy to establish that these papers were fake or nonsense one of the guidelines of COPE is to reach out to the office to find out more about why did they do it and etc. and of course in these instances you get very limited responses or no responses at all. Most people are embarrassed or ashamed. I think only a few apologized. So essentially when we have enough reason to move forward with retraction we then retract these papers and we publish a note saying that this paper has been retracted for whatever reason. In this case, I actually have to think really hard since it has been a few years ago now. But now a days we should publish a note saying why that article has been retracted. I think in those instances we did not, but I would have to check to be sure. But a normal retraction procedure is when you retract a paper that you clearly publish the reason for retraction. It could be for plagiarism, it could be for data fabrication or missing informed consent or they had no ethical approval - these kinds of reasons for all kinds of reasons to do retractions where the results cannot be relied upon any longer. So that would be the normal retraction process.

M.D: Okay. The last question that I have, is there anything that you haven't touched on or would like to tell me about CGP or the peer review process or even the ethics you think would be helpful for me?

T.W: Well, that depends a bit what your goal is. Because that is something. You are doing this for your thesis research correct?

M.D: Yes, for my masters thesis. I'm looking at the peer review process, kind of where it is lacking in certain areas.

T.W: The peer review process will definitely filter out these papers. I think in the time of 2014 there was too much trust in having the peer review process handled by others. This has led to this problem, thankfully for us only 18 papers and not a 100 papers like IEEE. But it could have happened to us too. But I think you also need to make a distinction between proceedings that are published as books and nonsense articles and journals. I mean retractions we did for SCIgen generated papers were proceedings in books. And we have not retracted any papers in journals being nonsense. I think that is something you need to be aware of in your thesis. Because journals have this very clear peer review process, essentially proceedings do but those are not handled by an editor in chief, those are usually

MD: Malea Dunne (Researcher)
TW: Tamara Welschot (Editor)

handled by a program chair or a conference chair. So that's the main difference between these two situations. So I think for the journals the peer review process is pretty good. But is can always be better because not related to fake papers, peer reviewers its very difficult for them to find out about data fabrication. This is really really hard to find out about. Sometimes even plagiarism is difficult, they get a paper for review, either they have read the text themselves and think "ah I know the author" or otherwise we rely on software like iThenticate to alert potential plagiarism. So essentially for journals we have not found instances of nonsense papers except for the one that was already rejected and also alerted to it. Maybe this is interesting for your files in the office, they have submitted another paper similar to the one that was rejected, but that one was already rejected as well. It makes me think the peer review process works and editors do find out about these nonsense papers.

M.D: Okay, great.

T.W: I just want to ask. Is your goal to find out where the peer review process could or should be changed or is it more to learn about how aware editors are because you would have more interviews with editors if they are aware of it?

M.D: Honestly, kind of both. I've spoken with a couple editors already some are aware and actually there are some that are unaware of the CGPs or they haven't come across any. So its kind of interesting to get each editors views on the topic. So its interesting the results I am getting back on this topic.

T.W: I think as I mentioned some editors in specific disciplines such as computer science, are much more aware of these kinds of things and when you go to editors in social sciences they will think "uh what?". So yeah depending on disciplines I think you are right. Some are more aware than others. And if I can give you a tip, this is really on fake papers. But you could extend your research into the issue of fake conferences or fake journals. They are in a way connected it all has to do with making money. Authors have a lot to gain. If they publish in a predatory journal they gain just a paper published. Its all connected to the "make believe".

M.D: Okay, great. Thank you very much for agreeing to speak with me and giving me information to add to my masters thesis.

7.10. Appendix F: Thematic Analysis Coding

Thematic Analysis Themes & Coding

1. Awareness
    - Prior knowledge on the CGP problem? before interviews
2. Becoming Aware
    - How the editor became aware of CGP problem
3. Issue
    - Are there any issues with the CGP topic?
4. Improvement Opportunities (Yes or No)
    - How can the CGP problem be improved
5. Line of Defense
    - The step-by-step process of peer review
6. Rejection Rate
    - How many papers have been rejected & why
7. COPE
    - Topic of *Committee on Publication Ethics.* Knowledge and/or involvement of COPE
8. Retraction of Papers
    - Removal of published papers & aftermath of removal
9. Publication Ethics
    - Ethical issues regarding publication in journals
10. Pressure to Publish
    - Authors and faculty members being pressured to publish academically
11. Scanning Tool
    - Tools developed to detect CGP

Awareness

- “I don’t know that much” (E1, 2019)
- “Yes” (E2, 2019)
- “Yes” (JJ, 2019)
- “In a small tight community of us, everybody knows about it [CGP] and worries about it. But no one really talks to us about it. It’s something of a concern been a concern on and off for probably twenty years in the interests you have in this topic” (DG, 2019)
- “Yes, I am aware of Computer Generated Papers” (TW, 2019)

Becoming Aware

- “I have heard this before” (E1, 2019)
- “A colleague told me about it many years ago. He actually tricked me to read one. It took me three paragraphs to notice it was a fake paper” (E2, 2019)
- “I became aware of these through the infamous SCIgen MIT Paper generator that was used to produce a paper accepted at a conference (SCI) without a peer review system” (JJ, 2019)
- “There was a period, roughly around 2007 through 2011, 2012, 2013 or something like that, when plagiarism and mechanical plagiarism became an issue that we were aware of in particular two things, that WE, being IEEE, we publish about 200 scholarly journals and I was at that point responsible for the computer journals” (DG, 2019)

- “We had one [CGP] published, not in a magazine but in a periodical of a conference proceedings. It [CGP] somehow got on the front page and someone levered it against us.” (DG, 2019)
- Um, so because I handle the complaints and questions from colleagues about policies like how to deal with plagiarism and these kinds of things. We had our public email where people could submit questions and complaints and indeed, I got a message from a scientist saying “ya know I would like to warn you..” because there was a couple of papers that were apparently very fake.” (TW, 2019)

Issue

- “I don’t think it is common in science and engineering publications” (E1, 2019)
    - NO ISSUE
- “If it happens, it is a very serious problem” (E1, 2019)
    - YES
- “Not really for high quality publications. In fact, it is bitter for low-quality journals and venues that they can be fooled this way. It shows that they are not rigorous” (E2, 2019)
    - NO ISSUE
- “The CGP papers are not a problem” (E2, 2019)
    - NO
- “The way I see it, a journal with a decent peer-review system should be able to spot SCIgen generated papers. However, SCIgen may only be the precursor of

more sophisticated approaches, that may become progressively harder to unearth" (JJ, 2019)

- "Um, in the engineering community the answer is largely no, but in the world of the humanities crowd has a different problem. Um, in the engineering and the scientific world as well, the biggest issue is ya know, the ability to replicate results" (DG, 2019)
    - NO ISSUE
- "From the point of view from the IEEE the big problem is the waste of time" (DG, 2019)
- "Ya know, by looking at the paper and reading it, you should get an idea if this is a hoax paper or a fake paper etc etc. Um, I think we should realize that algorithms are becoming much better and although I am personally thinking that still you can find out about some papers that are machine generated because there is something in the papers that are not matching up" (TW, 2019)

Improvement Opportunities (Yes or No)

- "Nowadays there are software detectors for similarities to identify plagiarism. This process is improving" (E1, 2019)
    - YES
- "There are always ways to improve" (E2, 2019)
    - YES
- "Yes, we are using more and more automated tools to detect plagiarism. Fake and plagiarized papers need to be fought with software tools" (JJ, 2019)
    - YES

- “Oh yeah, there is. It kind of goes from subject to subject. Ya know, peer-review used to be one at some level with a group of people sitting around a table passing notes and talking about what their students have done. And there was nothing. At this point, it is spread so widely, they really do no training or very limited training on it. And how you would bring reviewers on and sort of train them and get them to understand the process and think about things and what they should look for would be a good training method. How you get people trained and how you get them to accept them and get common standards, that would be hard. I think some of the professional societies could do it. I think IEEE and Computing ICM could do it. But one of the challenges again is getting the referees to do it and getting them to feel there is a sense of accomplishments there, that’s a tough bit” (DG, 2019)
    - YES
- “Well we have various models of peer-review. Essentially it really means that any papers being reviewed by two reviewers that’s the best practices and of course you have single lined and double lined and opened review and transparent peer review. But it actually boils down to two experts reading it or reviewing it and coming back with comments Um, I think that’s still the best way of making sure that what gets reviewed or published has some degree of screening” (TW, 2019)
    - NO
- “The reviewers need to read the paper carefully, they need to be experts in the field, and they should use some tools, such as iThenticate. I can also imagine

some Al tool to quickly discover if the paper has been automatically generated" (E2, 2019)

- "I could easily see a seminar that would teach at conferences to be a referee that this is what you do." (DG, 2019)
- "Um, I mean the strategy on how to get the incentive for this is having one journal or a set of journals where you viewed them as having a greater status than others and in particular if you took a top bit journal and said okay we are only going to have only trained referees and you worked up and built even a small staff and said this is what you'd get its part of building that of a premier product that if you went through that and your article was refereed we would know the quality of it and the author would get high quality of that comments back and the work would improve substantially and be able to build a community that this is important, this is good." (DG, 2019)
- "Yes. I think there's gotta be some sort of incentive" (DG, 2019)

Line of Defense

- "As long as the editors and then the reviewers do diligently their work, it should not be a problem" (E2, 2019)
- "We have a staff of roughly 20 associate editors all of them are currently 1-3 papers in the review process." (DG, 2019)
- "So far we have not come across that, so I think that the first line of defense would be the editor in chief and then the next step would be the reviewers." (TW, 2019)

Rejection Rate

- I currently reject about 20% of all submitted papers without sending them to reviewers for various reasons. The main one being out-of-scope papers. Reviewer fatigue is a serious problem, that can lead to bad papers being accepted" (JJ, 2019)
- "The peer review process will definitely filter out these papers. I think in the time of 2014 there was too much trust in having the peer review process handled by others. This has led to this problem, thankfully for us only 18 papers and not a 100 papers like IEEE." (TW, 2019)

COPE

- "Yes, we warn and punish certain authors whose papers are identified having plagiarism" (E1, 2019)
- "Yes I do" (E2, 2019)
- "Yes. I represent Elsevier multimedia publications in COPE" (JJ, 2019)
- "I have met some of the individuals on the team. The answer is no. I think I have been more aware of it than I apparently knew. This is the kind of organization you see in professional societies or a group of them banned together. It's the kind of thing that could ban together to do some work on problems such as this." (DG, 2019)
- "Absolutely. Actually I am one of the contact persons for the company" (TW, 2019)
- "Yes, so we register or most publishers register journals with COPE its to make everyone aware that we take research integrity or publication ethics very seriously, so that's one step. So they take care of making sure these journals are

registered with them and are available in their database so everyone can see which journals have been registered and essentially means that all our journals follow the COPE guidelines and as a company our own policies are very much towards the COPE guidelines, or the Code of Practice I should say and flow charts. I think they have really good discussions on lots of different topics as publishers we are very much involved and every publisher has particular topics that they work with COPE and so in the past the colleagues by the * cant understand * imprint they have worked with COPE on Techs recycling guidelines I think last year we as a company worked with COPE on flow charts on guidelines on manipulation on the submission process. There's currently a survey out that COPE is doing to get the * cant understand* so every publisher more or less works with COPE on particular topics. I think we cannot go around COPE, I know some of the society publishes actually have their own guidelines. They're probably a little more stricter on their policies or instructions I should say. And that means that it is sometimes said that whenever you have a complaint COPE is not really the one saying ya know to be a police officer or saying yes or no, they essentially look at whether the processes have been followed. In their positions it is probably also the best thing they can do. They are essentially not the police and the publisher is like a police officer or a judge. We have to follow due process or diligence to find out if something needs to be retracted or not. I think the whole focus of COPE is really making sure that research literature is ya know that there is integrity into what is being published." (TW, 2019)

Retraction of Papers

- We haven't done that [retracted] in the past, but I think we should publish the findings" (E1, 2019)
- "I think for our journal it is almost impossible. We really read each paper many times. However, if this [retraction] would happen we would start a formal investigation with the Ethic committee. If the result would be positive, we would recommend actions. First, we would inform the authors and ask them for a retraction if the retraction would be unsatisfactory, the journal would a) inform the authors about the final decision b) retract the paper with a mention that it was retracted, with full names of the authors, their affiliation, and the name of the paper c) depending if this was intent or an error, we would also inform the direct supervisors of the authors about the academic misconduct. D) the author could also receive a letter that they are not welcome to publish in our journal for certain period of time (usually around 3 years)" (E2, 2019)
- "This [retraction] can possibly be the most embarrassing event to happen to an editor. The text would have to be retracted As Soon As Detected, identifying the paper as a fake manuscript, with a personal note of apology from the Editor-In-Chief. Who would have to have to assume the blame at a personal level" (JJ, 2019)
- "Yes. There is basically a plagiarism committee and there is a process that I was involved in refining when it moved from 25-250 its this incredibly messy spreadsheet, because engineers love spread sheets. And it more or less explains it but you have got to remember the audience and if you go to the IEEE you will find regular papers that have been removed and there is a banner that

says “Paper has been removed” and is intellectual property. And that’s used a lot more than we would have liked. When I was involved we were dealing with 250 cases a year. My understanding is at some level the growth has stopped, but its still a really large number of cases.” (DG, 2019)

- “Yes. You are banned from publication. You could be banned for a year or you could be banned for up to 10 years. Sometimes universities will go after the professor for intellectual misconduct. Where the dean called me and we sat down and went through the papers and it was clearly a disciplinary hearing about the faculty member and I did not probe too deeply because it was not my business but I got enough to know more commonly and let him know the contract will expire and told him “you will never get this contract back”. I went through this once when I was the dean, there was one person who claimed they had done a certain amount of intellectual work. Those claims proved to be false and we talked it through and we said “you’ve got 18 months and you can finish your contract and that is it”” (DG, 2019)
- “The first step you take is usually to go to the author and say “what’s going on?” and then if that didn’t work then you’d say that this person seems too advanced” (DG, 2019)
- “Between 2007 and 2013 the number of plagiarism cases going completion, final judgement, it was roughly a factor of 10. So, lets say 10 to 250.” (DG, 2019)
- “I can’t recall how many papers we actually did retract. Because it was another publishing company, IEEE, they had to retract quite a few papers I think we had

about if I'm not mistaken 20 papers. It was 18 papers. And IEEE had more than 100 papers." (TW, 2019)

Publication Ethics

- "Also, there are many papers that have an incremental contribution and it is difficult to distinguish if it actually is self-plagiarism or a new idea. These papers require a huge amount of work to judge and it is very tricky" (E2, 2019)
- "The root causes need to be addressed to solve what is otherwise a very human problem. Until that happens I could see further and more capable strains of text generators appearing. Indeed there has been considerable research on computer-generated novels and poetry. Look at the high-profile retractions of manuscripts with fabricated results from Nature, Lancet and Science. Having a program to assist in Fabrication is just another way to make the process easier." (JJ, 2019)

Scanning Tool

- "Yeah, we now have a scanning tool that goes on to find these. What it is for us is several software tools that are combined and merged into managed systems and every paper that comes is scanned for a variety of things and against some other libraries and I believe the current version of it picks off the bibliography and scans against that." (DG, 2019)

Pressure to Publish

- "It's the ones at smaller schools that often don't have the background to do it and yet they are having pressure put on them to follow those standards" (DG, 2019)

- “When it becomes your job, in the whole case of plagiarism, um, that’s what you do. You cut corners and in particular all of these groups generally lacked the full skill to write a paper, as I saw it at the time” (DG, 2019)
- “And for many people particularly young people building their career that’s a very common theme.” (DG, 2019)
- “in particular because new professors and young processors in their first seven years of work are often trying to write papers to understand the tools and be able to apply them and when you have got people at work think there are papers that are meaningless and they are there solely to mock you are disrupting the process and you are making people question work that is in fact honestly done, but may not be particularly comprehensible” (DG, 2019)
- “it used to be, about a generation ago that the faculty that are involved in these colleges had much more limited publications on them then they do now, than someone does at Montana Tech or Montana State would be expected to write a paper or two to get tenure over a 7 year period and they would be expected to continue to write papers at that kind of pace for the rest of their career so they might end up having 10-12 papers that would be to their name and would mark their career. That is no longer the case. Graduate students coming out are expected to have used to be 1-2 papers now it’s about 5-6. You are expected to turnout as a new professor in the engineering world or scientific world a paper or two a year. If you are running a research lab, ya know more of a senior professor, you would like to see 8-10 come out of your lab a year” (DG, 2019)

- “Um, the other bit that’s involved in it that goes back to an earlier set of comments, is how deeply this work is tied up in both prestige and in income and those two things are not the same, but they are often tied up in it. And what people will do to keep their position and to build their sense of identity and authority and prestige on a fault basis.” (DG, 2019)
- “I see CGPs happening as a consequence of the pressure to publish. The submission of large quantities of low-quality manuscripts is motivated by a numbers-focused quantity-over-quality evaluation process for untenured faculty.” (JJ, 2019)

Ending

- “Will we be able to solve this? Maybe when we stop having people who have the wrong set of goals.” (DG, 2019)

SIGNATURE PAGE

This is to certify that the thesis prepared by Malea Dunne entitled "Computer Generated Papers as a New Challenge to Peer Review" has been examined and approved for acceptance by the Department of Technical Communication, Montana Technological University, on this 29th day of April, 2019.

Dr. Pat Munday, PhD, Professor of Science & Technology Studies
Department of Technical Communication
Chair, Examination Committee

Dr. Glen Southergill, PhD, Associate Professor of Technical Communication
Department of Technical Communication
Member, Examination Committee

Dr. Dawn Atkinson, PhD, Assistant Professor & Writing Director
Department of Writing
Member, Examination Committee

Dr. Charie Faught, PhD, HCI Associate Professor
Department of Health Care Informatics
Member, Examination Committee