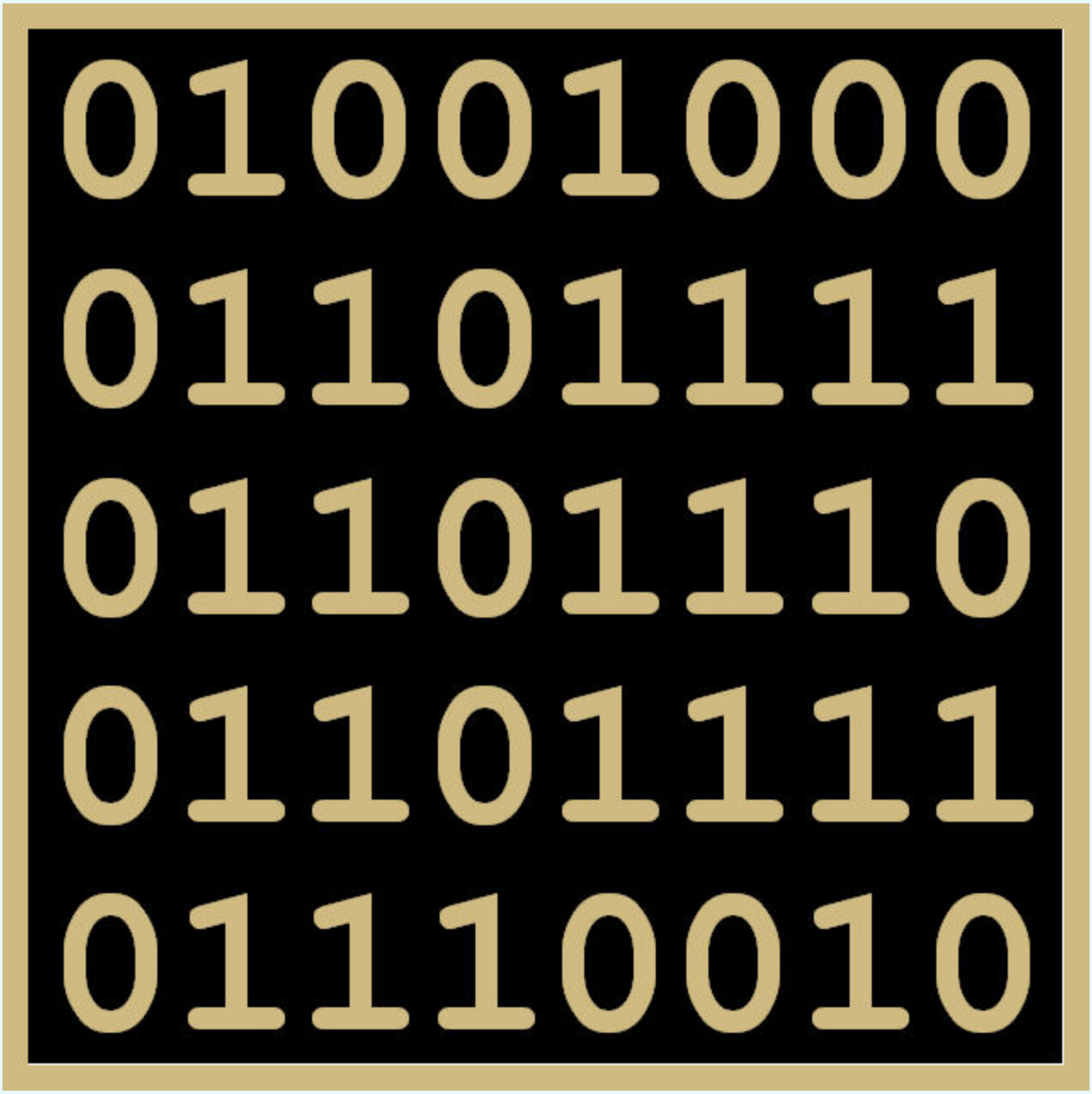
Computers & Society



August 2012
Vol. 42 I No.1

Contents

Chair's Message	4
Issue Introduction	5
The Pledge of the Computing Professional: Recognizing and Promoting Ethics in the Computing Professions	6
Gandhigiri in Cyberspace: A Novel Approach to Information Ethics	9
Using Moral Rules to Address Truth in Transition and the Demise Facts	21
Social Responsibility in the Information Society: A Potential Knowledge Gap for Tomorrow's Policy Makers	28
Deception Detection for the Tangled Web	34

SIGCAS Officers and Contact Information

Chair
Andrew A. Adams
SIGCAS E-Mail: chair_sigcas@acm.org
Other E-Mail: aaa@meiji.ac.jp

Vice Chair
Netiva Caftori
SIGCAS E-Mail: vc_sigcas@acm.org
Other E-Mail: n-caftori@neiu.edu

Executive Committee Member-at-large
Simon Rogerson
E-Mail: srog@dmu.ac.uk

Past Chair
Florence Appel
E-Mail: appel@sxu.edu

SIGCAS Computers and Society — The SIGCAS Magazine

SIGCAS Computers and Society is an online magazine accessible via the ACM Digital Library. The magazine aims to be an effective communication vehicle between the members of the group. The editors invite contributions of all types of written material (such as articles, working papers, news, interviews, reports, book reviews, bibliographies of relevant literature and letters) on all aspects of computing that have a bearing on society and culture. Relevant conference announcements and calls for papers will also be published. Submissions may be sent to any one of the three editors listed above, or to editors_sigcas@acm.org.

Readers and writers are invited to join and participate actively in this Special Interest Group. Membership is open to all, for US$20 per year, and to students for US$10 per year. The link to join up can be found on our web site, at http://www.sigcas.org .

Instructions to Authors
Writers of manuscripts are requested to observe the following guidelines:

- Electronic submission in either Microsoft Word format or RTF format;
- Harvard author-date citation and reference style;
- Author's name and email address placed below the title of the submission;
- Include several key words to facilitate retrieval from the Digital Library;
- Deadlines for submission: first day of February, May, August and November.

Copyright Notice

Chair's Message

Prof Andrew A. Adams

Chair, ACM SIGCAS

chair_sigcas@acm.org

I'd like to welcome Ben Gerber who has taken over as Newsletter Editor from this issue. He was appointed with the enthusiastic support of the Executive Committee and the Editorial Board and we're sure he will do a great job in the role.

As announced on the members email list earlier in the year we made two awards for 2011. The SIGCAS Making a Difference Award went to Oliver "Ollie" Smoot, jr. for his long service to the issue of standardisation, starting with his use by fraternity brothers as a measure of the Harvard Bridge between Boston and Cambridge in Massachusetts. The SIGCAS Outstanding Service Award went to Carol Spradling for her long and diligent work in promoting SIGCAS issues in the ACM computing curriculum development and otherwise representing SIGCAS on the ACM Education Council. A call for nominations for the 2012 Awards will be coming out in the autumn, so please keep an eye out for that and think about suitable recipients.

Vice-Chair Netiva Caftori and Past Chair Florence Appel ran a well-received pre-conference workshop at SIGCSE in March in Raleigh, NC, aided and abetted by long-standing SIGCAS member Don Gotterbarn who ran the other half of the day on behalf of ACM COPE. We're planning a similar event at SIGCSE 2013 as well as a Birds of a Feather session. We're also looking at expanding our activities to the SIGITE conference next year. Please do get in touch with the committee if you are attending either of these conferences and would like to help out or even just make suggestions for the program.

We're starting the process of redeveloping our web and other online presence. This has taken a long time to get going, but we're now starting up. We've got a number of members who have volunteered to help the committee develop and maintain this, but the more people involved, the better these things tend to be, so please do get in touch if you'd like to help. More information will be sent out to the members mailing list as we make progress. Our primary goal in developing this new site will be to enhance the ease and options for member-member interaction.

Issue Introduction

Ben Gerber
Editor
editors_sigcas@acm.org

Welcome to Volume 42 and the 42nd year of *SIGCAS Computers & Society*!

I would like to thank Andrew Adams, Netiva Caftori, Florence Appel (whose shoes are hard to fill!), the Executive Committee, and the Editorial Board for their confidence in appointing me Editor, as well as for their assistance in putting together this issue. I would also like to thank Don Gotterbarn for his extensive work in lining up a number of the submissions, and for himself contributing a thoughtful article.

This August issue is full of intriguing articles we hope you will enjoy. We are particularly pleased to be bringing you a truly fantastic advancement in the computing profession: "The Pledge of the Computing Professional: Recognizing and Promoting Ethics in the Computing Professions." We were sad to hear that one of our authors, William (Bill) Albrecht, passed away. Bill will be missed and remembered by family, friends, colleagues, and the innumerable students whose lives he touched.

Bill's colleagues (and co-authros), Ken Christensen, Venu Dasigi, Jim Huggins, and Jody Paul, shared with us a memorial note:

> It is with great sadness that we write this note. William (Bill) Albrecht tragically passed away on August 22, 2012 at age 51. Bill is survived by his wife, three children, sister, and uncle. Bill was an Associate Professor and Assistant Head in the Department of Mathematics, Computer Science, and Statistics at McNeese State University. Bill was also one of the founding members of *The Pledge of the Computing Professional*. *The Pledge* is a new organization intended to promote and recognize the ethical and moral behavior of graduates of computing-related degree programs as they transition to careers of service to society. Bill played a leading role in the formation of this organization. We will all miss Bill's commitment to Computer Science education and his leadership of *The Pledge*. Bill was the first author of the article "The Pledge of the Computing Professional: Recognizing and Promoting Ethics in the Computing Professions" that appears on page 6 of this issue.

The Pledge of the Computing Professional: Recognizing and Promoting Ethics in the Computing Professions

Bill Albrecht
McNeese State University, Lake Charles, LA
Ken Christensen
University of South Florida, Tampa, FL
Venu Dasigi
Bowling Green State University, Bowling Green, OH
Jim Huggins
Kettering University, Flint, MI
Jody Paul
Metropolitan State University of Denver, Denver, CO

Introduction

All of us in the computing community understand the importance of recognizing and promoting ethical behavior in our profession. Instruction in ethics is rapidly becoming a part of most computing-related curricula, whether as a stand-alone course or infused into existing courses. Both Computing Curricula 2005 and the current discussions on Computing Curricula 2013 recognize the significance of ethics, generally considering it a core topic across the various computing disciplines. Additionally, in their criteria for the accreditation of computing programs, ABET specifies that a student must attain by the time of graduation an understanding of ethical issues and responsibilities. What has been missing is a formal rite-of-passage ceremony to prompt student recognition and self-reflection on the transition from being a student to a computing professional. In 2009, seventeen faculty members and industry representatives from a wide range of institutions began to address this open problem by forming *The Pledge of the Computing Professional* [1], [2]. The Pledge exists to promote and recognize the ethical and moral behavior and responsibilities in graduates of computing-related degree programs as they transition to careers of service to society. The Pledge does not seek to define or enforce ethics – this is the role of other organizations. Specifically, The Pledge is modeled after the Order of the Engineer [3] and provides a rite-of-passage ceremony at the time of graduation.

A guest presentation by five members of The Pledge organizing committee – the authors of this short article – was given at the annual SIGCAS meeting at the 2012 SIGCSE conference held in Raleigh, NC in February/March. This short article summarizes this presentation. In this article, we describe the mission and history of The Pledge and its ceremony. We also describe what The Pledge is not, discuss possible future directions, and finally explain how to become a node (i.e., a chapter) in the organization.

The Mission of The Pledge

The mission of The Pledge of the Computing Professional is to create and administer a rite-of-passage ceremony for graduates of computing-related programs from academic institutions. The sole intent is to promote and recognize the ethical and moral behavior in these graduates as they transition to careers of service to society. There has been an increased awareness of ethics in society given recent high profile events such as the 2008 financial collapse, WikiLeaks, charges of falsified climate data, and the News International phone hacking reports. Thus, a proactive view on ethics is both important and timely. A rite-

of-passage ceremony is part of many professions, as it symbolizes the entry into a position of ethical and moral responsibilities. For example, nursing has a capping ceremony, doctors have the Hippocratic Oath, and Engineers have a ring ceremony. However, while computing, through ACM, has a Code of Ethics, this code is not adopted by graduates in any type of ceremony, nor does ACM have anything in place to formally symbolize the transition from student to practicing professional. Computing has matured from its origin as a subfield of mathematics or electrical engineering into a field in its own right, and has actually spawned many subdisciplines, such as software engineering, information technology, information systems, computer game design, information security, and others. Schools and Colleges of Computing are not uncommon today at many universities. Given the ever-increasing reliance placed upon software, graduates from computing-related degree programs need to be more aware than ever of their responsibility toward insuring that society is well-served through their creative works, hence the desire to develop an appropriate rite-of-passage ceremony.

The effort to define a new organization was initiated via a solicitation on the SIGCSE mailing list in March 2009. From this solicitation seventeen faculty members and industry practitioners joined an effort that led to formation of The Pledge. The first ceremonies were held in May 2011 at Ohio Northern University, University of South Florida, Metropolitan State College of Denver (now Metropolitan State University of Denver), and at McNeese State University. As of March 2012, four additional institutions have held ceremonies and several other institutions are planning ceremonies for the near future. The ceremonies have been held as part of Departmental and/or College pre-graduation ceremonies. At both the University of South Florida and Ohio Northern University, The Pledge ceremony is held with the Order of the Engineer ring ceremony. The Pledge ceremony entails a faculty member describing the history and importance of the Pledge. This is followed by the graduates standing-up and publically reciting the oath. Following the recitation of the oath, the students are individually recognized (for example, by walking across a stage) and presented a pin with the symbol of The Pledge. The students then sign a certificate containing the oath that they recited. The arrangement of the ceremony is intended to be flexible. The sidebar contains the oath and a picture (and explanation) of the symbol.

The Pledge of the Computing Professional

I am a Computing Professional.

My work as a Computing Professional affects people's lives, both now and into the future.

As a result, I bear moral and ethical responsibilities to society.

As a Computing Professional, I pledge to practice my profession with the highest level of integrity and competence.

I shall always use my skills for the public good.

I shall be honest about my limitations, continuously seeking to improve my skills through lifelong learning.

I shall engage only in honorable and upstanding endeavors.

By my actions, I pledge to honor my chosen profession.

The symbol is a matrix with the word "Honor" encoded in ASCII. The symbol is part of a lapel pin presented at the rite-of-passage ceremony.

We believe that inclusiveness is very important; the ideals of The Pledge reflect this. Graduates from any computing-related program at any type of post-secondary institution are eligible to join The Pledge. We

invite all institutions in the US and world-wide to become chapters (or "nodes") of the Pledge. The Pledge is thus to the computing professions what the Order of the Engineer is to the engineering professions – it is a rite-of-passage ceremony for computing graduates and an entry into (and recognition of) the profession. The Pledge is a means of validating the importance of ethics in the computing professions for all graduates of all computing programs.

What The Pledge is Not

Organizations are typically defined by what they are and what they do, but we believe that it is also important for an organization to understand what it is not. The Pledge is not intended to compete with ACM, IEEE-CS, etc. in any way. It has nothing to do with certification or licensure. It is not an honor society; membership is open to any graduating student. It is not exclusive to only specific sub-areas of computing; membership is open to graduates of all computing disciplines. It is not exclusive to only certain types of institutions; two-year, four-year, and graduate institutions are all welcome. And, very significantly, it is not a replacement for the ACM or IEEE-CS Code of Ethics; it is an affirmation of these codes. We do not seek to compete with any existing organization, but rather to fill a hole that is not addressed by any other organization or community. We do not take positions on what constitutes ethical behavior, nor do we seek to enforce ethical behavior among those who take The Pledge. These are positions and roles best handled by other organizations and communities.

Future Directions

Believing that The Pledge fills an important hole in the computing community, we seek for it to grow. We would like more institutions to adopt The Pledge ceremony for their graduates. We would like greater involvement from the computing community in determining future directions to better meet the mission of The Pledge. One possible direction might be to become embraced by ACM, under the auspices of SIGCAS, in somewhat the same way that the computing honor society – Upsilon Pi Epsilon – is now a part of (or under the umbrella of) ACM. The Pledge may be a means for increasing awareness of the ACM Code of Ethics and membership. The presentation made at the 2012 SIGCAS annual meeting and this article is a first exploratory step in this direction.

How to Become a Node in The Pledge

If you have read this far and you teach computing, you may have an interest in bringing The Pledge to your students. Instructions on how to join are given on the website:

`http://www.computing-professional.org`

There is a one time cost of $50 for chartering a node at an institution. The cost to a student is $10 (also one time) and covers the cost of the pin and certificate. There are no membership dues or additional costs. The script for a ceremony can be found on the website. Links to two YouTube videos of model ceremonies performed in 2011 at Ohio Northern University and University of South Florida are also on the website.

References

[1] J. Estell and K. Christensen, "The Need for a New Graduation Rite-of-passage," Viewpoints Column, Communications of the ACM , Vol. 54, No. 2, pp. 113-115, February 2011.

[2] The Pledge of the Computing Professional, 2011. URL: http://www.computing-professional.org/.

[3] Order of the Engineer, 2011. URL: http://www.order-of-the-engineer.org/.

Gandhigiri in Cyberspace: A Novel Approach to Information Ethics

VAIBHAV GARG
Indiana University

L. JEAN CAMP
Indiana University

The interpretation of the terms 'information' and 'ethics' is often culturally situated. A common understanding is contingent to facilitating dialogue concerning the novel ethical issues we face during computer-mediated interactions. Developing a nuanced understanding of information ethics is critical at a point when the number of information and communication technology (ICT)-enabled interactions may soon exceed traditional human interactions. Utilitarianism and deontology, the two major schools of ethics are based in a western perspective. We contribute to the existing discourse on information ethics by arguing for the inclusion of Gandhian notions of non-violence and confrontation. These are particularly relevant to cyberspace, which does not always lend itself to coercion due to legal, political and economic limitations. We address the applicability of ahimsa, satyagraha, and swaraj to cyberspace. We discuss a Gandhian approach to system design. Finally, we use case studies to illuminate the application of Gandhian notions as well as their limitations.

Introduction

We are perennially surrounded by pervasive and ubiquitous technology. Our interactions not only with each other but also with the components of the physical world are mediated by ICTs. Here, we refer to the ICT-enabled elements of our lives as occurring in cyberspace [Vlahos 1998; Floridi 2002]. Cyberspace is a place to earn a living, to enjoy life, to fulfill duties, and to obtain knowledge. Capurro [2008] frames the term digital ontology to describe the pervasive nature of digital technology in all dimensions of our existence. Bijker [2006] discusses the vulnerability inherent in technology that is not only desirable but is also an essential component of innovation. He argues that to analyze this vulnerability it is important to take a cultural perspective. A transcultural consensus requires a dialogue between and across cultures [Capurro 2008]. An example of such transcultural acceptance is in a Gandhian framework.

There are two major schools of thought in western ethics: utilitarian and deontological. Utilitarianism judges the moral worth of an action based on its ability to maximize utility for the actor. An action, which maximizes utility for some, might operate in a tyrannical mode by restricting the distribution of (positive) utility to specific sections of society while simultaneously denying benefits to others. Pareto improvement is seldom encountered beyond theoretical economic models. Deontology is a duty-based framework of ethics. To act morally the actor must act a priori from deon (duty) or what ought to be done. Thus, as long as the actor acts out of good will the act would be considered moral irrespective of the nature of the outcome. Current codes of ethical conducts such as ACM [Anderson, 1992] or IEEE, reflect the problems and tensions [Harrington, 1996] inherent to these approaches.

A Gandhian approach builds upon the Western ethical framework through the inclusion of the Vedantic philosophy of Hinduism[1]. Vedantic philosophy advocates a balanced pursuit of the four purusharthas (life goals): dharma (duty), artha (wealth), kama (pleasure), and moksha (enlightenment) [Parel 2006]. Ethical dilemmas can in practice and principle be reduced to a conflict between the four purusharthas. A Gandhian resolution is attained by finding a balance between the four purusharthas. Our task is to analyze this process of balancing and adapt it to information ethics.

We begin with a background of Gandhian philosophy and the Vedantic ideas that form its meta-theoretical foundation. We then examine the translation of Gandhian philosophy to cyberspace and demonstrate its suitability to information ethics on a culturally and politically diverse Internet. We do not argue that notions of non-violence and confrontation are novel to information ethics. We do, however, feel that the primary treatment of the subject has been through Western ethical framework. We argue for the expansion of the discourse by introducing Gandhian notions that spread across both western and eastern perspectives, not only in application but also in its inspirational grounding.

In the next section we discuss related work. Then we provide a brief introduction to a Gandhian approach and its key components: ahimsa, satyagraha, and swaraj. We then situate these Gandhian constructs in cyberspace. We discuss case studies to examine the application and limitations of a Gandhian approach. Furthermore, we note the limitations of this work and some of the open questions that need to be addressed in the future. Finally, we conclude.

Related Work

Maner [1980] coined the term "computer ethics". Computer ethics is "the analysis of the nature and social impact of computer technology and the corresponding formulation and justification of policies for the ethical use of such technology" [Moor 1985]. Floridi [1999], however, argues for information ethics over computer ethics. He says that computer ethics is difficult since traditional ethical theories are not easily adapted to computer ethics. In contrast to computer ethics, information ethics can be treated as a special case of environmental ethics or ethics of the cyberspace. Similarly, we argue that ethics in the cyberspace can also be viewed as a case for Gandhian discourse. Instead of focusing primarily on western philosophies (e.g. Rawls [1971]) here we will focus on the application of western philosophical frameworks to the cyberspace.

We face ethical dilemmas due to ever changing technology and the increasingly pervasive nature of computing [Moor 1998]. Johnson [1997] argues that the only solution is that we categorize and internalize ethical behaviors in the online world just as we have for the offline one. She also argues that online ethical issues themselves are not very different from ethical issues offline. Thus we can use the existing tools in

1 Gandhi's ideas were influenced by, amongst others, Western thinkers such as Thoreau and Ruskin, as well as Vedantic philosophy, which originated in India. This work concentrates on the Vedantic roots of Gandhian thought as we present an eastern, in particular Indian, ethical perspective on information ethics in the context of cyberspace.

Author's addresses: V. Garg and L. J. Camp, School of Informatics and Computing, Indiana University

ethics and apply them to the cyberspace. This work is essentially in agreement with that perspective and only presents a case that a Gandhian framework would facilitate these distinctions.

Lessig [1999] argues that there are differences between the physical space and cyberspace, which makes the latter both potentially more or less regulable. He notes that, unlike in the offline world where the architecture is defined by physical constraints, the online world offers a wider range of choices. Further, those choices, once embedded in the cyberspace, cause risk to become immoveable, invisible, ubiquitous and self-enforcing. His view is complementary to a Gandhian perspective on system design.

Nissenbaum [2001] at the same time suggests that the architecture of the system itself can hold values and thus can demarcate between the ethical and the unethical. She considers a dichotomy of controversies. The first deals with the aspect of social change and second deals with the underlying value system. The former deals with the idea of accountability and responsibility, the latter deals with a more radical change. An example would be a change in the way we think about privacy in light of large-scale data aggregation and data mining. She also says that this change goes both ways: at the same time that technology alters value systems, values also guide the evolution of technology. She suggests that these controversies can be resolved only through dialogue between the engineering community and those who study value systems and calls it 'engineering activism'. A Gandhian approach, as we will see later, facilitates this dialogue.

Electronic Civil Disobedience (ECD) [Ensemble 1994], similar to the above, actively attempts to counter those decisions that are seen as unethical. It is a reflection of how activists are embracing technology and hackers are becoming politically motivated [Wray 1999]. Since the capital is now mobile and electronic, the resistance must also have the same characteristics and adapt itself for cyber-activism. ECD has, however, evolved into hacktivism [Manion and Goodrum 2000], a more radical idea that is not grounded in resolution through dialogue and is seen by some as terrorism [Furnell and Warren 1999]. This classification makes it far less effective. It has been argued that hacktivism is not criminal behavior since criminals seek to profit from damage to individuals whereas hacktivists only target institutions [Ensemble 1994]. Hacktivists, however, may use technologically extreme measures and their actions may effect individuals who have no stake in the conflict [Furnell and Warren 1999].

Civil disobedience was also used by Gandhi but in a much different form, as we will see later in the case studies. His approach is more in line with Johnson [1997] in that it tries to internalize the norms of ethical behavior. At the same time Gandhi differs in not being an element of the school of philosophy that came out from the industrial revolution in the west but is instead an eastern outlook [Pantham 1983]. He encourages decentralization [Pantham 1983], which is a tenet on which the Internet has evolved and prospered [Post 2000] and whose benefits are well known [Brafman and Beckstrom 2006].

Gandhigiri: Ahimsa, Satyagraha, and Swaraj

Gandhi's philosophy is not only based on the lack of physical coercion but in fact rejects it under all conditions, even self-defense. Gandhian arguments therefore cannot apply to the coercion of the unwilling that underlies a solution based in jurisprudence. Cyberspace crosses all extant jurisdictional boundaries. That physical coercion is rarely possible over the network argues for the value of a Gandhian perspective. Another advantage is that Gandhian ideas have been proven in a diverse range of cultures. For example, this philosophy has contributed to the transformation of society in the United States of America and South Africa as exemplified by the popularity and success of mass movements led by Dr. Martin Luther King, Jr. and Nelson Mandela respectively. The Chipko movement also adapted Gandhian strategies of protest

to fight environmental degradation in India [Hardiman 2003]. These examples indicate the relevance and applicability of Gandhian principles to different cultures.

Another reason for applying a Gandhian framing to the analysis of conflicts in the cyberspace is the more dialogical[2] than monological[3] approach that Gandhi takes. For example, his most important work, Hind Swaraj [1909], is written in the style of a conversation rather than a set of declarations. A monological approach allows the reader to reflect on internalized ethical values [Nijhof et al. 2000] at the risk of reduced flexibility. Adaptability is important in the cyberspace not only because of its current dynamic nature but also because of the uncertainty in its future development. Gandhian ideology is flexible and subject to growth.

There are three essential components of Gandhian philosophy: ahimsa, satyagraha and swaraj. Ahimsa is the journey, satyagraha is the path, and swaraj the destination. Before we can see how these terms translate in the cyberspace, we first need to understand these components in the context of his time.

Ahimsa

Literally translated, 'ahimsa' means 'non-violence'. Gandhi believed in non-violence. There are three aspects to Gandhi's conceptualization of non-violence. First, he believed that true force is not brutish. It is the strength of an individual's mind. Second, while ahimsa as a notion is common in eastern religions, but Gandhi's ahimsa differed from that preached by the Buddhists or the Jains. Gandhi believed that true ahimsa does not mean just accepting whatever persecution is brought upon us. He believed in action. He believed that if the persecutor can be shown how they cause suffering and how their victims bear it without complaint, the persecutor would have a change of heart and would eventually realize the folly of his actions. Finally, Gandhi believed that true ahimsa meant standing up, not only to the injustices done to oneself, but also those done to others. Such meaningful suffering requires partaking in their resistance leading to the idea of satyagraha.

Satyagraha

Satyagraha is ahimsa in action. Formed from two words, 'Satya' meaning truth and 'Agraha', translated as force, the word satyagraha, literally translated, means the force of truth. It is a peaceful form of civil resistance. Satyagraha enables not only the strong but also the weak, because it calls on mental strength instead of physical strength. This is not a new notion. There are examples of it being used long before Gandhi introduced it into the Indian political movement. In fact, these precedents and their success may have convinced him of the power of this movement. Some examples of these protests that Gandhi came across are: passive resistance by the Hungarian nationalists against the Hapsburgs from 1849 to 1867 and by Sinn Fein against British rule in Ireland during the early years [Gandhi 1907].

Apart from examples abroad, the Indian social structure, especially the Gujarati one (to which he belonged) was filled with examples of non-violent resistance. An example would be the practice of traga: threats of self-harm to motivate others [Hardiman 2003].

Swaraj

The idea of swaraj can be literally translated as self-government. Some find swaraj, the greatest good for all and not just the greatest number, to be self-contradictory. This is because most people believe that there

2 A dialogical approach uses a Socratic style question and answer format between the fool and the wise man.

3 A monological approach is presented in a style in which there are no interruptions and ideas are conveyed as declarative statements.

is no such optimization. The idea of swaraj can be understood from Gandhi's seminal work called Hind Swaraj [Gandhi 1909]. He treats swaraj as an open-ended question and does not give a specific definition. He does, however, define a few characteristics.

The first characteristic that swaraj must have is the idea of the governing body as servants. Those who govern are there to serve the people and not to rule them (e.g. public servants). Swaraj can not be achieved by giving up on what is old, what is not swaraj, since swaraj is a dynamic which builds upon its own self, each time becoming better and more wholesome [Gandhi 1909]. Swaraj requires peer production through collective action. Swaraj is a process and not an end point. Thus, the fact that at one time there is no optimal solution, which provides the greatest good for all, does not mean that the process should cease at that moment.

The Cyberspace Gandhian

In this section we translate Gandhian ideas of ahimsa, satyagraha and swaraj to the cyberspace. The application of Gandhi to cyberspace is straight forward. Cyberspace is culturally diverse and crosses myriad jurisdictional boundaries. It is extremely difficult, and arguably potentially undesirable, to enforce laws that transcend these boundaries and ensure ethical behavior. Gandhi does not advocate forcing ethics onto people but instead inculcating them.

Gandhi's conception of swaraj [Gandhi 1909] is similar to Floridi's notion that information, being a source of nourishment, should be available to all [Floridi 2005]. It is necessary for a person to fulfill his/her purusharthas, for example a person would not be able to fulfill his or her duty (to the best of their potential) if the information access is restricted [Wagner 2003; Stallman et al. 2002]. In that sense Gandhian ethics would also incorporate disclosive ethics [Introna 2007].

The idea of Gandhian cyberspace can also be explained using his idea of Oceanic Circles [Parel 2008]. Oceanic Circles refer to a strong unified society that would be free from violence and aggression. Gandhi described the current society as a pyramid, i.e. a hierarchy based society. Every step is built upon a lower step, thus irrevocably suppressing the one beneath. Some argue that the Internet and related network technology is based on the similar principles of control and hierarchy [Froomkin 1997]. For example, access control is basically one account (Administrator) having all the privileges and others having various levels of privileges assigned by the administrator [Zittrain 1996]. In an organization this might lead to resource starvation if lower ranked processes are denied access to resources. The alternative is an Oceanic Circle approach.

"Life will not be a pyramid with the apex sustained by the bottom. But it will be an oceanic circle whose center will be the individual always ready to perish for the village, the latter ready to perish for the circle of villages, till at last the whole becomes one life composed of individuals, never aggressive in their arrogance, but ever humble, sharing the majesty of this oceanic circle of which they are integral units" [Parel 2008].

This idea is based on the notion of completely independent and self-sustainable villages. They must be capable of taking care of their affairs including defending themselves. Each village, in itself a circle with certain area of influence, will be flanked by other villages who would be equally powerful, thus drawing a larger (oceanic) circle of influence. Thus the process is repeated, every time drawing a larger circle. Since the outer circle derives power from the inner ones it would not wield the power to crush the inner circle but will give strength to all within and derive its own strength from it (Gandhi, 1946). It is easy to see how Oceanic circles would correspond to the Internet. The individual is the node that forms villages

or subnets that join together to form bigger subnets and thus a bigger oceanic circle.

Peer to peer technologies and open source movements [Stallman et al. 2002] run parallel to the idea of oceanic circles. In Gandhi's words they allow "free and voluntary play of mutual forces". Such a society is necessarily highly cultured in which every man and woman knows what he or she wants and, what is more, knows that no one should want anything that others cannot have with equal labor" [Parel 2008]. In the sense of peer-to-peer groups a user can't simply download songs unless they contribute by sharing some songs themselves.

The concept of each village being able to defend itself also makes sense. If every subnet in the world was capable of protecting its resources then it would make the Internet a much more secure place. Most security incidents happen because individual systems are not patched[4]. Most of the time people are not even aware that they are leaving their systems vulnerable to attacks. From this comes the idea of an all-aware society leading to Swaraj. As Gandhi says "Swaraj can only be achieved through an all round consciousness of the masses" [Batra et al. 1984]. Similarly, our shared privacy can only be met if we all implement security to avoid data loss. Of course, security is a prerequisite for privacy but does not in any way guarantee it.

While Gandhi had reservations about technology, he also said, "I would prize every invention of science made for the benefit of all" [Bose 1962]. Gandhi's dislike was towards technologies that would lead to unemployment and concentration of wealth in the hands of the few. His dislike was for mass production. Instead, ICT-enabled technologies facilitate distributed production by the masses. One example of this can be seen in the music industry with increasing number of artists choosing to release their work on independent labels. ICTs, like telemedicine, are also helping people with limited resources get access to better healthcare. Several ICT enabled project have been promising in increasing education [Mitra and Rana 2001]. It may be argued that these technologies are expensive and only available to few. However, ICTs such as cellular telephones vastly increase connectivity and can increase economic opportunity [Camp and Anderson 1999]. Paradigm shifts in thinking about engineering are making production of new technology not only locally but also economically[5]. This new engineering paradigm is also based in Gandhian philosophy. Thus ICTs would, based on their usage and context, be acceptable in a Gandhian setting.

Case Studies

In this section, we provide case studies of how behaviors can be understood as ethical or unethical from a Gandhian perspective.

iPhone

Consider the Apple iPhone, which was launched in 2007. At the time it was only available on the AT&T network. Customers were also not allowed to install breakthrough designs for ultra low cost products third party software on the iPhone. Customers were not happy about relinquishing control in either of these areas. At that time, AT&T was subject to an attempt at boycott due to its role in the NSA wiretapping practices of the Bush Administration [Sugiyama and Perry 2006]. Customers also believed that since they had paid for the phone, it was their prerogative to decide whether they wanted to install third party

4 http://www.schneier.com/crypto-gram-0406.html#4, Retrieved April 12th, 2012

5 http://www.ted.com/talks/r_a_mashelkar_breakthrough_designs_for_ultra_low_cost_products.html, Retrieved on April 12, 2012

software or not.

Some unsatisfied customers hacked the iPhone to allow third party software and network service through other providers. This resulted in a cycle between Apple and hackers wherein Apple would release patches to make the hacks ineffective and hackers would break the new patch. While iPhone users were locked in to AT&T for a long time, they got a more immediate respite for third party apps. Within a few months AT&T announced an SDK allowing users to develop third party applications, effectively allowing them to install third party software on the iPhone.

This is an instance of how Satyagraha can be practiced online. It runs parallel to the Dandi March [Weber 1997] undertaken by Gandhi to protest against the salt tax imposed by the British government. Salt is an important ingredient in the Indian diet. A lot of people, especially the ones near the coastal areas would simply make salt from the sea. The British government, however, made it illegal for anyone to manufacture salt except the government. Gandhi appealed to amend the salt law but to no avail; Gandhi and his supporters openly broke the law by making salt at Dandi. This was a widely publicized event and a large number of people were arrested. Eventually the British government gave in and called Gandhi for talks.

There are many possible points of illumination with the iPhone story. The disobedience was about economic self-control. The actions of the protestors were non-violent. The actions were widely publicized and the authorities were aware of the mechanisms of the protest. The authorities escalated the measure to curb protest AT&T locked the phones via upgrade but the protestors did not. Disinterested customers were not affected by the actions of the protestors.

Metallica

The second case study can also be illuminated using a Gandhian perspective. Metallica is a thrash metal band that came out as an offshoot of the new wave of British Heavy Metal. They were the first band to sue Napster, a file sharing system based on peer-to-peer architecture. This was not acceptable to their fan base. Metallica fans argued that the reason for Metallica's popularity was bootlegs. By suing Napster Metallica denied upcoming bands the same freedom, which gave them success. Fans revolted by destroying Metallica CDs, paraphernalia and other items in public. This, however, did not elicit any response from the band as the tactics to boycott them only brought more publicity. Eventually, though, the fans adopted a new strategy and simply stopped buying Metallica's music and concert tickets. Thus experiencing genuine revenue loss, Metallica settled out of court. The bands final statement was that they opposed Napster not because it shared their music but because Napster should have asked Metallica before doing so.

This case study draws parallels with the Non-cooperation movement. Non-cooperation was a protest against the Rowlatt Act that overruled Habeas Corpus. This was started by Gandhi and was mobilized in 1920. In the Non-cooperation movement, Indians simply quit their jobs and did not go to work. The machinery of the British government relied on Indians working. Without them the machinery lacked its foot soldiers in clerks, teachers, doctors etc. Metallica too lost its fans and thereby lost not only money and reputation but also its identity. The band was previously respected for being true to its roots, but once the movement began, Metallica became a symbol for the self-serving record companies. All this happened without any violence on the part of the protesters. They broke no law, but by not buying Metallica's music they exercised a right and reached a solution without affecting anyone who was not a stakeholder in the dispute.

Kathy Sierra

The third case is more difficult as it addresses threats of horrific violence. Kathy Sierra is a programming instructor and game developer. She was also a prominent blogger. In 2007, she stopped blogging and cancelled her appearance at the O'Reilly Tech Conference due to death threats from various sources via e-mail and blog posts. The threats forced her to make significant changes to her life. This is one case where someone's virtual world persona can have serious and potentially fatal consequences in the physical world. Similar threats using the Internet were also witnessed in the case of Students Against War (SAW) vs. Michelle Malkin. In this case Michelle Malkin reposted the names and contact information of individuals involved with SAW. This information was originally posted on the SAW website, but they decided to remove it after getting various threats. Michelle Malkin, however, did not agree to SAW request to remove this from her blog, reposting it several times and claiming that SAW needed to take responsibility. This form of cyber-harassment has been widely studied [Campbell 2005; Ellison and Akdeniz 1998; Servance 2003]. In these cases Gandhian philosophy would probably be an ineffective way to protest against the actions of the individuals involved. However, the Gandhian perspective clearly illuminates the many ethical contributions to the greater harm.

Such cases are different from the first two case studies in many ways. First, individuals are singularly targeted and even though Kathy Sierra was supported by many other bloggers, she alone would have had to bear the consequences if the threats against her were realized. Secondly, the people involved in doing harm are not the authorities but fellow bloggers and Internet surfers. Also, in the above two case studies, both Metallica and AT&T took legal recourse and exercised what were their legal rights. The consequences were mostly economic. The problem of the first two case studies is of ethics, whereas here the problem is not only ethical but also legal. When the problem becomes criminal, we have the option of using the legal framework to solve the issues. For example, in the case of Kathy Sierra an investigation was conducted by law enforcement. If the events had escalated and there were proof to show physical danger, she might have considered police protection.

There is, however, a clear contribution of Gandhian philosophy to these cases. A Gandhian would clearly identify these as unethical behaviors. From a Gandhian perspective, the use and threat of force is never justified, so those who threatened the students, those who advocated threatening them, those who threatened Sierra, and those who justified the threats as freedom of speech can all be identified as unethical actors. Justification of threats of violence, while legal, is clearly unethical. Using a Gandhian framing, such an identification of the unethical as unethical contains no threat, and thus there can be no pretense that such an identification is itself a threat. Thus a Gandhian framework could mitigate rather than escalate the situation. First, it denies unethical actors the claim of legal justification. Second, it refutes reliance on the importance of the ends to justify means. Finally it identifies the diminishing of themselves and their victims in the case of threatening violence on advocating the right to make these threats.

Open Questions

The purpose of this work is to argue that a Gandhian perspective can better enable us to distinguish ethical and unethical online behaviors. From this comes a potential increase in awareness by society and technologists of the nature of the (un) ethical. This may result in changes in technology and patterns of behavior. Understanding which changes best motivate ethical behaviors requires the study of social structure. Such study would be concerned with relationships among groups, as enduring patterns of behavior by participants in the social system are understood in relation to each other. Social patterns may become institutionalized norms or cognitive frameworks. At the same time there may constantly be new

emergent behaviors [Giddens 1984; Orlikowski 2008; DeSanctis and Poole 1994]. We, however, do not expand on how this addition of Gandhian insight to information ethics would animate technology, its artifact, or associated patterns in this work. This paper is rather intended to begin such a dialogue.

Gandhigiri also promises insights about the nature of common good in the cyberspace. Recall the discussion of utilitarian and de-ontological perspectives. In Gandhigiri both the ends and the means must be more than simply not unethical. Both ends and means must serve the common good. According to Aristotle [Nichols 1992], the common good concerns itself with the relative equality of outcomes where all citizens can flourish, similar to Gandhi. While the ideal of the common good is transcendent, how the material and manifest work of serving the common good are worked out is a matter of collective action [Limayem and DeSanctis 2000; Jankowski and Nyerges 2001]. Such collective action may also be informed by the addition of a Gandhian perspective. As such, again, this process online may influence social and technological architectures. While these questions are beyond the scope of this immediate work, we propose that the addition of Gandhigiri to structuration theory would enhance the dialogue.

Conclusion

The case studies demonstrate both the potential and the limits of a Gandhian approach. The iPhone users were allowed to install third party software but were for a long time locked in with AT&T as the solitary service provider. Allowing third party software on the iPhone was trivial, as it dealt primarily with the individual and his or her own phone. Allowing owners of hardware to migrate to another provider was a bigger problem with more stake-holders and consequences leading to contract breaches. The same, however, is true for the offline scenarios. For example in the first study, while people were allowed to make salt by the government they were not given freedom. Allowing people to make salt has smaller consequences, but giving them freedom and right to self-governance would have had much wider implications.

The last case study also shows that not all problems can be solved by a Gandhian approach in practice. It is not reasonable to demand the level of self-awareness shown by Gandhi from those who are threatened and harassed on the network. Thus, legal recourse in some cases is a critical, or potentially, even life-saving option. Gandhian philosophy does, however, contribute to the understanding of this by identifying the ethical failures of those who advocated violence, supported this advocacy, or failed to stand against them.

There are four methods to satyagraha [Pantham 1983]: Purificatory, Non-Cooperation, Civil Disobedience and Constructive Programs. The case studies cover only two. Their solutions cannot necessarily be generalized over all the myriad problems in the domain of information ethics. They do, however, provide an insight into how a resolution might be reached using a Gandhian framework and the benefits of such a framework. There is also need for a deeper more comprehensive analysis comparing the Gandhian approaches to western thought as applicable to cyberspace.

The practice of Gandhian ideals requires patience, courage, and faith in the goodness of human nature [Gandhi 1907]. It believes in an open dialogue between the involved parties. It is like Floridi advocating inclusion over discrimination [Floridi 2005]. There is direct impact of this school of thought on ethical issues including informed consent, anonymizing datasets for research, developing codes of conduct for ethical research and ethical system design. This paper does not provide a solution to these problems but argues that a Gandhian framework of ethics is a powerful source of insight, and should be brought to bear. In particular a Gandhian approach empowers the end user and allows them to confront the regulating institutions. It protects the entities not involved in the dispute from the stakeholders and

reaches a meaningful resolution by enabling dialogue. The results from the case studies reflect existing notions like net neutrality [Marsden, 2008] and indicate that this framework can accommodate existing value systems. Thus the Gandhian framework can serve to incorporate eastern value systems with western ethics and provide insights that are more global and thus more applicable to an increasingly international and multicultural cyberspace. In conclusion we agree with Johnson on the need to facilitate dialogue but argue that clearly distinguishing ethical behaviors from the unethical requires drawing on the entire global range of ethical systems [Johnson 1997].

References

ANDERSON, R. 1992. ACM code of ethics and professional conduct. Communications of the ACM 35, 5, 94-99.

BATRA, S. ET AL. 1984. The quintessence of Gandhi in his own words. Madhu Muskan.

BIJKER, W. 2006. The vulnerability of technological culture. Cultures of Technology and the Quest for Innovation, 52–69.

BOSE, N. 1962. Studies in Gandhism. Navjivan Publishing House, Ahmedabad, India.

BRAFMAN, O. AND BECKSTROM, R. 2006. The starfish and the spider: The unstoppable power of leaderless organizations. Vol. 332. Portfolio (Hardcover).

CAMP, L. AND ANDERSON, B. 1999. Grameen phone: empowering the poor through connectivity. Information Impacts Magazine, 13–27.

CAMPBELL, M. 2005. Cyber bullying: An old problem in a new guise? Australian Journal of Guidance and Counselling 15, 1, 68–76.

CAPURRO, R. 2008. Intercultural information ethics: foundations and applications. Journal of Information, Communication and Ethics in Society 6, 2, 116–126.

DESANCTIS, G. AND POOLE, M. 1994. Capturing the complexity in advanced technology use: Adaptive structuration theory. Organization science, 121–147.

ELLISON, L. AND AKDENIZ, Y. 1998. Cyber-stalking: the regulation of harassment on the internet. Criminal Law Review 29, 29–48.

ENSEMBLE, C. 1994. Electronic civil disobedience. Critical Art Ensemble.

FLORIDI, L. 1999. Information ethics: on the philosophical foundation of computer ethics. Ethics and information technology 1, 1, 33–52.

FLORIDI, L. 2002. On the intrinsic value of information objects and the infosphere. Ethics and Information Technology 4, 4, 287–304.

FLORIDI, L. 2005. Information ethics, its nature and scope. ACM SIGCAS Computers and Society 35, 2, 3–3.

FROOMKIN, A. 1997. Empire strikes back, the. Chicago Kent Law Review 73, 1101.

FURNELL, S. AND WARREN, M. 1999. Computer hacking and cyber terrorism: The real threats in the

new millennium? Computers & Security 18, 1, 28–34.

GANDHI, M. K. 1907. Benefits of passive resistance. Indian Opinion 7, 183-185.

Gandhi, M.K. 1909. Hind Swaraj or Indian home rule.

GIDDENS, A. 1984. The constitution of society: Outline of the theory of structuration. University of California Press.

HARDIMAN, D. 2003. Gandhi in his time and ours. Orient Blackswan.

HARRINGTON, S. 1996. The effect of codes of ethics and personal denial of responsibility on computer abuse judgments and intentions. MIS Quarterly 20, 3, 257-278.

INTRONA, L. 2007. Maintaining the reversibility of foldings: Making the ethics (politics) of information technology visible. Ethics and Information Technology 9, 1, 11–25.

JANKOWSKI, P. AND NYERGES, T. 2001. Geographic information systems for group decision making: towards a participatory, geographic information science. CRC.

JOHNSON, D. 1997. Ethics online. Communications of the ACM 40, 1, 60–65.

LESSIG, L. 1999. Code and other laws of cyberspace. Basic books.

LIMAYEM, M. AND DESANCTIS, G. 2000. Providing decisional guidance for multicriteria decision making in groups. Information Systems Research 11, 4, 386–401.

MANER, W. 1980. Starter kit in computer ethics. Originally self-published by the author in 1978.

MANION, M. AND GOODRUM, A. 2000. Terrorism or civil disobedience: toward a hacktivist ethic. ACM SIGCAS Computers and Society 30, 2, 14–19.

MITRA, S. AND RANA, V. 2001. Children and the internet: Experiments with minimally invasive education in india. British Journal of Educational Technology 32, 2, 221–232.

MOOR, J. 1985. What is computer ethics? Metaphilosophy 16, 4, 266–275.

MOOR, J. 1998. Reason, relativity, and responsibility in computer ethics. Computers and Society 28, 14–21.

NICHOLS, M. 1992. Citizens and statesmen: a study of Aristotle's Politics. Rowman & Littlefield Pub Inc.

NIJHOF, A., FISSCHER, O., AND LOOISE, J. 2000. Coercion, guidance and mercifulness: The different influences of ethics programs on decision-making. Journal of Business Ethics 27, 1, 33–42.

NISSENBAUM, H. 2001. How computer systems embody values. Computer 34, 3, 120–118.

ORLIKOWSKI, W. 2008. Using technology and constituting structures: A practice lens for studying technology in organizations. Resources, Co-Evolution and Artifacts, 255–305.

PANTHAM, T. 1983. Thinking with Mahatma Gandhi: beyond liberal democracy. Political theory 11, 2,

165– 188.

PAREL, A. 2006. Gandhi's Philosophy and the Quest for Harmony. Cambridge University Press.

PAREL, A. 2008. Gandhi and the emergence of the modern Indian political canon. The Review of Politics 70, 01, 40–63.

POST, D. 2000. Of black holes and decentralized law-making in cyberspace. Vand. J. Ent. L. & Prac. 2, 70.

RAWLS, J. 1971. A theory of social justice.

SERVANCE, R. 2003. Cyberbullying, cyber-harassment, and the conflict between schools and the first amendment. Wisconsin Law Review, 1213.

STALLMAN, R., GAY, J., AND LESSIG, L. 2002. Free software, free society: selected essays of Richard M. Stallman. Free Software Foundation Boston, MA.

SUGIYAMA, T. AND PERRY, M. 2006. NSA domestic surveillance program: An analysis of congressional over-sight during an era of one-party rule, the. U. Mich. JL Reform 40, 149.

VLAHOS, M. 1998. Entering the Infosphere. Journal of International Affairs 51, 2.

WAGNER, R. 2003. Information wants to be free: Intellectual property and the mythologies of control. Columbia Law Review 103, 995.

WEBER, T. 1997. On the salt march: The historiography of Gandhi's march to Dandi. HarperCollins Publishers India.

WRAY, S. 1999. On electronic civil disobedience. Peace Review 11, 1, 107–111.

ZITTRAIN, J. 1996. Rise and fall of sysopdom. Harvard Journal of Law & Technology 10, 495.

Using Moral Rules to Address Truth in Transition and the Demise Facts

Don Gotterbarn
Professor Emeritus, East Tennessee State University

Abstract
Many elements in current social media have led to the technological devolving of the concept of truth. It is argued that some of these problems can be mitigated by the application of moral "Rules".

Introduction
Twitter and other forms of social media have exacerbated confusion about the nature of truth and have had a negative impact on the relation between many of the elements of society. The Internet has generated some significant modifications in the understanding of truth and information. Looking at these transitions and how they are compounded by Twitter shows some ethical difficulties in addressing abusive inaccurate claims made about individuals and corporations. Corporate social media policies would be improved by incorporating ethical principles.

"Truth" in Transition
Internet 0 to 1.0
The upsurge of social media has been so rapid that there have been many changes from the early stages of the Internet. These changes raise significant ethical issues that have not been addressed. In Internet 1.0 there were several concurrent concepts of "Truth". Normally when people speak of 'truth' they have a correspondence theory of truth in mind; in which a statement is true when it bears a direct correspondence to some fact or state of affairs in reality. Rather than revisit the philosophical debates about this theory, I use it as a point of departure to look at some variations to this common sense meaning of truth; changes I believe which are encouraged by the development of the Internet and social media.

Evolving Truth
Prior to the Internet there was a kind of evolving truth related to developing one's identity. This is seen in the belief that individuals can mature and change and things that were "true of their personalities" at one point may no longer be true. The notion of defining one's self in existentialism and Hindu texts assumes this kind of "truth". In this context 'current truth' and 'future truth' are more important than 'past truth'. With this sense of truth in mind Mayer-Schönberger [2010] and Kioshi Murata [2011] talk about the importance of forgetting. Forgetting is important to human development and the ability to move on from past mistakes. Internet 1.0 had been criticized because it does not forget and casts our digital past in concrete. This "externalisation of human memory…which is continually updated by 24/7 electronic surveillance systems" prohibits forgetting the past [Murata]. This externalized data/"memory" was awarded a certain level of credence- truth over people's verbal claims about their past. This digitally memorialized past means that any error of your past (recorded event on the Internet) colors your future.

This imposes the correspondence theory of truth over any "evolving truth"

A simplistic description of the transition from Internet 1.0 to 2.0 as "being a transition to a system where most of the information is user generated" misses some significant ethical issues. The rapid changes in technology facilitated the easy importation of data to the Internet. Many of our actions are now recorded by some digital media and quickly become universally accessible through social media. This is true not just for our current behavior but technology is expanding to make it possible for the Internet to serve as a complete externalized memory for all behavior in the future (Maturata, 2011). This external memory is also being used to digitize recollections of earlier events that had been observed but not necessarily recorded. The previously unrecorded information is made available through audio histories- people recoding their (perhaps accurate) memories and thereby turning them into absolute permanent truths. The veracity of these documents is being elevated by calling them "Oral-History" rather than "tales from the past". Mayer-Schönberger points out how this restricts our ability to remake ourselves and emphasize different portions of our identities in different contexts. Our digital persona is no longer as malleable as was our physical persona.

But, paradoxically, this transition also has some negative effects on the concept of "truth" and how our understanding of it affects behavior. There was some presumption of correspondence truth in Internet 1.0. The optimism about truth of content was carried into Internet 2.0. The attitude was that truth could be 'user generated' and was more trustworthy then corporate information on the web. For example user product reviews were considered more trustworthy than reviews provided by professional product reviewers.

Self-Correcting truth: Wikipedia

This concept morphed into a belief in that the user self-governing Internet would generate self-correcting truth. It is based on participatory information sharing in the development of "true" contents. In this the description is modified to more closely approximate the fact; the description changes unlike in 'evolving truth' where the object of the truth changes. This concept of self-correcting truth is exemplified in Wikipedia. Users are asked to be bold writing down the basics for an article anticipating its later correction by a broad base of readers. This contrasts with the Encyclopedia Britannica model of initial careful writing and reviewing by a dedicated cadre of professionals editors used to monitor its objective content.

Wikipedia started with a belief in the self-correcting wisdom of the crowd, but it also recognized the problems with the Internet forgetting so Wikipedia allows correction of over emphasis on past mistakes by allowing individuals to petition to have past transgressions made less prominent. Of course such after the fact removal is too late. Once something is on the Internet there is no way to be sure it is removed [Gotterbarn 2009].

Manipulating self-correcting truth:

There are difficulties with the optimism of 'self-correcting truth'. The 'self' who is correcting the article may not be motivated by an interest in truth. All users even those with special interests can and have rewritten Wikipedia articles. A computer at Exxon Mobil made substantial changes to a description of the 1989 Exxon Valdez oil spill in Alaska playing down its impact on the area's wildlife [Hafner 2007]." An American politician incorrectly described a significant event in US history. Her supporters attempted to rewrite the Wikipedia entry to make it conform to the politician's version of history. "Dozens of similar examples of insider editing…Many Wikipedia entries are in a constant state of flux as they are edited and re-edited, and the site's many regular volunteers and administrators tend to keep an eye out for bias [Hafner 2007]."

The misdirected optimism in discovering the truth in a self correcting way on the web led one of Wikipedia's founders, Larry Sanger, to develop a new site Citizendium. Citizendium, the Citizens' Compendium is a wiki encyclopedia project that is expert-guided, public participatory, and requires real names of the editors. This is based on the need for some expert oversight.

The wisdom of the mob

The wisdom of the mob has become truth by mob rule

The original Wikipedia notion has been corrupted as the technology has been modified. Jimmy Wales, co-founder of Wikipedia , believed that any user should be able to change any entry and if enough users agree with the person who changed the entry, the entry becomes "true". Notice we are no longer talking about seeking a correspondence with fact but a consensus of opinion, or more accurately a majority opinion. Larry Sanger [2010] understood this change and called it an 'anarchist notion of truth'. Sanger's concerns about Wikipedia were interestingly enough rejected by some who used the number of hits on Wikipedia as evidence for its 'truth'. For some in social media, quantity is equated with truth.

The notion of a correspondence theory of truth has taken a further beating. The importance of a web page is assessed without human evaluation of the content. The content does not count. According to Google's PageRank algorithm the "value" of a page is determined by the structure of the Web, by the number of hyperlinks to that node. Those nodes with a higher Page-Rank will be returned to a search first. Google optimistically claims that this produces an unbiased result. But such an automated process falls in the same way as Wikipedia's optimism. PageRank and other ranking algorithms [Austin 2011] are used by organizations to modify the way the Internet represents the world and also represents individual reputations. Manipulation of those algorithmic systems by organizations like Reputation Defender to change the "truth" and use the internal structure of the Internet and manipulate electronic search results by multiplying links to existing ones or create new positive web pages. This does not "make the web forget, but it can make positive links return high in search results and negative sites hardly return at all. 'Truth'-location high enough in a returned search so it is read- is determined by the structure of the web, and not determines by any direct correspondence to reality. "Truth" has become a correspondence to VIRTUAL reality; or more accurately correspondence to many inconsistent virtual realities.

Notice how truth, or at least how it is represented, on the web is equated with numbers in terms of hyperlinks or people who don't modify Wikipedia entries. This numerical emphasis is formalized using a software product, a semantic engine based on LSA (latent semantic analysis), which is an advanced form of statistical language modeling which does a frequency analysis. The promoters of this product claim that this counting yields "the meaning and truth around any topic or subject, which in this case is what consumers are saying about a company's product, service or brand"[Robert, 2011]. "Truth" is in the numbers not in the facts. Removing the concept of correspondence has led to a separation of fact from truth.

On social media many people are only listening to like minded people; creating information silos. This selective attention contributes to the individual's unjustified beliefs that their position is what most people believe and is therefore true. This selective listening is encouraged and supported by Internet filtering designed to reduce the complexity of searching for information on the Internet. Algorithm ease defines the truth.

Google also uses filters which are based on your location when you make a search and filters which have been personalized based on your browsing history. These filters are not generally known about and their

standards are not public. In response to some criticisms about personalization [Schwartz, 2011] a Google representative asserts that "We do have algorithms in place designed specifically to promote variety in the results page". This automated process uses unspecified standards to put you in certain information silo and only provides information favored by that group. Based on an automated process of a non-human gatekeeper I will get a narrow band of information.

Facebook uses a filter on its friends list. Friends that we do not click on frequently will disappear from our view. Often we click on things we agree with, we then get more and more of what we agree with. Beyond the basic concerns with censorship, this automated filtering has several difficulties. Not only does it support intellectual isolation and a distortion of truth; limiting our exposure to opposing information; because of the tendency to associate truth with quantity of confirmations this puts us in an information silo convincing us that our view is correct. This is not a good foundation for truth.

Not only have these changes in our understanding of truth affected corporation but changes to the social media have had significant impact.

New Social Media

Improvements in technology (wireless communication, miniaturization, etc) and the change in our understanding of the ways we communicate generally referred to as 'social media', have caused many new and significant problems. First, convergence of computing and communication has caused some confusion. Computing and communication have been blending and raise questions like: Is the cell phone a computer, is an internet search on your phone while at work business or personal use of a computer. These technological changes have also facilitated radical changes in the acceptable use patterns of technology. Individuals are now almost in continuous contact through social media. Both the technology and its usage patterns in social media require careful ethical evaluation. Among the problems are: a failure to see that the nature of the medium sometimes significantly distorts the messages, it is wrong to transmit from some locations, the equation of degree of repetition with truth and, the failure to understand the impact of messages beyond its video screen representation.

Giving a Twittersworth

Social media raise some ethical tensions for an individual. The new technology has increased your audience; instead of gossip being one on one conversation you can now gossip with a bull horn. Your worth is calculated in the number of 'friends on your page' and the more people who listen to you or the higher the number of hits you have then the greater your worth. Your importance in social media is not determined by credentials, licenses, or experience but by popularity in terms of followers. You are valued by the number of tweets and followers of your every tweet.

Tweeted truth

Oddly this generates a tension between your 'value' - tweet count or number of friends- and the 'veracity' of what is said. Problems with the accuracy and impact of tweets are beginning to be recognized. The new media requires and is developing standards to evaluate the content versus the number of times it has been repeated. There are web sites and standards developed by journalist to help substantiate the content of tweets. The Canadian Association of Journalists has tweeting guidelines. There are recommendations for what and how to re-tweet.

Mistweet

Many people now use Twitter's 140 characters messaging without thinking how shortening the message

may cause the loss of significant information as when the words "is indicated" were deleted from a re-tweet about the occurrence of a second Icelandic volcano. The instantaneous exponential repetition of this tweet added to its credibility and caused an unjustified panic.

Sometimes it is inappropriate to tweet from certain locations like a war zones. During the attacks in Mumbai, Twitter was so effective in providing up to date information that there was a concern that the tweeting would reveal critical information to the terrorists.

A single tweet can be re-tweeted increasing its impact, be it negative or positive. A significant repetition increases the credibility of a claim. The original April fool's day joke about President Obama's birth location had the date removed thus significantly changing its information content and was re-circulated. Its near infinite recirculation added so much to its credibility that significant numbers of people still believe that he was not born in the USA. No hard evidence like a birth certificate has been strong enough to sway their belief in the repeated message- the wisdom of the mob. Truth is in the numbers and the media facilitates a rapid unthinking increase in the numbers.

Just as there is a gap between the speed of technological development and changes in the law to help manage the technology, we are also in catch up mode with technology.

Truth versus Offense

Unfortunately social media can also be used to intensively attack individuals from multiple directions: web pages, Twitter, Facebook and MySpace accounts. There have been law suits to get websites to remove slanderous or false posts. But in U.S. even if you win such lawsuits the Web site doesn't have to take the offending material down any more than a newspaper that has lost a libel suit has to remove the offending content from its archive. These attacks are particularly problematic because as we have seen the 'truth' of such attacks is judged by the number of times and places it is repeated. Once someone has looked at such attacks several times, automated features of the web take over building a silo and curtail access to differing views. Unfortunately if a one's image is tarnished by a flock of tweets, no matter what evidence is provided it will be difficult to fix because social media has shifted trust away from institutions and invested it in the "crowd".

Some problems arise in part because an individual's use of social media blurs the distinction between public information and private information and between work information and personal information. Notes on LinkedIn, MySpace and Facebook are a blend of private and public information.

None of the policies speak of the above identified problems with social media. The above items – changes in the understanding of truth, convergence of communications and computers, and social media systems- have combined to significantly alter our socio-technical context. This revised context requires focused attention to explicit ethical issues. It is woefully inadequate to address ethical communications problems with "be polite" and "use common sense". There is little attention to the negative non-libelous impact irresponsible use of this media may have on others or the new socio-technical conditions.

Moral Responsibility

In 2010 an Ad Hoc Committee for Responsible Computing was formed to develop a set of rules describing the Moral Responsibility for Computing Artifacts [Miller, 2010]. The Rules currently consists of five rules as a normative guide for people who design, develop, deploy, evaluate or use computing artifacts. The document focuses on "the importance of moral responsibility for these artifacts" and encourages

"individuals and institutions to carefully examine their own responsibilities with respect to computing artifacts." The document includes a preliminary definition of "moral responsibility" as indicating "that people are answerable for their behavior when they produce or use computing artifacts, and that their actions reflect on their character. [Davis 2011] "Moral responsibility" includes an obligation to adhere to reasonable standards of behavior and to respect others who could be affected by the behavior."

I think these rules capture some significant common elements of ethical action. Although they were not developed to address the specific problems identified above, I think they have identified some essential elements of moral responsibility that could be used to help address some of the issues about social media if people were aware of these rules. This can be seen by some minor modifications of these rules. I inserted the words in italics.

Rule 1*: The people who** *communicate via social media* **are morally responsible for that** *communication* **and for the foreseeable effects of** *it. This responsibility is shared with other people who** *have affected and contributed to that communication* **as part of a sociotechnical system.**

(This identifies moral responsibility for both those who create the message and those who have developed a system which misleadingly alters our understanding of the credibility of that message.)

***Rule 2*: The shared responsibility of** *a social media communication* **is not a zero-sum game. The responsibility of an individual is not reduced simply because more people become involved in designing, developing, deploying or using the artifact. Instead, a person's responsibility includes being answerable for the behaviors of the artifact and for the artifact's effects after deployment, to the degree to which these effects are reasonably foreseeable by that person.**

(This emphasizes the relevance of all participants –tweeters, followers, re-tweeters, mis-tweeters, bloggers, and subscribers for the affects of a message. The one who unthinkingly re-tweets every message is responsible for its increased credibility. The one who designs or modifies the Page Rank algorithm is responsible for the censorship and impressions it produces. The use of the word 'foreseeable' indicates that a morally responsible person should give pause and think about the consequences of each use of social media.)

***Rule 3*: People who knowingly use a particular computing artifact are morally responsible for that use.**

(The moral responsibility of a user includes an obligation to learn enough about the social media and its effect to make an informed judgment as in the Tweet about the Icelandic volcano cited above)

***Rule 4*: People who knowingly design, develop, deploy, or use a computing artifact can do so responsibly only when they make a reasonable effort to take into account the sociotechnical systems in which the artifact is embedded.**

(This requires that one try to understand the relevant system and the nature of the system and its context will impact others.)

***Rule 5*: People who design, develop, deploy, promote, or evaluate a computing artifact should not explicitly or implicitly deceive users about the artifact or its foreseeable effects, or about the sociotechnical systems in which the artifact is embedded.**

Incorporating the sense of these rules in a social media policy would help address the socio-technical

problems of social media identified above and an awareness of these rules would help address the problem of the wisdom of the mob and provide reason to adhere to social media policy which is not based on corporate self-interest. *

* This use of the moral rules was first addressed in an earlier work; Gotterbarn, D. "Tweeting is a beautiful sound, but not in my backyard: Employer Rights and the ethical issues of a tweet free environment for business." *Ethicomp 2011 Conference Proceedings*, eds. Bissett, A, et al., Sheffield Hallam University Press, Sheffield UK.

8 References

Austin, David (2011) "How Google Finds Your Needle in the Web's Haystack" American Mathematical Society, online at http://www.ams.org/samplings/feature-column/fcarc-pagerank accessed 3/15/2012

Davis, M (2011) "'Ain't no one here but us social forces:' Constructing the professional responsibility of engineers," Science and Engineering Ethics.

Gotterbarn, D (2009) "When soon after is way too late: the deception of 'opt-out' systems." Inroads SIGCSE Bulletin 41(4): 6-8.

Hafner, Katie (2007) "Corporate editing of Wikipedia revealed" on line at http://www.nytimes.com/2007/08/19/technology/19iht-wiki.1.7167084.html accessed 3/15/2012

Miller, Keith, (2010) "Ad Hoc Committee for Responsible Computing. Moral Responsibility for Computing Artifacts: Five Rules, Version 27" online at https://edocs.uis.edu/kmill2/www/TheRules/moralResponsibilityForComputerArtifactsV27.pdf accessed 3/15/2012.

Murata, K. and Orito, Y. (2011) "The right to forget/be forgotten" CEPE 2011: Crossing Boundaries, Ethics in Interdisciplinary and Intercultural Relations. Proceeding of Milwaukee Conference

Robert, Jen (2011) "Semantic analytics serves the truth & vegetables from a social media diet," online at http://www.collectiveintellect.com/blog/semantic-analytics-serves-the-truth-vegetables-from-a-social-media-diet accessed 3/15/2012.

Sanger, Larry (2010), "How I started Wikipedia part 2" online at http://www.youtube.com/watch?v=Sqb-DhgkTTI&NR=1 accessed 3/15/2012

Schwartz, Barry (2011) "Duck! Google's Cutts Responds to Search Filter Bubbles," found online http://www.seroundtable.com/google-search-bubble-response-13591.html accessed 3/15/2012

Viktor Mayer-Schönberger (2010) Delete: The Virtue of Forgetting in the Digital Age, Princeton University Press E book

Social Responsibility in the Information Society: A Potential Knowledge Gap for Tomorrow's Policy Makers

Richard Taylor

Diploma Group Subject Manager (Information Technology in a Global Society, Computer Science), International Baccalaureate, Cardiff

richard.taylor@ibo.org

Abstract

The technical advances that have enabled the development of "the cloud" have resulted in an exponential increase in the speed of information dissemination. Policy makers, sponsors of IT infrastructure and users of information and communication technologies, while being aware of the benefits of "the Cloud" as a mechanism to facilitate more efficient access to information, do not always appreciate that these developments may not always have the positive outcomes intended.

University courses such as those in Social Informatics have managed to keep pace with this rapid evolution of information and communication technologies and their societal impacts, but within the secondary education sector this has not always been possible, potentially creating a knowledge gap for tomorrow's policy makers.

Introduction

The rapid development and proliferation of information and communication technologies has irrevocably changed the way we live and work. The "Cloud"[1] as well as providing a backbone to our digital society, brings both opportunities and challenges that must be appreciated by the citizens of tomorrow. With citizens submitting ever increasing amounts of information into the Cloud, concerns are being expressed both about the ethical issues that may arise in the governance of these critical IT infrastructures (systems) and whether future citizens will be able to make informed decisions as they interact with these systems. One contributory factor in the ability of citizens to effectively interact with these IT systems may result from their study of ICT[2] during their Secondary education.

Within the UK Secondary[3] Education Sector the study of IT systems and the way that human beings interact with them has been addressed in through a myriad of ICT courses, considered to 'academic' or 'vocational', with varying degrees of success. Regardless of the type of ICT course that is followed by the students, there is a tendency to have a greater emphasis on the skills required to develop IT systems at

1 The term "the cloud" may be used as a metaphor for the Internet or a more expansive description of anything beyond the firewall http://www.infoworld.com/d/cloud-computing/what-cloud-computing-really-means-031

2 In this paper ICT is defined as a subject area with secondary education. IT is used for to refer to any IT system or infrastructure.

3 Secondary Education is defined in this case as any education or training provided received between the ages of 11 to 19.

the expense of the ability to view these IT systems in a wider context.

The ICT Debate

There has been much debate about what constitutes an appropriate ICT education in UK schools. Michael Gove, the Education Secretary, has indicated that he would like to see a reworking of the ICT National Curriculum[4] to include more computing (Michael Gove to scrap 'boring' IT lessons, 2012). Stephen Twigg, his opposition counterpart, also called for an overhaul in ICT lessons: "For too many pupils computer teaching can be little more than a glorified typing course (Burns, 2012)". The effect of such interventions could lead to a greater focus on the technical side of the subject, potentially sidelining the overarching social impacts and ethical issues and creating a potential "skills" mismatch as the vast majority of future citizens will be IT users rather IT developers. In addition to this, political and commercial considerations may restrict the teaching of themes relating to the relationship between human beings and IT systems into schools. ICT Is not currently included in the eBacc[5], and as a consequence not seen as a priority for inclusion in the post 14 curriculum. In the 16-19 (11th and 12th Grade) curriculum, it is not recognised by the Russell Group universities as a facilitating subject.

The concerns about the nature of the ICT curriculum has led to a number of interest groups such as NAACE[6], the Royal Society and Computing at School (CAS)[7] have publishing recommendations for the future direction of ICT and Computer Science which could help guide politicians and shape the development of future ICT curricula. Unfortunately, due to the lack of clarity in defining the relationship between the two subjects, where they are often seen as substitutes rather than complements, has retarded the development of a coherent strategy.

In 2011, the [UK] Government has proposed to disapply ICT from the National Curriculum Programme of Study [at Key Stage 3[8]], along with the associated Attainment Targets and statutory assessment arrangements, from September 2012. This means that, while ICT will continue as a subject within the National Curriculum (pending the outcomes of the Government's review of the National Curriculum in England), schools and teachers will have much more freedom to teach it in ways that are creative, innovative and inspirational (Department for Education, 2012). One response to this announcement was the NAACE Draft Key Stage 3 ICT discussion document has been recently published. This focuses on a broad range of topics that may provide a suitable framework for the discussion of the wider implications of IT systems. If implemented, this would allow teachers in schools to use the framework as a springboard to develop the themes in a local context further. However, the innovative nature of some of the content may be problematic by the difficulties in ensuring ICT teachers are able to keep their subject knowledge up to date, something that can be difficult with the current workload and shortage of funds for dedicated professional development.

The Royal Society paper predominantly directed at the Secondary Education sector attempts to define ICT. It considers that the subject consists of three discreet strands, shown in Figure 1; Information

4 National Curriculum - The National Curriculum is a framework used by all maintained schools to ensure that teaching and learning is balanced and consistent. (Crown Copyright, 2012)

5 eBacc - The measure recognises where pupils have secured a [GCSE] C grade or better across a core of academic subjects – English, mathematics, history or geography, the sciences and a language (Department of Education, 2011)

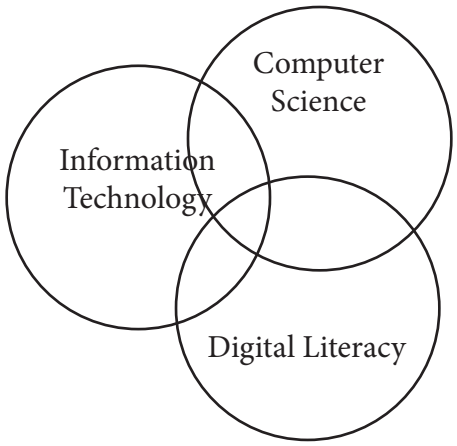
6 NAACE – National Association for Advisors for Computers in Education

7 Computers at School – Group formed to promote the teaching of Computing in UK Secondary Schools.

8 Key Stage 3 – this equates to Years 7 – 9 (6^{th} – 8^{th} Grade), the first three years of secondary education

Technology as "the assembly, deployment and configuration of digital systems to meet user needs for particular purposes", Computer Science as "the rigorous academic discipline encompassing programming languages, data structures, algorithms …" and Digital Literacy as "the general ability to use computers" (Royal Society, 2012). However, there is almost no explicit identification of the wider social, ethical and legal implications of these technologies.

Figure 1: The interrelationship of the three constituent parts of ICT as proposed by the Royal Society



Adapted from "Shut down and restart" (Royal Society January 2012)

In the 16-19 age range, ICT courses are provided by exam boards such as OCR[9] and AQA[10]. These courses are predominantly based on the design and development of IT systems at the expense of the associated social/ethical considerations which are usually addressed as single assessment statements[11]. Examples of these assessment statements include "discuss the impact of ICT on society, organisations and individuals" (OCR, 2008) and "discuss with examples the consequences of the use of ICT for different groups of individuals and society" (AQA, 2007). Courses such as these are developed over cycles that may run over a number of years and unless mechanisms are included to enable updates to the course as technology evolves, sections of these courses which are specifically linked to IT systems will become obsolete relatively rapidly.

The NAACE press release of September 2011 included reference to the emergence of new technologies where it stated "Consumer use of technology is being increasingly used for semantic and behaviour analysis; for example consumer profiling by supermarket loyalty cards …. (NAACE, 2011)" and "workforce CPD and Standards, need to be shifted in order to address the current mismatch between human web-influenced behaviour and educational practice" (NAACE, 2011). Unless ICT is able to adapt to accommodate the emerging technologies, there will always be a lag between technologies used and when they are discussed in Secondary education. An example of this can be seen with data mining which has yet to be included in any UK specification yet is a significant issue for most IT users.

A New ICT

As technologies continue to advance and societal expectations of IT systems change; ubiquitous computing

9 OCR - Oxford Cambridge and RSA Examinations, a UK awarding body,

10 AQA – Assessment and Qualifications Alliance, a UK awarding body

11 Assessment statement – a statement relating to a particular part of a specification that provides guidance to teachers about the depth at which it will be examined.

becoming the norm, the cloud becomes increasingly pervasive and the use of social networking services such as Facebook lead to a blurring of different aspects of a citizens life, the developers and managers of ICT curricula need to investigate ways to make the subject evolve to accommodate these changes. This should include an explicit framework that enables students to have a deeper understanding of the interaction between human beings and IT systems, and the social, ethical and legal considerations that arise. This essential and overarching part of any ICT course may be encompassed under the umbrella term "digital wisdom[12]". One representation of this relationship is indicated in Figure 2 below.

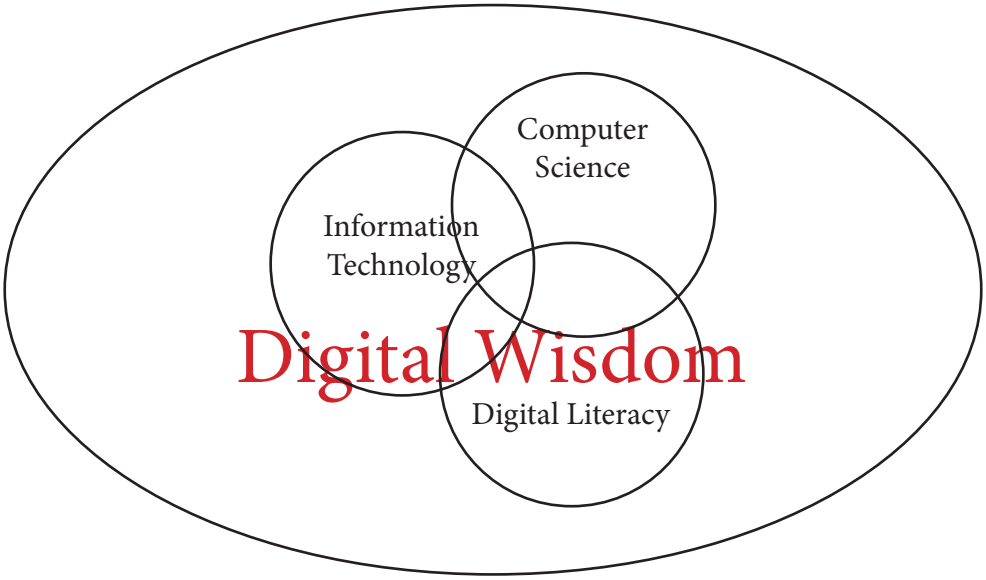


Figure 2: An alternative overview of ICT

Digital wisdom encompasses; responsible and informed use, awareness of risks, respect of relevant ethical and legal positions, the expectations and needs of different user groups.

Conclusions

The pervasive and ubiquitous nature of information and communication technologies has not only irrevocably changed the way in which we live and work, but also the nature of our interaction with them. It is possible in the university sector to deliver courses that are able to respond rapidly to these changes, but within the secondary education sector this has not been possible. The political landscape has emerged where transformative change of ICT is achievable. Any new ICT course, in order to reflect the society in which we live, would need to reflect the ever changing inter-relationship between the technology and society, within an explicitly defined ethical framework. The transformative change that ICT requires will be a major undertaking and require effective marketing the new course, making sufficient resources available for the professional development of current and new ICT teachers, engaging in discussions with awarding bodies (exam boards) and universities to develop continuity of ideas and best practice as well as raising the profile of the subject so that it is seen by key external groups such as the Russell Group as a facilitating subject. A new ICT that focuses on the relationship between human beings and IT systems has the potential to be at the centre of 21st Century education rather than at the periphery, it is an opportunity that has to be taken.

12 Digital wisdom – based on an interpretation of the term "media wisdom" that encompasses; technical competence, creativity, analysis and reflection (Martens 2011).

References

Associati, Casaleggio (2011). "The evolution of Internet of Things. online at www.casaleggio.it/Focus_internet_of_thinds-eng.pdf accessed 1 Dec. 2011.

BBC (2012) Michael Gove to scrap boring IT lessons online at http://www.guardian.co.uk/politics/2012/jan/11/michael-gove-boring-it-lessons - accessed 11 Jan 2012

Department for Education (2012) English Baccalaureate online at http://www.education.gov.uk/schools/teachingandlearning/qualifications/englishbac/a0075975/theenglishbaccalaureate - accessed 11 Jan 2012

Department for Education (2012) English Baccalaureate online at http://www.education.gov.uk/schools/teachingandlearning/qualifications/englishbac/a0075975/theenglishbaccalaureate - accessed 24 Jan 2012

Department for Education (2012) ICT Curriculum consultation online at http://www.education.gov.uk/a00202110/ict-curriculum-consultation - accessed 24 Jan 2012

DirectGov (2012) Understanding the National Curriculum online at http://www.direct.gov.uk/en/Parents/SchoolslearninganddevelopmentExamsTestsAndTheCurriculum/DG_4016665 - accessed 26 Jan 2012

Doyle J (2012) Extradition pact with the U.S. is being misused to send computer student for trial, says Sir Ming online at http://www.dailymail.co.uk/news/article-2087135/Richard-ODwyer-US-extradition-pact-misused-says-Sir-Menzies-Campbell.html?ito=feeds-newsxml – accessed 19 Jan 2012

Duquenoy P., Martens B. and Patrignani N., eds. (2010) Embedding ethics in European Information and Communication Technology Curricula. Online at Lirias.kuleuven.be/handle/123456789/269376 accessed 24 Jan. 2012.

ETICA (2011) ETICA wiki online at http://www.ccsr.cse.dmu.ac.uk/ETICA-Wiki/index.php/Main_Page - accessed 5 Jan 2012

Furber, S, (2012). Shut down or restart? Online at royalsociety.org/education/policy/computing-in-schools/report/ accessed 15 Jan 2012.

GCE Specification – ICT A/S, A2. Cambridge: OCR,(2008). Online at http://www.ocr.org.uk/qualifications/type/gce/ict_tec/ict/ accessed 30 Dec 2011.

GCE Specification - Information and Communication Technology. Manchester: AQA, (2007). Online at http://web.aqa.org.uk/qual/gce/ict/ict_materials.php?id=04&prev=04 accessed 30 Dec 2011.

Heinrich, P, Allen, A, (2012) Draft Naace Curriculum Information and Communication Technology (ICT) at Key Stage 3, Nottingham online at www.naace.co.uk/ accessed 18 Jan 2012

Johnson, L, Smith, R, Willis, H, Levine, A, Hayward, K, (2011). The 2011 Horizon Report, Austin, Texas: The New Media Consortium. online at wp.nmc.org/horizon2011/ accessed 30 Dec 2011.

Kleinig, J., ed. "Privacy and Security" . ANU E Publishing, (2011). Online at http://epress.anu.edu.au/wp-content/uploads/2011/12/whole5.pdf accessed 24 Jan 2012.

Knorr E, Gruman G (2011) What cloud computing really means online at http://www.infoworld.com/d/

cloud-computing/what-cloud-computing-really-means-031 - accessed 5 Jan 2012

Naace, (2011) “Press release Issues & Expert Recommendations from The Future of Technology in our Schools.”. online at http://www.naace.co.uk/pressrelease accessed 24 Jan 2012.

Russell Group, Informed Choices. Russell Group, (2011). Online at http://russellgroup.ac.uk/informed-choices accessed 10 Jan 2012.

The Council of European Union, Council conclusions on media literacy in the digital environment. Brussels: EU (2009). Online at ec.europa.eu/culture/media/literacy/docs/recom/c_2009_6464_en.pdf accessed 17 Jan 2012.

Deception Detection for the Tangled Web

Anna Vartapetiance
PhD Student in the Department of Computing, University of Surrey
a.vartapetiance@surrey.ac.uk

Lee Gillam
Senior Lecturer in the Department of Computing, University of Surrey
l.gillam@surrey.ac.uk

Abstract
Deception is a reasonably common part of daily life that society sometimes demonstrates a degree of acceptance of, and occasionally people are very willing to be deceived. But can a computer identify deception and distinguish it from that which is not deceptive?

We explore deception in various guises, differentiating it from lies, and highlighting the influence of medium and message in both deception and its detection. Our investigations to date have uncovered disagreements relating to the measurements of such cues, and variations in interpretations, as could be problematic in building a deception detection system.

Introduction
Suppose we wished to create an intelligent machine, and the web was the choice of information. More specifically, suppose we relied on Wikipedia as a resource from which this intelligent machine would derive its knowledge base. Any acts of Wikipedia vandalism, as being explored in an international workshop (PAN 2011) amongst other places, would impact upon the knowledge base of this intelligent system, and the system might develop confidence in entirely incorrect information. Ethically, should we develop a machine which can craft its own knowledge base without reference to the veracity of the material it considers? If we did, what kinds of "beliefs" might such a machine start to encompass and would these be acceptable? How might a learning machine distinguish between generally agreed positions, those over which there were debate, and those in which there were deceptions and lies? What kinds of conclusions might be derived about our world as a consequence of addressing such questions? If trying to construct an ethical machine, how appropriate can ethical outcomes be considered in the presence of deceptive data? And, finally, how much of the Web might be deemed deceptive?

In this paper, we investigate the nature and, importantly the detectability, of deception at large, and in relation to the web. Deception appears to be increasingly prevalent in society, whether deliberate, accidental, or simply ill-informed. Examples of deception are readily available, from individuals deceiving potential partners on dating websites, to surveys which make headlines about "Coffee causing Hallucinations" with no medical evidence and very little scientific rigour [BBC, 2009], to companies which collapsed due to allegedly deceptive financial practices (e.g. Enron, WorldCom), and segments of the financial industry allegedly misrepresenting risk in order to derive substantial profits [Telegraph, 2010]. We envisage a Web Filter which could be used equally well as an assistive service for human readers, and as a mechanism

within a system that learns from the web, and are concerned with whether relevant literature presents a suitably computable basis for such a system. In section 2, we define deception as something intentional and more broad than lying; section 3 looks at various key features of deception and gears towards the detection of deception. Section 4 provides a brief view of the relationship between deception and ethics, and offers a few examples characterised with respect to both. In section 5, we attempt to adopt some existing approaches to deception detection, with varying successes. Finally, we highlight related questions about deception detection and the possibilities for such a system.

Defining Deception

Like it or not, deception is a reasonably common part of daily life. Society sometimes demonstrates a level of acceptance of it, and occasionally even seeks to be deceived. Most sociologists believe that people engage in deceptive behaviour in order to avoid tension and conflicts, manage impressions, and minimize negative feelings [Goffman, 1959; Lippard, 1988, Metts, 1989]. DePaulo et al. [1996] argues that deception in everyday life is less about pursuit of goals such as financial gain and material advantage and more about pursuit of rewards such as esteem, affection and respect. As a result, she believes that lies we hear everyday are more often regarding feeling, preferences and opinions. But what do we understand by deception? Mahon [2007] defines deceiving as:

> *"To intentionally cause another person to have or continue to have a false belief that is truly believed to be false by the person intentionally causing the false belief by bringing about evidence on the basis of which the other person has or continues to have that false belief."*

Here, deception is an *intentional* act, which distinguishes it from simply being ill-informed: a person deceived may well have false beliefs but lack awareness of the falsehood. If they broadcast that falsehood, they cannot really be perceived as deceptive people. So, believing the Earth to be flat and the Sun to rotate it, may well have been the case at a particular historical juncture, but there was not necessarily an intention to deceive.

So, when we use the term *deception* do we simply mean *lie*? Many researchers use both terms interchangeably, such as Mahon [2007], Vrij [2000] and Ekman [1988]. However, if we consider a definition such as:

> *"A lie is a statement, believed by the liar to be false, made to another person with the intention that the person be deceived by the statement [Bok, 1978]."*

we must characterise lies as a specialised form of deception which may exist alone, or which may be accompanied by other actions and behaviours that might be in agreement with the lies by being deceptive themselves, or, likely, in disagreement. Essentially, a lie is "told", whilst deceptions can occur without requiring such verbalisation – as we will explore later in this paper.

A basic classification of lies has been proposed by Erat & Gneezy [2009], defining 4 types based on their consequences (Figure 1).

1. "Selfish black lies" increase the player's payoff but decrease the others payoff
2. "Spite black lies" decrease both sides' payoffs.
3. "Pareto white lies" increase both sides' payoffs.
4. "Altruistic white lies" increase the others payoff but are costly to the deceiver

Whilst we have suggested lying as a specialisation of deception, we could also generalise (Figure 1) to deception, with similar payoffs.

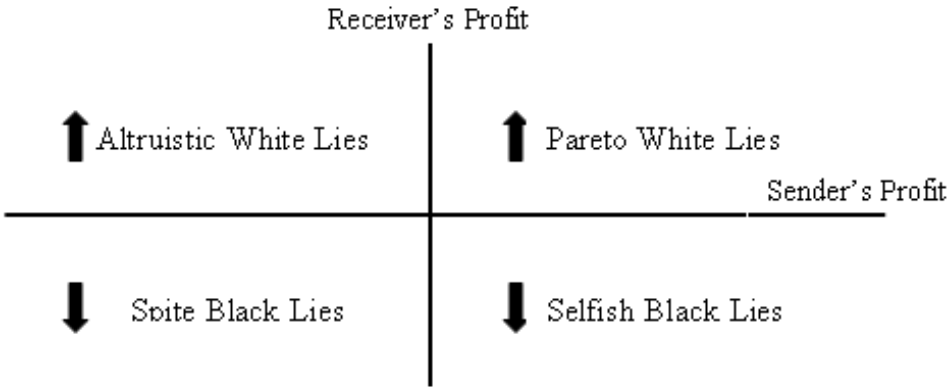


Figure 1: Taxonomy of Lies Based on Change in Payoffs

The above figure may also be related to the degree to which deceptive behaviour is planned. The most harmful deceits– those that can in lower half of the figure – are those where prior thinking is more likely to be required, as well as some level of preparation. By contrast, actions with little or no negative consequences for receivers – white lies – are more likely spontaneous. Researchers are not necessarily in agreement over such a characterisation (see, for example, Camden, Motley, & Wilson, 1984; Lindskold & Walters, 1983; Hample, 1980].

Dimensions of Deception

The nature of deception has encouraged study by psychologists, sociologists, criminologists, philosophers and anthropologists over many for centuries. "A History of Lie Detection", Trovillo [1939], covers different methods of detecting deception from African witchcraft to invention of polygraphs which dates the importance of recognising cues of deception as far as 900 B.C.

In this section, we provide a brief overview of research into deception and its detection, covering various studies into aspects of communication and occasionally identifying variations in interpretations which may be awkward to reconcile.

Medium

Different mediums are used for different sorts of communication, with variations in the nature of interaction, capabilities for feedback, number of available channels, language variety, and so on. "media richness" can be used to refer to a medium which can offer scope for larger numbers of potentially deceptive cues. The medium can be characterised by [Hancock, Thom-Santelli & Ritchie, 2004]:

- **Synchronicity:** extent of real-time communication.
- **Distribution:** degree of physical co-location.
- **Recordability:** extent to which the medium is automatically recordable.

Based on these elements, a rich medium is more "synchronous, concentrated and non-recordable", and so face-to- face interactions offer a rich medium, while numeric documents offer less richness [Daft &

Lengel, 1986; Daft & Wiginton, 1979].

Using such characteristics, researchers have tried to highlight preferable mediums for deception in Media Richness Theory [MRT, Daft & Lengel, 1984], Social Presence Theory [SPT, Short et al. 1976], Social Distance Theory [SDT, DePaulo et. al. 1996], Feature Based Model [FBM, Hancock, Thom-Santelli & Ritchie, 2004], and Channel Expansion Theory [CET, Carlson and Zmud, 1994 and 1999]. According to MRT and SPT, richer mediums should be preferred for deceptive behaviour, so face-to-face communication has more potential to contain deception than emails. However, according to SDT deceivers will prefer emails to face-to-face communication as deceivers will feel uncomfortable during the deception. In FBM, most deception emerges in synchronous, concentrated and non-recordable media, so telephones are preferred to instant messaging and face-to-face interaction, while emails are less preferable. CET highlights familiarity with the medium, topic of communication, environment, and many more elements, as changing the richness of the medium which can affect the choice for deceptive behaviour. A summary of medium, features of the medium, and predictions of being lie-bearing (excluding CET which does not follow any specific rule) can be seen in Table 1.

Table 1: Deception Media Preferences and Theories

	Face to Face	Phone	IM	Email
Features of the medium				
Synchronous	X	X	X	
Recordless	X	X		
Distributed		X	X	X
Prediction of lies				
Feature Based	2	1	2	3
Media Richness	1	2	3	4
Social Distance	4	3	2	1
Pertinent features of communication media for deception and predictions for lying (1 = highest, 4 = lowest) from Hancock, Thom-Santelli and Ritchie [2004]				

Given varying interpretations, it is difficult to determine an overall expectation just from the medium. And, of course, researchers may have come to various conclusions depending on the nature of analysis undertaken. There is certainly scope for more comprehensive research in this area.

Deception in the 3 Vs

Telling a lie or being engaged in a deceptive behaviour is reported to be mentally, emotionally and physically more challenging than being truthful [Miller & Stiff, 1993; Zuckerman et al., 1981; Vrij, Edward, & Bull, 2001]. Cues for deception can be carried in communication, depending on the medium (see above) by any or all of the 3 Vs:

- **Visual** (non-verbal): physical behaviour; reactions, movements.
- **Verbal**: anything said or written.
- **Vocal**: elements that accompany verbal communication.

Visual

"They [the movements of expression] reveal the thoughts and intentions of others more truly than do words, which may be falsified [Darwin, 1872, p 359]".

Ekman and Friesen [1969] define non-verbal behaviour as "Any movement or position of the face and/or the body" defining 3 body areas that participate in the information transfer regarding deception [Ekman, 1965]:

1. **Body act:** movements of body parts, particularly intensity and emotion. There is a cognitive challenge (or cognitive load ,Vrij et al. 2011, or complexity, Newman et al. 2003], in controlling body language whilst deceiving, occasionally exemplified in people sitting on their hands to prevent hand gestures evidencing deception.

2. **Posture:** taking up specific body positions. Researchers have developed lists of postures they believe to be indicative of deception [Mehrabian, 1971; Horvath et al.1994], suggesting that truth-tellers are more likely to lean forward, in a comfortable and open display. Deceivers will usually take defensive and frozen postures, such as leaning back and crossing their arms or legs.

3. **Face:** movements of parts of the face or posing certain expressions. Porter & ten Brinke [2008] have demonstrated that faking negative emotions is harder than faking positive emotions because eyebrows, mouth, and eye movements (eyelashes, blinks, eyeballs, etc) can leak cues, but it might be possible to control these (see 1), whilst pupillary size changes may be considered reliable as they are harder to control, although harder to see [Lubow and Fein, 1996].

Most non-verbal cues appear to exist across cultures, such as changes blinking, facial micro expressions and body movements. However, most non-verbal cues will be absent when the communicators are not physically or virtually co-located (distributed), but can be present in both synchronous and asynchronous communication, depending on the medium being used.

Vocal Deception

There are two main features of vocal deception: nature of voice, and the way the words are being said. Nature of voice refers to tone, vocal tension, and pitch because tension causes vocal cords to tighten resulting in such changes. When considering how words are said, focus is on characteristics such as speed, number of errors, pauses and length of them. But these changes may also be dependent on the context, so cues to vocal deception in interpersonal interactions will differ from cues to vocal deception formal interviews [Mann, Vrij & Bull, 2002], and may also vary according to the participants.

Verbal Deception

Verbal deception covers significant ground, and is our principal area of interest amongst the 3 Vs. In contrast to the language-independence of non-verbal cues, verbal cues have language-dependence with variations in lexical and grammatical combinations potentially offering up cues. When someone is engaged in verbal deception, language patterns will change. These may be accompanied by gestures (non-verbal elements) and vocal elements depending on the medium.

Three main types of interaction are important:

1. **Spoken** (in face-to-face, in audio & video records, in video & audio conferences, and so on)

2. **Written** (in blogs, emails, testimonies, academic articles, and so on)

3. **Transcripts** of spoken

Spoken communication combines verbal cues with vocal (and non-verbal cues if the participants are face-to-face), either in parallel or as complementary [Ekman & Friesen, 1969]. However, in a written context vocal and non-verbal cues are missing [Gupta & Skillicorn, 2006]. To identify deception requires either comparisons amongst lexical and grammatical forms used by the same person in deceptive and non-deceptive communication, or analysing such forms across samples from different people.

A number of researchers [Newman et al. 2003; Keila & Skillicorn, 2005b; Burgoon et al. 2008] have been investigating the lexical, syntactic, and meta-content features of verbal deception, classifying pattern changes by three main dimensions: (a) quantity (b) quality and (c) overall impression. Quantity changes relate to number of words being used; e.g. in a sentence, in a passage, and so on. Qualitative changes relate to lexical selection but still depend in part on the number of nouns, verbs, and sentences. Overall impression appears to relate to human judgement [DePaulo et al. 2003], but due to the subjective nature of such assessments we have decided to set aside such cues from our research and focus on those which are more readily measurable.

Regarding quantity and quality, there are different perspectives with many overlaps that highlight the most important or frequently studied cues. Pennebaker's approach has been adopted widely, and is based on style (word-by-word), accuracy, and flexibility, for both written and spoken text (such as: Newman et al. 2003; Toma & Hancock, 2010]. Pennebaker characterizes deceptive behaviour on the frequency of four kinds of words in three main categories:

1. **Self-references:** Use of first-person singular shows speaker-ownership of statements and presents a link between the reality and the speaker. Self awareness reportedly leads to honesty, which will increase the number of self references [Voraurer & Ross, 1999; Davis & Brock, 1975]. Deceivers may use fewer self-references to "distance" and "dissociate" themselves from the ownership of a statement [Knapp, Hart & Dennis, 1974], or because it is not a personal experience.

2. **Negative words:** Emotions such as guilt, shame and fear can result in leakage of deception [Ekman, 1985/1992; Vrij, 2000]. As most of these feelings are negative, this can lead to discomfort for the deceiver [DePaulo et al., 2003]. The effect of these negative emotions is reported to increase the number of negative words in deceivers' statement.

3. **Cognitive complexity:** increased cognitive complexity/load occurs through (a) exclusive words and (b) motion/action verbs. Deceivers should use fewer "exclusive words" - such as except, but, without and exclude [Pennebaker & King, 1999]. Fewer exclusive words leads to an increase in motion and actions verbs (e.g. go, lead, walk).

DePaulo et al. [2003] offer a list of 158 cues. From this list, we consider just 25 cues to relate to verbal and to be measurable, and these relate to just 10 research papers over that period. The cues include: Response length, Talking time, Cognitive complexity, Unique words, Generalising terms, Self-references, Mutual and group and other references, Word and phrase repetitions, Negative statements and complaints, and Extreme descriptions.

In contrast, Burgoon's group have 45 cues in 8 categories, with 20 shared with DePaulo, but some of which are variously elaborated, expanded, and contracted, elsewhere [Burgoon & Qin, 2006; Qin et al. 2005; Zhou et al. 2004; Zhou, Burgoon & Twitchell, 2003; Zhou et al. 2003; Burgoon et al. 2003]. Burgoon's categories are:

1. **Quantity** (syllables, word, sentence, verbs, simple sentences, noun phrase)
2. **Complexity** (big words, syllables per word, short sentences, long sentences, avg. clauses, avg. length of noun phrase, flesh-kincaid grade level, syntactic complexity, sentence complexity, conjunctions, lexical complexity, pausality)
3. **Diversity** (lexical diversity, content word diversity, redundancy)
4. **Specificity** (sensory details, modifiers, first-person singular/ plural pronouns, second/ third person pronouns, temporal and spatial details, Over all specificity, perceptual information)
5. **Affect** (affect, pleasantness, imagery, positive and negative affects)
6. **Activation** (emotiveness index , activation,)
7. **Verbal non-immediacy** (passive voice, reference, modal verbs, uncertainty, objectification, generalizing term)
8. **Informality** (typo errors)

It is unclear whether Burgoon's set of cues are comprehensive and fixed, and related publications offer up those that are relevant to the specific research being presented, or whether they can be added to, deleted from, or move to other categories as appears to have happened within his publications.

Relating Ethics and Deception

Knowing certain important characteristics of deception, is it ethical to deceive? Although we do not aim at comprehensive treatment of the ethics of deception, perspectives offered by ethical theories are still interesting.

Deontology, and here we may specialise to Kantianism, emphasizes the goodness of rules. An absolute duty not to be deceptive would not consider whether the deception is well-intentioned. If somebody asks where your friend is because they want to harm him, and you point in the opposite direction, you have *deceived* but you have not *lied* because you have not verbalised the action[1]. From this perspective, Erat and Gneezy's [2009] classification would be entirely unacceptable. Consequentialists and Utilitarians, on the other hand, would tend towards agreement with Erat and Gneezy's classification as utility (happiness and benefits) is considered - with Pareto White Lies appearing to offer the best option. So, deception may offer a better alternative except in relation to "Spite Black Lies". Prima Facie Duty [Ross, 1930; Garrett, 2004] includes Fidelity, which requires avoiding deception unless other duties are more pressing – in which case, deception may be acceptable. This consideration of Erat and Gneezy's [2009] classification suggests that deceptions of various kinds can be readily acceptable – but not by Kantianism.

Characterising Deceptive Acts

How might we characterise examples of deception with respect both to Erat & Gneezy [2009], and ethical theories? We offer a few examples here.

Placebos: Those given placebos them are not likely to be told, so the deception is not verbalised: Sherman

1 Kant can probably be extended to allow for such a broader interpretation than examples with which he is typically identified.

and Hichner's [2008] study shows that although 45% of clinical practice involves some kind of placebo, only 4% of the patients are informed. Doctors are participating in an act of deception that benefits both themselves and their patients, increasing the total happiness. This suggests a "Pareto White Lies", acceptable by consequentialists and utilitarians.

Online dating: In exploring an online dating website, it was apparent that certain profiles evidenced various kinds of deceptions geared towards convincing others to become interested in them without knowing the reality from the beginning - "Selfish Black Lies". Depending on severity and over-riding duties, some might agree with such behaviour. Figure 2, below, shows an example of potential deception in an online dating website. A user appears to have created two different profiles – the image has been mirrored, and the wall adornments appear to offer good confirmation of this, but variations in nationality and height (either a 5cm growth or shrinkage has been experienced within a 1 year period by this person – age is the same for each – or there is some preference for height being played to that depends on national preferences). Unfortunately, examples of such an attempt to deceive are frequent but only a comprehensive analysis of the database could pick out all cases of it.

Figure 2: Example of verbal and image deception in online dating websites

Surveys which make headlines about "Coffee causing Hallucinations" with no medical evidence and very little scientific rigour might also be characterised as "Selfish Black Lies". Plagiarism might also be an example of "Selfish Black Lies", where the sender intends to deceive another into believing knowledge or work he/she does not own. But, then, what if researchers offer references to a work that does not exist, believing that it does? [Dubin, 2004]

Financial practices: Companies which collapsed due to deceptive financial practices (e.g. Enron, WorldCom), and difficulties relating to the so-called "credit crunch", may have initially been conceived as "Selfish Black Lies". However, those responsible may not have emerged unscathed – though from the credit crunch, some appear to have escaped rather better than others – from what now appears to be "Spite Black Lies". Utility is not increased, and so the only remaining mitigation would relate to duties more pressing than Fidelity.

Examples of "Altruistic white lies" can be seen in familiars and friends who care for each other where someone deceives the other to believe in something because they would like to protect time emotionally or physically; just like pretending to enjoy a meal.

Deception detection – a system?

Given the extent of research on verbal deception, are we able to use this to good effect? We have tried to ascertain the potential for extant research on verbal deception to lead to a Web Filter. Such a filter should flag deceptive material, whilst allowing non-deceptive material to pass through – low numbers of false positives should be generated.

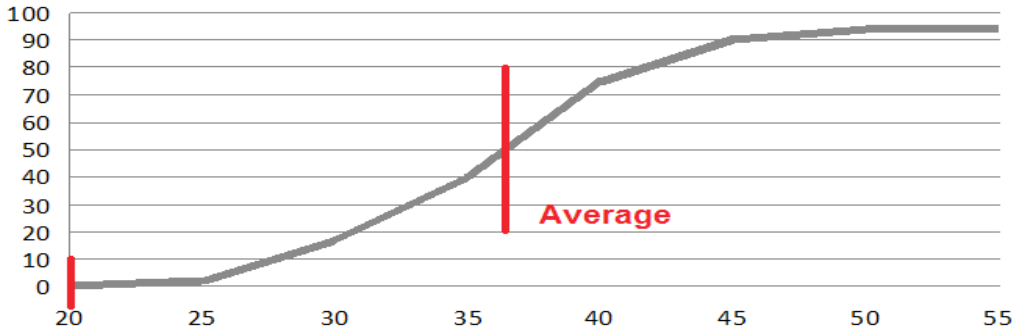


Figure 3: Cumulative distribution of proportion of "big words" according to LIWC on 100 MuchMore scientific abstracts showing a distribution rather higher than that indicated.

In Burgoon's research, neither the cues nor the threshold values for the cues are stable. For example, in Zhou et al. [2004] number of words, sentences and emotiveness index are shown to increase in cases of deception, but in Burgoon et al. [2003] and Zhou et al. [2003] all three are shown to decrease. Similar inconsistencies can be seen across their papers for cues relating to first person singular, perceptual information, positive affects, redundancy, modal verbs and uncertainty. To implement a system capable of weighting its own parameters would prove to be a challenge, and further work will be needed to unpick these inconsistencies.

Pennebaker's group offer a free trial version of the "Linguistic Inquiry and Word Count (LWIC)" software online. It is claimed that LIWC can flag deceptive text on any sort [Tausczik & Pennebaker, 2009]. We have tested the online version using 100 articles (scientific abstracts) from the MuchMore Springer English Corpus (plain version)[2], which we have no reason to believe are deceptive. Figure 3 shows the frequencies obtained only for "big words", which – if we interpret LIWC correctly – may be indicating that all these texts are deceptive. A threshold for "formal" of 19.6 is suggested by LIWC, presumably indicating an acceptable percentage of big words. These abstracts have an average of 36.15 (55.26 max, 21.15 min – all texts above 19.6). With few self-references, few positive or negative emotions, and fewer articles (e.g. determiners) than expected, additional efforts will be required in interpreting these results appropriately to avoid all such text being flagged as a false positive.

Keila & Skillicorn [2005b] have apparently adopted concepts from LIWC, yet they tried to use SVD to – according to their abstract - determine a distinction in language use by those involved with criminal activities at Enron (using the Enron email corpus). However, the paper does not appear to elaborate this claim further. We have used the same approach (in R), and agree with their conclusions that they can differentiate between long and short emails, but it is unclear what relationship this bears to criminality, and hence it would not appear appropriate for system development.

To date, despite all of the well-grounded research undertaken by those for whom we must have great

2 MuchMore Springer Bilingual Corpus, available at: http://muchmore.dfki.de/resources1.htm

respect, it would seem that the development of a Web Filter for deception is going to be a significant challenge indeed.

Conclusions

Deception is a reasonably common part of daily life and the Web offers many new opportunities to make it more so. Characteristics of the Web, and related technologies, such as asynchronicity and distribution can offer the right kind of environment for this to propagate without detection. With this in mind, we reviewed deception research, looking in part at ways it might be detected, and relating it to the ethics of deception. We considered whether extant approaches might offer the potential for the development of a Web Filter for deception, but have yet to discover a fixed set of deception cues with well-understood thresholds which might be applied to this. It is possible that many of the results claimed in literature depend entirely on the researcher and their cues, and also on the context of the research; these might not extrapolate across text types, genres, and so on. Whilst searching for text corpora upon which to evaluate extant deception approaches in the week prior to submission of this full paper, we discovered Potts' [2010] course material on deception. We are encouraged by the fact that our work is proceeding along related lines, with some of the same core references, although there are clear points of deviation which will doubtless be clarified in time.

Elsewhere, researchers are building intelligent systems that use the Web as the main source of information[3]. However, if humans don't have a good ability to detect deception, can't agree on the cues of deception, and have differences of opinion on how deception might be detected, how could we possibly craft such systems and rely on their outputs? Alternatively, what happens if we deliberately create intelligent machines to deceive [Arkin, 2010], could we detect them? How would we prevent them? Should we be able to?

Earlier versions of this paper were presented at the 2012 IFIP Social Accountability and Computing Workshop "ICT critical infrastructures and social accountability: methods, tools and techniques" at Middlesex University in London, and in Proceedings of the 12th ETHICOMP International Conference on the "Social and Ethical Impacts of Information and Communication Technology" at Sheffield Hallam University.

References

Arkin, R. (2010), The Ethics of Robotics Deception, 1st International Conference of International Association for Computing and Philosophy, Aarhus, DK, pp. 1-3.

BBC (2009), 'Visions link' to coffee intake. BBC News, online at: http://news.bbc.co.uk/1/hi/health/7827761.stm. Accessed 10.06.2011

Bok, S. (1978), Lying: Moral choice in public and private life, New York: Pantheon.

Burgoon, J.K., Blair, J.P., Qin, T. & Nunamaker, J.F., Jr. (2003), Detecting deception through linguistic analysis, Proceedings of First NSF/NIJ Symposium on Intelligence and Security Informatics (ISI), June 2-3, 2003, Tucson, AZ, pp. 91-101.

Burgoon, J.K., Blair, J.P. & Strom, R.E. (2008), Cognitive Biases and Nonverbal Cue Availability in Detecting Deception, Human Communication Research, 34(4), pp. 572-599.

Burgoon, J.K. & Qin, T. (2006), The dynamic nature of deceptive verbal communication, Journal of

3 Examples can be seen at: http://wikipediavandalism.tumblr.com/

Language and Social Psychology, 25(1), pp. 76-96.

Camden, C., Motley, M.M. & Wilson, A. (1984), White lies in interpersonal communication: A taxonomy and preliminary investigation of social motivations, Western Journal of Speech Communication, 48, pp. 309-325.

Carlson, J.R. & Zmud, R.W. (1994), Channel Expansion Theory: A Dynamic View of Media and Information Richness Perceptions, Proceedings of Academy of Management Best Papers, pp. 280–284.

Carlson, J.R & Zmud, R.W. (1999), Channel expansion theory and the experiential nature of media richness perceptions, Academy of Management Journal, 42(2), pp. 153-170.

Daft, R.L. & Lengel, R.H. (1984), Information richness: a new approach to managerial behavior and organizational design. In Cummings, L.L. & Staw, B.M. (eds.), Research in organizational behaviour, 6, pp. 191-233, Homewood, IL: JAI Press.

Daft, R.L. & Lengel, R.H. (1986), Organizational information requirements, media richness and structural design, Management Science, 32(5), pp. 554-571.

Daft, R.L. & Wiginton, J.C. (1979), Language and Organization, Academy of Management Review 4 (April), pp. 179-191.

Darwin, C. (1872), The Expression of the emotions in man and animals, London: John Murray.

Davis, D. & Brock, T.C. (1975), Use of first person pronouns as a function of increased objective self-awareness and performance feedback, Journal of Experimental Social Psychology, 11, pp. 381-388.

DePaulo, B.M., Kashy, D.A., Kirkendol, S.E., Wyer, M.M. & Epstein, J.A. (1996), Lying in everyday life, Journal of Personality and Social Psychology, 70, pp. 979–995.

DePaulo, B.M., Lindsay, J.J., Malone, B.E., Muhlenbruck, L., Charlton, K. & Cooper, H. (2003), Cues to deception, Psychological Bulletin, 129(1), pp. 74-118.

Dubin, D. (2004), The most influential paper Gerard Salton never wrote, Library Trends 52/4, pp. 748-764.

Ekman, P. (1965), Communication through nonverbal behavior: A source of information about an interpersonal relationship, In Tompkins S.S. & Izard C.E. (eds.), Affect, Cognition, and Personality, New York: Springer.

Ekman, P. (1985), Telling lies, Clues to deceit in the marketplace, politics, and marriage, New York: W. W. Norton & Company.

Ekman, P. (1988), Lying and nonverbal behavior: Theoretical issues and new findings, Journal of Nonverbal Behavior, 12, pp. 163-175.

Ekman, P. (1992), Telling lies. Clues to deceit in the marketplace, politics, and marriage (2nd. ed.), New York: W. W. Norton & Company.

Ekman, P. & Friesen, W.V. (1969), The repertoire of nonverbal behavior: Categories, origins, usage and coding, Semiotica, 1, pp. 49-98.

Erat, S. & Gneezy, U. (2009), "White Lies", Rady Working paper, Rady School of Management, UC San Diego.

Garrett, J. (2004), A Simple and Usable (Although Incomplete) Ethical Theory Based on the Ethics of W. D. Ross. Online at http://www.wku.edu/~jan.garrett/ethics/rossethc.htm. Accessed 10.06.2011.

Goffman, E. (1959), The presentation of self in everyday life, Garden City, New York: Doubleday Anchor.

Gupta, S. and Skillicorn, D. (2006), Improving a Textual Deception Detection Model, Proceedings of the 2006 conference of the Center for Advanced Studies on Collaborative research, October 16-19, 2006, Toronto, Canada.

Hample, D. (1980), Purposes and effects of lying, Southern Speech Communication Journal, 46, pp. 33-47.

Hancock, J.T., Thom-Santelli, J. & Ritchie, T. (2004), Deception and design: The impact of communication technology on lying behaviour, Proceedings of the Conference on Human Factors in Computing Systems (ACM SIGCHI), pp. 129-134.

Horvath, F., Jayne, B. & Buckley, J. (1994), Differentiation of truthful and deceptive criminal suspects in behavior analysis interviews, Journal of Forensic Sciences, 39, pp. 793–807.

Keila, P.S. & Skillicorn, D.B. (2005b), Structure in the Enron email dataset, Computational & Mathematical Organization Theory, 11(3), pp. 183-199.

Keila, P.S. & Skillicorn, D.B. (2005), Detecting Unusual Email Communication. Proceedings of CASCON 2005 (IBM Centers for Advanced Studies), Toronto, Canada, pp. 117-125.

Knapp, M.L., Hart, R.P. & Dennis, H.S. (1974), An exploration of deception as a communication construct, Human Communication Research, 1, pp. 15-29.

Lindskold, S. & Walters, P.S. (1983), Categories for acceptability of lies, The Journal of Social Psychology, 120, pp.129-136.

Lippard, P.V. (1988), Ask me no questions, I'll tell you no lies: Situational exigencies for interpersonal deception, Western Journal of Speech Communication, 52, pp. 91–103.

Lubow, R.E. & Fein, O. (1996), Pupillary size in response to a visual guilty knowledge test: new technique for the detection of deception, Journal of Experimental Psychology: Applied, 2(2), pp. 164–177.

Mahon, J.E. (2007), A Definition of Deceiving, International Journal of Applied Philosophy, 21, pp. 181-194.

Mann, S., Vrij, A. & Bull, R. (2002), Suspects, lies, and videotape: an analysis of authentic high-stake liars, Law and human behavior, 26(3), pp. 365-76.

Mehrabian, A. (1971), Nonverbal betrayal of feeling, Journal of Experimental Research in Personality, 5, pp. 64-73.

Metts, S. (1989), An exploratory investigation of deception in close relationships, Journal of Social and Personal Relationships, 6, pp. 159–179.

Miller, G.R. & Stiff, J.B. (1992), Applied issues in studying deceptive communication, In R. S. Feldman (ed.), Applications of nonverbal behavioral theories and research, pp. 217-237.

Newman, M.L., Pennebaker, J.W., Berry, D.S., & Richards, J.M. (2003), Lying words: Predicting deception from linguistic styles, Personality and Social Psychology Bulletin, 29(5), pp. 665-675, Sage Publications.

Pan, (2011), Pan 2011 Lab Uncovering Plagiarism, Authorship and Social Software Misuse, 19- 22 September 2011, Amsterdam.

Pennebaker, J.W. & King, L.A. (1999), Linguistic styles: Language use as an individual difference, Journal of Personality and Social Psychology, 77, pp. 1296-1312.

Porter, S. & ten Brinke, L. (2008), Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions, Psychological Science, 19(5), pp. 508-514.

Potts, C. (2010), Deceptive Language, Handout for Extracting Social Meaning and Sentiment, Department of Linguistics University of Stanford, online at: http://www.stanford.edu/class/cs424p/materials/ling287-handout-11-02-deception.pdf. Accessed 16.06.2011.

Qin, T., Burgoon, J.K., Blair, J.P. and Nunamaker, J.F. (2005), Modality Effects in Deception Detection and Applications in Automatic Deception Detection, The 38th Annual Hawaii International Conference on System Sciences (HICSS), January 03, 2005, pp. 1-10.

Ross, W.D. (1930), The Right and the Good. Oxford: Clarendon Press.

Sherman, R. Hickner J. (2008), Placebos: current clinical realities, The Journal of Clinical Ethics, 19, pp. 62-65.

Short, J. A., Williams, E. & Christie, B. (1976), The social psychology of telecommunications, New York: John Wiley & Sons.

Tausczik, Y.R., & Pennebaker, J.W. (2009), The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods, Journal of Language and Social Psychology, 29(1), pp. 24-54.

Telegraph (2010), RBS 'deceived' Highland Capital over collateralised debt obligation. The Telegraph, online at: http://www.telegraph.co.uk/finance/newsbysector/banksandfinance/8190075/RBS-deceived-Highland-Capital-over-collateralised-debt-obligation.html. Accessed 16.06.2011.

Toma, C.L. & Hancock, J.T. (2010), Reading between the Lines: Linguistic Cues to Deception in Online Dating Profiles, Proceedings of the ACM conference on Computer-Supported Cooperative Work (CSCW 2010),pp. 5-8.

Trovillo, P.V. (1939), A history of lie detection, Journal of Criminal Law and Criminology, 29, pp. 848-881.

Vorauer, J.D. & Ross,M. (1999), Self-awareness and feeling transparent: Failing to suppress one's self. Journal of Experimental Social Psychology, 35, pp. 415-440.

Vrij, A. (2000), Detecting lies and deceit: The psychology of lying and its implications for professional practice, Chichester: John Wiley and Sons.

Vrij, A., Edward, K. & Bull, R. (2001), Stereotypical verbal and nonverbal responses while deceiving others, Personality and Social Psychology Bulletin, 27, pp. 899-909.

Vrij, A., Granhag, P.A., Mann, S. & Leal, S. (2011), Outsmarting the liars: Towards a cognitive lie detection approach, Current Directions in Psychological Science, 20, pp. 28-32.

Zhou, L., Twitchell, D. P., Tiantian, Q., Burgoon, J. K. & Nunamaker, J. F., Jr. (2003), An exploratory study into deception detection in text-based computer-mediated communication. Proceedings of the 36th Annual Hawaii International Conference on System Sciences, Waikoloa, HI, pp. 10

Zhou, L., Burgoon, J.K. & Twitchell, D.P. (2003) A Longitudinal Analysis of Language Behavior of Deception in Email in: Intelligence and Security Informatics, Proceedings of First NSF/NIJ Symposium, ISI 2003, Tucson, AZ, USA, June 2-3, 2003, H. Chen, R. Moore, D. Zeng and J. Leavitt (eds.), Springer Berlin / Heidelberg, 2003, pp. 102-110.

Zhou, L., Burgoon, J. K., Zhang, D. & Nunamaker, J. F., Jr. (2004), Language dominance in interpersonal deception in computer-mediated communication, Computers in Human Behavior, 20(3), 381-402.

Zuckerman, M, DePaulo, B.M. & Rosenthal, R.(1981), Verbal and nonverbal communication of deception, In L. Berkowitz (Ed.), Advances in experimental social psychology, 14, pp. 1-59.