"I Don't Know Where He is Not": Does Deception Research yet offer a basis for Deception Detectives?

Anna Vartapetiance
Department of Computing
Faculty of Engineering & Physical Sciences
University of Surrey
a.vartapetiance@surrey.ac.uk

Lee Gillam
Department of Computing
Faculty of Engineering & Physical Sciences
University of Surrey
l.gillam@surrey.ac.uk

Abstract

Suppose we wanted to create an intelligent machine that somehow drew its intelligence from large collections of text, possibly involving the processing of collections available on the Web such as Wikipedia. Does past research in deception offer a sufficiently robust basis upon which we might develop a means to filter out texts that are deceptive, either partially or entirely? Could we identify, for example, any deliberately deceptive edits to Wikipedia without consulting the edit history? In this paper, we offer a critical review of deception research. We suggest that there are a range of inconsistencies, contradictions, and other difficulties in recent deception research, and identify how we might begin to address deception research in a more systematic manner.

1 Introduction

Deception exists in various forms, and there can be acceptance in society of deceits of various kinds - typically geared towards personal gain (self-deception) or protection from harm. Often termed "white lies", these differ significantly from those likely to be of a more harmful nature ("black lies") – and here we would include the misrepresentation *of* science and the misrepresentation *as* science; the latter is prevalent in, for example, the advertising of cosmetic products. It remains difficult to discern, however, whether the portrayal by an apparently trusted media outlet of a reported survey of 200 students' responses to a question of whether they thought they had hallucinated after drinking coffee fits the former or the latter when characterized by the BBC as "'Visions link' to coffee intake". Could we rely on existing deception research to enable us to distinguish amongst the presentation of such things on the Web?

In this paper, we present a critical review of deception research, seeking to answer the questions outlined above. We first explain our preferred definition of deception, to disentangle deceptions from lies, and then clarify the impact that selection of a specific medium (text) has on the likely nature of deception. We then review the features that researchers tend to focus on as "cues" that might be used for detecting deception, focus these down to a set as may be detectable in text, and then demonstrate that in treatments of such cues by leading deception researchers there are various inconsistencies, contradictions, and other difficulties. We further consider how we might begin to address these difficulties, such that a more systematic approach might emerge from this research, and what future work might emerge.

2 Defining Deception

To understand deception, it is important to establish what we mean by it. Out of various definitions for deception, (e.g. Masip, Garrido & Herrero, 2004; Hall & Pritchard, 1996; Russow, 1986), we settle on Mahon (2007):

> *"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*

Proceedings of the EACL 2012 Workshop on Computational Approaches to Deception Detection, pages 5–14,
Avignon, France, April 23 - 27 2012.

the other person has or continues to have that false belief."

This particular definition leads with intent, which offers contrast with unintended actions as might lead to deception, and also allows us to distinguish from the ill-informed (e.g. believing the Earth is flat, the centre of the Universe, and so on). This also covers a deception occurring through a variety of actions or inactions. Some researchers equate lies with deceptions and have a tendency to use both terms interchangeably (Ekman; 1985; Vrij, 2000); we consider lies to be a specialized subgroup of deception and again highlight Mahon (2008):

"… to make a believed-false statement with the intention that that statement be believed to be true".

Hence, lies have an essentially narrow scope to specific false statements. For example, deliberately pointing the wrong way without saying which way to go would be a deception, but only becomes a lie through a speech act. Being "very economical in his information" and hence concealing the truth leads to a deception but not a lie.

Given these differences between deception and lies, it then becomes interesting to see how actions and statements can be constructed in order to bring about such "false beliefs".

3 Structure and Media

Just like any other human interaction, deceptive behaviour can be divided into two main groups: planned and unplanned. In planned interactions, people have time to think, reflect and compare situations with past experiences. They know or have time to consider knowing the person who they interact with (DePaulo, 2003). In unplanned interactions, people are not necessarily aware of actions that will happen which might need to be controlled. They are not fully aware of the person they will interact with and cannot guarantee the outcomes. Planned deceits should be harder to detect simply because the deceivers have time to rehearse their words and behaviours in order to present the impression of being truthful, or at least being more compelling.

Moreover, the choice of medium for communication can force the type of interaction. Based on Hancock, Thom-Santelli & Ritchie, 2004), deceptiveness in media relates to three main elements:

- **Synchronicity:** to what extent the medium provides real-time communication.
- **Distribution:** whether the people who are communicating are in the same physical location or not.
- **Recordability:** to what extent the medium is automatically recordable.

By knowing these, it is possible to argue that synchronicity and unplanned interactions are directly related, so media that are synchronous should be avoided for planned deceits as they give opportunity to discuss whilst deceivers might need time to rehearse their answers so will prefer asynchronous communication – for example, email.

If we focus on running text as the medium for deception, then while synchronicity and distribution are variable, recordability is certain. This will mean that most of the deceits can be planned well in advance, which could well make their detection somewhat more challenging. On the other hand, social media tends to assume greater degrees of synchronicity and a notionally lower distribution, so deceptions in social media may be more prevalent, not least because there can be less opportunity for planning. The next question, then, is what might be detectable. This brings us to the notion of deception "cues".

4 Deception Detection Cues

Possibilities of being able to formulate human deception processes have encouraged experts in many fields such as psychology, sociology, criminology, philosophy and anthropology to study such behaviour and look for cues as might indicate it. Researchers have shown that telling a lie or being engaged in deceptive behaviour is mentally, emotionally and physically more challenging than being truthful (Miller & Stiff, 1993; Zuckerman et al., 1981; Vrij, Edward, & Bull, 2001). It is emotionally challenging because deceivers might experience Fear and Threat (of being caught), Guilt and Shame (of deceiving someone and of having their trust questioned) or even Duping Delight (joy of deceiving someone). It is mentally challenging as deceivers need to create a story that is believable and consistent and try to remember what they are saying just in case they are questioned later (Miller & Stiff, 1993; Vrij, 2000; Zuckerman et

al., 1981; Cody, Marston & Foster, 1984; Vrij, Edward & Bull, 2001). It is physically challenging as deceivers usually attempt to control the physical signs of their deceptive behaviour (Buller & Burgoon 1996; Vrij & Mann, 2004). These attempts can give away a deception or a lie as it is not easy to hide nervousness and fear/guilt, remember lies in detail, and try to manage all of these to make an honest impression at the same time. These will result in behaviours which would be different from truthful actions, giving *Cues of Deception*.

In principle, almost any aspect of human existence that is involved in any action and behaviour may be carrying a cue to flag up deception; that can be eye movement, choice of words, arm positions or motions, and much more. One or many of these may be involved in a single communication, but some will be more specific to certain types of communication. For example, body language and eye movements are mainly considered in synchronous, non-distributed communication, while the structure of the sentence will be more apparent in recordable, distributed communication such as IMs and emails. Such cues can be readily grouped by the *3Vs*:

Visual (Non-verbal): any physical behaviour; reactions, movements, etc in three main groups of Body Acts, Postures and Face.

Vocal: elements that accompany verbal communication with two main features involved: Nature of voice (e.g. Tone/ Tension, Pitch) and Rhythm (e.g. number and the length of pauses).

Verbal: anything said or written (e.g. wording and structure).

However, it is important to note that the physical signs, the visual and vocal, cannot entirely be trusted since specific conditions may lead to similar effects. In certain circumstances, people will be naturally nervous or may feel fear simply because of a situation. For example, in an interview, and in particular in interviews with law enforcement officers, a cue to deception may be out of the normal for *that* interaction, whilst all parts of the interaction could indicate deception in contrast to everyday interactions (Navarro, 2008).

With our interest in detecting deception in text, we focus towards Verbal and in particular Written. Here, the deceiver must make words and patterns of those words do the work, and there is some expectation that this leads to different word usage and language patterns from those that might be considered, somehow, normal.

5 Verbal Deception Detection Cues

Three main types can be defined for verbal deception:

- **Spoken** (e.g. face-to-face, audio and video recordings)
- **Written** (e.g. blogs, emails, testimonies, academic articles)
- **Transcripts of spoken** (phonetic transcription, orthographic transcription)

However, recordings of speech will retain vocal elements which may offer cues, and transcripts may offer surrogates for pauses and retain the speech disfluencies ("ums", "ahs", "like", and so on). Written text, then, is possibly hardest to treat as the visual and vocal cues are missing in contrast to spoken and transcripts (Gupta & Skillicorn, 2006). Interestingly, this suggests that Web content could offer ready source material but with the significant challenge in terms of detecting deception in it as the deceiving authors of written content will have the opportunity to plan.

Many researchers have investigated the lexical, syntactic, and meta-content features of verbal deception, classifying pattern changes into three main dimensions: (1) Quantity; (2) Quality; and (3) Overall impression. *Quantity* changes relate to the number of words being used. *Quality* change focuses on the difference between the word choices but still on a quantitative basis. However, *Overall Impression* is based on human judgment from deceivers' verbalizations including such elements as friendliness, sounding helpful, serious, uncertain, and so on (DePaulo et al. 2003). We discard these cues due to reasons of subjective interpretation - judges (detectors) would need to be trained, and while something seems believable and helpful to one, it may not appear the same to others, and exploring inter-annotator agreement would become a distraction. We focus only on existing measurable cues that should be independent of a judge's training and so could be used by both humans and machines.

For Quantity and Quality measurements there are various hypotheses, different lists of cues, and even different expected changes. We have focussed more on studies where ideas have gained traction through adoption (citation and derivative exploration) by others. For example, Pennebaker's research has been adapted based on its style (word-by-word), accuracy and flexibility for both written and spoken text (e.g. Toma &

Hancock, 2010; Little & Skillicorn, 2008; Gupta and Skillicorn, 2006; Newman et al. 2003).

5.1 Generalized Cues

DePaulo et al. (2003) developed a list of 158 visual, verbal and vocal deception cues, extracted from an analysis of 116 research papers between 1920 and 2001. 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. As we will show, research since 2001 picks up on several of these cues, and we have been able to use DePaulo's coding system to cross-reference subsequent papers for our own purposes.

5.2 Frequency-based Cues

A number of researchers appear to make use of Pennebaker's Linguistic Inquiry and Word Count (LIWC) system to support their experiments and claims (e.g. Gupta & Skillicorn, 2006; Hancock et al., 2004; Keila & Skillicorn, 2005a, b, c). They mention that the cues defining deception according to Pennebaker involve:

Self-references: Using first-person singular (e.g. me, I and my) shows speaker ownership of a specific statement or event. This offers a link between the reality and the speaker, and as deceivers haven't experienced that link they will reduce the use of self-references.

Negative words: Emotions such as guilt, shame and fear may be attributed to the deceivers' discomfort (DePaulo et al., 2003) and the effect of negative emotions on the pattern of language is believed to lead to an increase in the use of negative words.

Cognitive complexity: As suggested earlier, cognitive complexity increases while deceiving. These effects become apparent in statements in various ways, which directly affects the structure of the text by changes in two main categories. **(a)** Exclusive words: Statements grounded in reality are more likely to highlight the details, including what happened and related reactions. Deceivers, lacking these details, use fewer exclusive words such as except, but, without and exclude. **(b)** Motion/action verbs: A decrease in exclusive words can result in an increase in action verbs (e.g. go, lead, walk) while trying to sound more assuring and convince others to take actions based on their words. Moreover, cognitively, it is easier to use simple and concrete actions in stating false stories compared to fake evaluations and retaining details.

However, we have so far found little evidence that Pennebaker has proposed cues for deception except for one research paper by Newman, Pennebaker, Bery & Richards (Newman et al., 2003). In that paper, the authors discuss cues previously offered by others (that relate to categories in LIWC) along with the reduction in the number of 3rd person pronouns, which contradicts previous studies such as Knapp et al., (1974). Subsequent authors have referenced such articles ambiguously, which may give the impression that LIWC itself offers the answer, for example, Hancock et al., (2004):

> *"[LIWC] was used to create empirically derived statistical profiles of deceptive and truthful communications (Pennebaker et al., 2003),…"*

and Gupta & Skillicorn (2006):

> *"Pennebaker et al. have constructed a model (LIWC) (Newman et al., 2003; Pennebaker, Francis & Booth, 2001) for deception based on the frequencies of various classes of words."*

Whilst LIWC can offer analysis of data, when it comes to understanding the behaviours of cues as might indicate deception by "increase" or "decrease" in frequency, there is no clear baseline. So, to be able to detect any deception, work would first need to be done in order to (1) establish the frequency ranges for different elements within a specific collection, (2) set thresholds of deception per-collection and per cue, and then (3) manually verify those above and below the deception threshold. Relationship to some collection-specific average is unlikely to readily produce appropriate results.

5.3 Category-based Cues

Burgoon and colleagues have categorized deception cues. However, Burgoon and other researchers have, without much explanation that we can find, varied the number of categories and also reported cues in different categories in different research papers (Burgoon & Qin, 2006; Qin et al. 2005; Qin, Burgoon & Nunamaker,

2004; Zhou et al. 2004; Zhou, Burgoon & Twitchell, 2003; Zhou et al. 2003; Burgoon et al. 2003). Indeed, they appear to add, delete, or otherwise emphasise different cues throughout their work. Neither the cues nor the threshold related to their deceptiveness appear stable. A set of cues that have been moved around categories is represented by Black cells in Table 1. Table 1 also shows, in gray, certain inconsistencies amongst these researchers: in Zhou et al. (2004), the number of words, sentences and the emotiveness index show an increase in cases of deception, but in Burgoon et al. (2003) and Zhou et al. (2003) all three are shown to decrease.

Burgoon and colleagues are not alone in offering a categorization; Pennebakers' LIWC categories would be related, modulo terminological and category variation. However, indications of expected values for such cues remain elusive and we only have information that some may rise whilst others may fall.

Cues	**(1)**		**(2)**		**(3)**		**(4)**		**(5)**	
Word	**	Q	+**	Q	+**	Q	-**	Q	-**	Q
Sentence	**	Q	+**	Q	+**	Q	-**	Q	-**	Q
Modifiers	-**	U	+**	U	**	Q	**	Q	--	--
First-person singular	--	--	-**	V	+**	V	-**	V	--	--
2nd person pronouns	--	--	-**	U	**	V	--	--	--	--
3rd person pronouns	--	--					**	V	--	--
Temporal details	**	S	+**	S	--	--	-**	S	--	--
Spatial details	**	S			--	--			--	--
Perceptual information	--	--	+**	S	--	--	-**	S	--	--
Affective terms	**	A	--	--	--	--	--	--	-**	S
Positive	--	--	+**	S	+**	A	-**	S	--	--
Negative	--	--	+**	S	+**	A	+**	S	--	--
Emotiveness index	--	--	+**	E	--	--	+**	E	-**	S
Lexical diversity	-**	D	-**	D	-**	D	-**	D	--	--
Redundancy	-**	D	-**	D	**	D	+**	D	--	--
Passive voice	**	V	+ **	V	**	V	+ **	V	--	--
Modal verbs	-**	U	+**	U	**	V	+**	V	--	--
Uncertainty	+***	--	-**	U	**	V	+**	V	--	--
Objectification	--	--	-**	V	+**	V	**	V	--	--
Typo errors	-***	--	+**	I	+**	I	+**	I	--	--

Quantity = Q; Complexity = C; Specificity = S; Affect = A; Activation /Expressiveness = E; Diversity = D; Verbal non-immediacy = V; Informality = I; Uncertainty = U; Vocabulary Complexity = VC; Grammatical Complexity = GC;
(1) Qin et al. 2005 (2) Zhou et al. 2004 (3) Zhou, Burgoon & Twitchell, 2003 (4) Zhou et al. 2003 (5) Burgoon et al. 2003
Gray= inconsistency in expected results; in Black= inconsistency in categories
[**] included, [***] mentioned but not highlighted

Table 1: Contradictions in Cues and Expectation

5.4 Evaluating the Cues

Despite commonalities in what can be and is being studied amongst DePaulo, Pennebaker and Burgoon, it is apparent that there is not yet a clear set of cues with predefined expected values that could be used for detecting verbal deception. However, without clear descriptions of how to interpret results it is also possible that results could have been misreported. To address this, we undertook a number of small experiments – mainly geared around repetition of previous reported experiments – to try to understand the behaviour of deception cues.

Our experiments involve analysis of the BBC article "'Visions link' to coffee intake" mentioned previously (BBC, 2009) with cues identified by Pennebaker, Mehl, and Niederhoffer, 2003), tests on academic work (we used 100 scientific abstracts[1], which we have no reason to believe are deceptive), and attempting to repeat an analysis of the Enron email corpus including the emails of the executives (Keila, and Skillicorn, 2005a, b, c). The latter of these is made all the more difficult by offering three differing numbers of emails for the analysis without

[1] MuchMore Springer Bilingual Corpus, Available at: http://muchmore.dfki.de/resources1.htm

details of how to obtain such a number from the full collection. Unfortunately, experiments all tended to support the idea that it would be hard to detect deception "in the wild" reliably, in part because deceptive texts may "hide" amongst non-deceptive. We can see how this might happen with a simple experiment using the online version of LIWC. We use the 7 LIWC categories, scaled by the maximum of each, for the 100 texts from the MuchMore Springer corpus. We then select the closest matching text (Nearest) from the first 10 to the coffee article (Coffee), and note that values for 5 of these 7 are already close together with differences for social words and cognitive words more marked, but still well within the ranges. A broad grain such as this is unlikely to be revealing.

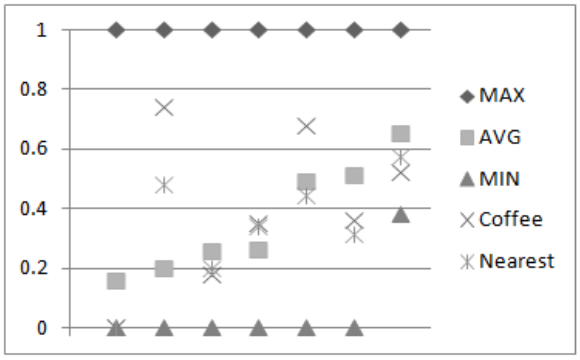

Figure 1: How a deceptive article might hide amongst scientific articles

6 Readability and Deception

Given variation in cues and expectations of values for those cues, a question arises of whether it is possible to provide some common, relatively well understood, and static baseline from which it would be possible to consider the variations in the values. Interestingly, various cues used in relation to deception also feature in Readability research, so might Readability scores offer such a baseline for comparison? Daft and Lengel (1984) argue that more ambiguous texts are more likely to contain deception, and such a claim has been supported in relation to fraudulent financial reports that contained more complex words, while truthful reports attained scores indicating better readability (Moffit and Burns 2009).

Historically, readability measures have been used to indicate the proportion of the population that would be able to understand a given text, but it has become apparent that word familiarity, cognitive load/complexity, cohesion, and other features of text contribute to its readability (Newbold and Gillam, 2010, Gray and Leary, 1935) and are also features considered in deception research.

Given the apparent overlap, we consider whether we might use readability measures to point more reliably to deceptive texts. Table 2 shows the cues covered by Gray and Leary (1935) for readability which are *also* studied as verbal cues for deception, along with expected direction of change in relation to readability and to deception (direction for the latter as suggested in e.g. Burgoon et al., 2003; Qin, Burgoon & Nunamaker, 2004)[2]. Not only is there an overlap with readability, but there seems to be a clear suggestion that more difficult texts are more likely to be deceptive.

Could such a clear relationship hold in practice? What would happen with articles such as "'Visions link' to coffee intake" or the 100 scientific abstracts? Scientific texts, and texts offering a misrepresentation *of* or *as* science, will probably both contain Big words, likely Nouns, may contain Rare words in contrast to general language, and possibly have relatively complex sentences. The writing style is also likely to impact on pronoun count. So systematic differences amongst such values might offer an indication of deception.

Cues	**R**	**D**
Big words	-	+
Nouns	- *	+
Verbs	*	+
Rare words	-	+
Sentence complexity	-	-
Number of first person pronouns	+	-
Number of second person pronouns	+	-
Number of third person pronouns	+	-
Average syllables per word and sentence	-	+
* may vary depending of the structure of the sentence and the words before and after them		

Table 2: Readability features and their relationship to Deception

Also, the online version of LIWC has a category for Big words (those with more than 6 letters). The values from this follow a similar pattern to that of Grade level for readability. For

[2] There are contradictions for expectancy rate for these cues so chosen expectations might conflict with other theories.

Coffee against 10 Springer articles, dividing Big word by Grade level provides the lowest value for the Coffee article. So, it is possible – indicatively, but not conclusively - that the ratio of Big words to Grade level could offer an indication.

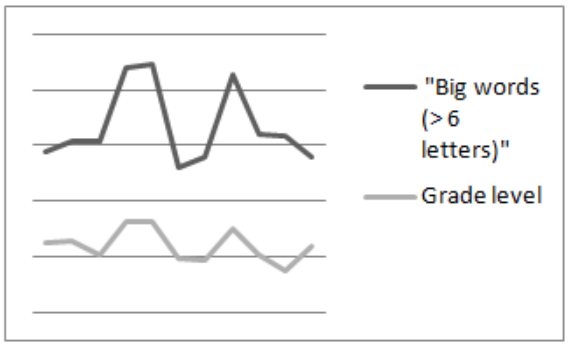

Figure 2: How Big words and Grade level tend towards indicating each other.

To explore how such a relationship might hold in practice, we consider a small experiment comparing essentially the same core content, but which results in different readability scores. A document that has been (supposedly equivalently) translated several times, albeit with particular variations, offers such a basis, and a good example of this is the Bible. We selected the Gospel of John (because it contains first person pronouns) from the following four Bibles[3];

- New International Version (NIV)
- New King James Bible (NKJB)
- King James Bible (KJM) and
- New America Standard Bible (NASB)

We are not suggesting here that the Gospel of John is general representative for the English language, nor that it should be seen to be deceptive per se, but as translations from a single source it should help to demonstrate any effect.

We choose four Pennebaker categories for our comparison. Since we have yet to find complete lists in Pennebaker's research, and since Newman et al. (2003) does not offer up full lists of words, we make use of the list of 86 words cited by Little & Skillicorn (2008) as being from Pennebaker. If all versions of the Gospel of John essentially contain the same content, and if we can use these categories for ranking purposes, we might expect to either see equal ranks for all four cases or to have the old versions (KJM and NASB) flagged up with higher ranks of deceptiveness.

Table 3 shows the scores for First Person (FP), Negative Words (NW), Exclusive Words (EW) and Motion Verbs (MV) as well as Grade Level and Reading ease score which shows, in terms of readability, NIV and NKJB are the better.

	FP	NW	EW	MV	Grade Level	Reading ease[4]
NIV	1.96	0.24	0.53	0.47	6.48	78.11
NKJB	3.44	0.12	1.00	0.48	7.24	77.21
KJB	2.29	0.28	0.69	0.37	7.78	74.48
NASB	2.04	0.23	0.65	0.44	8.38	73.88
Newman et al. (2003): Light Gray Little and Skillicorn (2008): Dark Gray						

Table 3: Variables for Deception and Readability for Gospel of John in 4 Bibles

For Newman et al. (2003), the deceptive text will have:

- Decreased frequency of first person singular pronouns ➔ NIV
- Increased frequency of negative emotion words ➔ KJB
- Decreased frequency of exclusive words ➔NIV
- Increased frequency of action verbs ➔NKJB, NIV

On the other hand, Little and Skillicorn (2008) expect a deceptive text should show:

- Increased frequency of first person singular pronouns ➔ NKJB
- Increased frequency of negative emotion words ➔ KJB
- Increased frequency of exclusive words ➔NKJB
- Increased frequency of action verbs ➔NKJB, NIV

Interestingly, these results suggest that the New International Version (NIV) and New King James Bible (NKJB) score higher on deception despite both having higher readability values. These results contradict what we would expect in relation to readability, further underlining the

[3]Accessed from link below for stability in structure and sentencing http://www.biblegateway.com

[4] Readability values from: http://www.online-utility.org/english/readability_test_and_improve.jsp

difficulty in relying entirely on the existing literature and leading us to question whether even readability offers gain at this grain.

7 Further critique

Analysis and experiments presented above suggest that difficulties emerge from present considerations of cues of deception – at least in relation to verbal deception. However, it is unclear whether this is a consequence of how the cues are being treated, or whether there are other biases which have a telling effect. In much of this research, conclusions have tended to be drawn on specific datasets, many of which are not readily available for inspection or use in repeat experiments by others.

The datasets were usually collected in one of three ways:

Role Playing: some are asked to deceive others (e.g. Hancock et al., 2005; Burgoon et al., 2003; Qin et al., 2005; Qin, Burgoon & Nunamaker, 2004).

Diary Keeping: individuals are asked to document their own interactions (e.g. DePaulo et al., 1996; Hancock et al., 2009; Hancock, Thom-Santelli and Ritchie, 2004). In this type of study, participants take time to document, once per day, their lies and to self score them based on dimensions such as seriousness, feelings while lying, and fear of getting caught.

Obtained as-is: (e.g. Keila, and Skillicorn, 2005a, b, c). Most such studies adopt Pennebaker's approach. This means that any classificatory thresholds have to be manually set and evaluated, via the means of the Human Eyeball.

All three approaches suffer from potential experimental effects. For the first two, it would be important to control for the *Hawthorne effect* which highlights that "observation and studies can change the behaviour of the participants" regardless of whether they should have really changed anything specifically in diary based studies (Franke, 1978; Jones, 1992). We believe during such studies peoples' behaviours might change, intentionally or otherwise. This might be because they become more cautious about perceptions of them by the researchers, want to avoid fear, shame, and so on, or feel uncomfortable with undertaking or documenting such an act. Indeed, the researchers may even be being deceived about the deceptions by the subjects. The third approach leaves the decision regarding actual deceptiveness of the text or statement open to the possibility that the researcher has been "primed by expectations". (Doyen et al., 2012).

8 Conclusion and Future Work

This paper has outlined a critical review of previous research in deception detection in order to assess whether it is possible to create deception detection. On present evidence, whilst there may be various important findings, there are too many areas open to question to believe that such a system could readily be constructed.

We still believe that previous deception detection research has a significant role to play, but many of the difficulties outlined in this paper need to be addressed first. Essentially, this requires a more systematic approach towards both datasets and treatment of cues. The public availability of deception-bearing texts covering different text types and genres would offer an ideal basis for such an approach, and a similar rigour in identifying cues tested, following DePaulo, would be highly beneficial. From this, it may be possible to identify specific cues as worth study in certain genres, whilst of little interest in others – irrespective of their relative frequency of use.

In absence of this, in our own near-future work we intend to explore the extent to which deception cues have also featured in tasks of plagiarism detection. Here, the datasets of PAN, and in particular as relates to authorship attribution and intrinsic plagiarism detection are of interest. Since the act of plagiarism is a deliberate attempt to deceive, such collections – albeit of a synthetic nature - offer us ready grounds for repeatable explorations and might lead to further insights into the general nature of the cues themselves.

9 References

BBC, (2009). Visions link' to coffee intake. *BBC News*. Retrieved 10.0d.2011 from http://news.bbc.co.uk/1/hi/health/7827761.stm

Buller, D.B. and Burgoon, J.K. (1996). Interpersonal Deception Theory. *Communication Theory*, 6, 203-242.

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, 91-101.

Burgoon, J.K. & Qin, T. (2006). The Dynamic Nature of Deceptive Verbal Communication. *Journal of Language and Social Psychology*, 25(1), 76-96.

Cody, M.J., Marston, P.J., & Foster, M. (1984). Deception: Paralinguistic and Verbal Leakage. In Bostrom, R.N. and Westley, B.H. (Eds.). *Communication Yearbook 8*. Beverly Hills: Sage. 464-490.

Daft, R.L. & Lengel, R.H. (1984), Information Richness: a New Approach to Managerial Behavior and Organizational Design. In Cummings, L.L. and Staw, B.M. (Eds.). *Research in organizational behaviour*. 6, Homewood, IL: JAI Press, 191-233.

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

Doyen S , Klein O , Pichon C-L, & Cleeremans A , (2012) Behavioral Priming: It's All in the Mind, but Whose Mind? PLoS ONE 7(1): e29081. doi:10.1371/journal.pone.0029081

Ekman, P. (1985). Telling lies, Clues to Deceit in the Marketplace, Politics, and Marriage. New York: W.W. Norton & Company.

Franke R.C. & Kaul J.D. (1978). The Hawthorne Experiments: First Statistical Interpretation. *American Sociological Review*, 43(5), 623-643.

Gray, W.S. & Leary, B (1935). *What Makes a Book Readable*. Chicago: Chicago University Press.

Gupta, S. & 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, 1-4.

Hall, H. V. & Pritchard, D.A. (1996). Detecting Malingering and Deception. Forensic Distortion Analysis (FDA). Boca Raton, FL: St. Lucie Press.

Hall, H. V. & Pritchard, D.A. (1996). Detecting Malingering and Deception. Forensic Distortion Analysis (FDA). Boca Raton, FL: St. Lucie Press.

Hancock, J.T., Birnholtz, J., Bazarova, N., Guillory, J., Amos, B., & Perlin, J. (2009). Butler Lies: Awareness, Deception and Design. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2009).*

Hancock, J.T., Curry, L., Goorha, S. & Woodworth, M.T. (2004). Lies in Conversation: An Examination of Deception Using Automated Linguistic Analysis. *Proceedings of Annual Conference of the Cognitive Science Society,* 26, 534-540.

Hancock, J. T., Curry, L., Goorha, S., & Woodworth, M.T. (2005). Automated linguistic analysis of deceptive and truthful synchronous computer-mediated communication. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS-38)*, Los Alamitos, CA: IEEE Press.

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)*, 129-134.

Jones S.R (1992). Was There a Hawthorne Effect? *American Journal of Sociology*, 98(3), 451-468.

Keila, P.S. & Skillicorn, D.B. (2005a). Detecting Unusual and Deceptive Communication in Email. *Centers for Advanced Studies Conference*, 17-20.

Keila, P.S. & Skillicorn, D.B. (2005b). Detecting unusual email communication. *Proceedings of the 2005 Conference of the Centre for Advanced Studies on Collaborative Research*, 117-125.

Keila, P.S. & Skillicorn, D.B. (2005c). Structure in the Enron Email Dataset. *Computational and Mathematical Organization Theory*, 11(3), 183-199.

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

Little, A. & Skillicorn, B. (2008). Detecting Deception in Testimony. *Proceeding of IEEE International Conference of Intelligence and Security Informatics (ISI 2008)*, June 17 - 20, 2008, Taipei, Taiwan, 13-18.

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

Mahon, J. E. (2008). Two Definitions of Lying. *International Journal of Applied Philosophy, 22*(2), 211-230.

Moffitt, K. and Burns, M.B. (2009). What Does that Mean? Investigating Obfuscation and Readability Cues as Indicators of Deception in Fraudulent Financial Reports. Proceedings of Americas Confernece on Information Systems (AMCIS 2009) , 399.

Masip, J., Garrido, E. & Herrero, C. (2004).Defining Deception*, Anales de Psicologia*, *20*(1), 147-171.

Miller, G.R. & Stiff, J.B. (1993). *Deceptive Communication*. Newbury Park, CA: Sage.

Navarro, J. (2008). "What Every BODY is Saying: An Ex-FBI Agent's Guide to Speed-Reading People." New York. Harper-Collins.

Newbold, N. & Gillam, L. (2010). The Linguistics of Readability: The Next Step for Word Processing. *Workshop on Computational Linguistics and Writing: Writing Processes and Authoring Aids (CLandW 2010)*. June 6, 2010, Los Angles, 65-72.

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), 665-675.

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

Pennebaker, J.W., Francis M.E. & Booth, R.J. (2001) *Linguistic inquiry and word count (LIWC)*. Erlbaum Publishers.

Pennebaker, J.W., Mehl, M. & Niederhoffer, K. (2003). Psychological Aspects of Natural Language Use: Our Words, Our Selves. *Annual Review of Psychology*, 54(1), 547-577.

Qin, T., Burgoon, J.K. & Nunamaker, J.F., Jr. (2004). An Exploratory Study on Promising Cues in Deception Detection and Application of Decision Trees. *Proceedings of the 37th Hawaii International Conference on System Sciences*, January 5-8, 2004, Waikoloa, HI, 23-32.

Qin, T., Burgoon, J. K., Blair, J. P., & Nunamaker, J. F. (2005). Modality Effects in Deception Detection and Applications in Automatic-Deception-Detection. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, 23-23.

Tausczik, Y.R. & Pennebaker, J.W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29, 24-54.

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*, 899-909.

Vrij, A., & Mann, S. (2004). Detecting Deception: The Benefit of Looking at a Combination of Behavioral, Auditory and Speech Content Related Cues in a Systematic Manner. *Group Decision and Negotiation (special deception issue),* 13, 61-79.

Zhou, L., Burgoon, J. K., & Twitchell, D. P. (2003). A longitudinal analysis of language behavior of deception in e-mail. *Proceedings of Intelligence and Security Informatics, 2665,* 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.

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*, January 6-9, 2010, Waikoloa, HI, 10-19.

Zuckerman, M, DePaulo, B.M. & Rosenthal, R. (1981). Verbal and Nonverbal Communication of Deception, In Berkowitz, L.(Ed.). *Advances in Experimental Social Psychology, 14*, 1-59.