

Revealing Online Deception:
The discrepancy between deceptive belief and practice online

Honors Thesis

Presented to the College of Agriculture and Life Sciences, Social Sciences
of Cornell University
in Partial Fulfillment of the Requirements for the
Research Honors Program

by
Cameron Wilton Hall
May 2008
Jeffrey Hancock

Abstract

Do we really encounter as many lies as we think online? The present study examines the possibility of a *prevalence paradox*, or the discrepancy between one's beliefs about deception online and actual encounters with deception. This research tests several factors that may influence the prevalence paradox in mediated communication, including characteristics of the user (e.g., experience) and of the communication environment (e.g., synchronicity and evanescence). Participants reported on their beliefs and experiences with deception in email, instant-messaging, blogs, and social networks. Although the results from this study did not support the predicted factors for the prevalence paradox, they did confirm the existence of the paradox in each of the communication environments.

Revealing Online Deception:

The discrepancy between deceptive belief and practice online

People tend to hold strong beliefs about deception, from what kinds of cues may reveal deception to what actually counts as a lie. For the purposes of this discussion, we will define deception as either a successful or an unsuccessful deliberate attempt, without forewarning, to create in another a belief, which the communicator considers untrue (Vrij, 2000). One belief or assumption held by many deception researchers is that deception occurs with great frequency in our everyday communication (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996). It is difficult, however, for people to believe that their daily interactions contain deception. When communicating, we are trying to make sense of a message based on the assumption that the message is true. Without this assumption, fostering any type of relationship or meaningful conversation would be difficult. It follows naturally then that people in a functioning society believe that lying takes place less than it actually does. The *truth bias* explains the tendency to assume that a communicator's message is truthful (Vrij, 2000).

While the truth bias explains a predisposition to trust our communication partners, it does not explain why we continue to maintain a certain level of trust. In other words, if we are lied to so much, why do we continue to be so trusting? The main explanation is that people typically do not know when they are the object of deceit. We are poor lie detectors in large part because we rely heavily on non-verbal cues to detect deception. People believe cues, such as diverted eye gaze and fidgeting, can reveal deception. This is in part because these cues differ from normal interactions. Prior research defines perceived nonverbal deviations from normative expectations as expectancy-violations

(Bond, Omar, Pitre, Lashley, Skaggs, & Kirk, 1992). Vrij (2000) argues that not only do deceivers, in general, tend to focus more on the content of their deception, thus ignoring their nonverbal behaviors; they are simply not skilled at suppressing these cues. Despite these findings, research shows that nonverbal cues are unreliable indicators when it comes to catching a liar (Vrij, 2000). Thus, the reliance on these cues is misleading and inaccurate.

The Internet has opened a new world of communication and has changed the way we look at deception. In most computer-mediated communication (CMC), we cannot see our communication partners and nonverbal cues are eliminated. George, Marett, and Tilley (2006) argue that in CMC the medium acts as a filter, even for leaked cues, blocking the receiver from perceiving them. Without the nonverbal cues that people believe facilitate deception detection, CMC users feel somewhat handicapped when it comes to detecting deception online. This may result in a certain level of mistrust, and for this reason, it may be reasonable for people to believe that lies are more prevalent in CMC than face-to-face (FtF). We believe that people follow an *ease of lying heuristic* where people believe that the easier it is to tell a lie the more likely a person will lie. As stated before, because CMC does not permit the leakage of nonverbal cues, it appears that lying would be easier in CMC than FtF. It follows then that in a CMC environment, without the presence of nonverbal cues, people may be more likely to expect deception because we assume it is easier to accomplish.

Despite any distrust of the Internet, lying online is not as prevalent as people might believe. In one study of deception, Hancock, Thom-Santelli, and Ritchie (2004) explored the frequency of deception across various communication technologies. Their

results showed that lying on the Internet is not particularly new or exciting. In fact, Hancock et al. (2004) found that the overwhelming majority of lies occurred in a basic FtF setting. In addition, telephone conversation had the highest rate of lying, followed by FtF, instant-messaging (IM), and least of all email.

Despite these empirical findings, Caspi and Gorsky (2006) found that 73% of all participants believed deception is widespread online, while only 29% reported that they themselves lie online. Furthermore, participants reported that there were few, if any, attempts were to deceive them online. This evidenced a clear discrepancy between people's beliefs and actual practice of deception online.

As Caspi and Gorsky (2006) established, people believe the internet is rife with deception. However, evidence from Hancock et al. (2004) suggests that lies may not even occur with the greatest frequency online. Furthermore, according to Caspi and Gorsky (2006), despite the fact that people believe lying is widespread online, they do not report lying themselves with a proportionate frequency. As a result, there is a significant difference between how much people believe lying occurs online and how much they personally experience it. The *prevalence paradox* defines this particular phenomenon, which suggests that people are distrustful of the online environment despite the world's increasing dependence on the medium both professionally and socially.

As stated, we believe that the primary reason for such a belief rests in the fact that the environment is stripped of non-verbal cues, which people believe they can depend on for detecting deceit. This makes intuitive sense, because in the absence of these cues people believe lying is easier and should therefore occur with greater frequency (i.e., the ease of lying heuristic). Importantly however, research to date has focused only on

frequency and not on the magnitude of lies online or the different possible communication environments available online (e.g., chat vs. email, blogs). For instance, although Caspi and Gorsky (2006) concluded that people believe the online environment is inherently deceitful, they base their conclusion on findings drawn from a study focused only on newsgroup participants. As a result, they fail to make a distinction between various types of environments. However, communication environments have varying features that distinguish them from one another. Accordingly, this study will focus specifically on the beliefs people hold about deception on differentiated CMC environments — namely email, instant-messenger (IM), blogs, and social networks. This deeper focus will allow us to not only demonstrate the presence of the prevalence paradox, but also offer further insight as to why it exists.

As mentioned, we believe the lack of non-verbal cues online is the primary factor in producing an inherent distrust of the online world. However, the different CMC environments, for the most part, uniformly strip non-verbal cues. Therefore, the lack of non-verbal cues cannot be a driving factor in predicting or determining where people believe deception will occur with the greatest frequency. What becomes important then is determining the factors associated with a lack of non-verbal cues that breed distrust online, and cross apply those factors to differentiated CMC environments.

Given the ease of lying heuristic, if we determine what people believe makes deception easier online, we can infer which communication environments they believe will be the most deceitful based on characteristics of those environments and characteristics of the sender and receiver. Carlson and Zmud (1994) took a similar approach when expanding on the media richness theory. They found that personal factors

could expand the potential bandwidth provided by a communication environment. We believe there are three factors that significantly influence the perceived ease of deception online; experience with the communication environment, synchronicity of the communication environment, and perceived social distance.

Experience with a communication environment may play the greatest role in predicting where people believe that most lying will occur online. As experience increases within a certain communication environment, people develop a set of normative expectations for how a conversation or exchange will progress. As one repeatedly confirms these expectations through experiences, a person's confidence in detecting deception increases. They expect to be able to recognize any deviation from their expectation and in turn expect to be able to catch most liars. As a result, they do not expect lying to occur with any great regularity in the communication environment in which they have significant experience. Furthermore, as experience increases, they may automatically filter out deceptive behavior without even knowing it. For example, someone who uses email all the time may automatically disregard spam mail without even considering it deceptive. The effect due to experience is hard to predict because experience with various communication environments change from person to person. However, it may be explainable in part by the communication environment. Email was the first wide spread communicative communication environment, followed by IM, blogs, and finally social networks. Applying this historical progression might predict that people have the most experience with email and the least experience with social networks.

Each communication environment has a defined level of synchronicity. The user may control exactly how synchronous a specific exchange may be, but the communication environment allows for a particular range of synchronicity. Accordingly, IM is the most synchronous, followed by email, blogs, and social networks. IM and email both encourage a response from a specific person, but conversations in IM essentially take place in real time, as long pauses between exchanges are typically inappropriate. Email, on the other hand is more of a message awaiting reply at the convenience of the receiver. As such, the sender may expect a considerable wait before a response, depending on the nature of the message. Exchanges on blogs and social networks are often not directed at any particular person and there is no expected time frame for a response. Therefore, we are required to evaluate synchronicity based on social norms and perceived synchronicity. We expect that people will report IM as the most synchronous, while blogs and social networks are the least synchronous. This builds the basic premise is that IM grant the least opportunity for planning or tailoring deception, while blogs and social networks allow for much more.

Perceived social distance is also largely a function of the specific communication environment. The further removed a CMC environment makes two people feel from each other, the more they will believe deception occurs. Essentially, the more disconnected people feel relative to social distance the less confident they will feel in judging character and detecting deception. The effect follows the idea that as cues people rely on to detect deception are stripped away, people feel more vulnerable to lies and expect more deception. As people feel further and further removed from the source of a message, they feel more stripped of cues and their expectation for deception increases.

Previous research not only failed to differentiate between CMC environments (e.g., Caspi & Gorsky, 2006), but also between differing types of deception. It is important to know not only whether or not deception is happening, but also the nature of the deception. People will often refer to a deceptive statement as “just a little white lie.” This type of deception is usually innocuous and would be something like attributing your tardiness to traffic, when you really just woke up late. Deception increases in magnitude from there. The consequences of a lie increase with the magnitude of the lie. If your mother cannot find her necklace and you lie and say, you have not seen it when you actually broke it; you have left the realm of white lies. This time, the consequences are more severe. Suppose your father cannot find your mother and ask if you have seen her. If you tell him no, when you left her at the grocery store, the consequences could be severe. Obviously, these types of lies are very different, but in Caspi and Gorsky’s (2006) study, they are treated as equal. The present study will differentiate between lies of varying magnitudes to see which CMC environments people believe will facilitate lies of the greatest magnitude.

We believe that the most influential factor in determining where people think the lies of greatest magnitude will occur is verifiability of the lie. Essentially, the environment wherein people believe it is the hardest to be caught lying will be the environment wherein people believe the most serious lies will occur. We believe there are characteristics specific to the CMC environments that will affect people’s beliefs about potential consequences, such as the evanescence of the message contained in the specific communication environment. Unlike face-to-face communication, online communication maintains a written record of the message. The length of time the message is stored and

the nature in which it is stored vary across different communication environments.

Communication environments that users believe do not store messages, or do so for a short period, and where the messages are not readily accessible, should promote lies of greater magnitude.

In IM, the conversation is conventionally stored only during the current interaction. At the closing or termination of an interaction, the records are lost. On the contrary, email messages are stored until deletion. However, there is a record for both the sender and receiver of an email. For a record to be lost, both parties must delete it. Blogs and social networks share similar properties: they both display content until their removal by the author or administrator. However, this does not ensure complete removal because external sources may download the information and store it without notice. Assuming these conventions for message, storage and retrieval we expect to see the lies of greatest magnitude in IMs, followed by blogs or social networks, and of the least magnitude in email. To disambiguate the difference between evanescence in a blog post and social network profile, we will examine people's beliefs about retrievability for the communication environments.

Other factors may influence where people believe lies of larger magnitude occur. Nonetheless, the evanescence of the message should be the strongest predictor for beliefs about lies of greater magnitude because this is what people will think of as the most important factor in getting caught or not. If a message disappears when the conversation ends, one can be confident that if they successfully deceived during an instance, the lie will not reappear later. However, if the lie sits in cyber space, accessible to anyone, the lie could reappear at anytime.

Using the previously presented theories and heuristics, this study proposes four hypotheses. First, increased experience with the communication environment leads users to believe they can better detect deception. As experience increases, the user feels more confident they can rely on cues hidden to a user with less experience. Thus, an experienced user will believe it is harder for him or her to be deceived and subsequently that lying must occur less. Second, the more synchronous a CMC environment is the less people will believe that lying occurs because longer response times provide a better opportunity for someone to plan their deception. This makes it easier to deceive, thus leaving the user to feel more vulnerable to deception. Third, the perceived social distance inherent to communication environment will play a significant role in determining where people believe the most deception will occur. The further removed someone feels from the person they are communicating with, the more susceptible to deception that person feels. As a result, if a CMC environment increases social distance between people, it in turn increases the belief that more deception is occurring. Finally, as the evanescence of a communication environment increases the magnitude of the lies in the environment will decrease. If it is easier to check a deceptive statement, people will assume the deceit less significant. Thus, if it is easy to retrieve a previous conversation, the lies in the conversation should be little white lies.

To test these hypotheses, the present exploratory study conducted a survey of college students to examine their beliefs about deception. The subjects reported on their use and perceived characteristics in each of the aforementioned communication environments. They then reported on their beliefs and experiences with deception in the

communication environments provided they claimed to have previously used the environment.

Method

Participants

Since this topic has remained rather untested and we are examining beliefs, conducting an exploratory online survey was the best option. The exact population would include all internet users that have used email, instant messaging, blogs, and social networks, therefore we decided on a convenience sample. Participants were undergraduate students from Cornell University. Most, but not all of the subjects, received participation credit in an Information Science or Communication course for completing the study. Although it is impossible to calculate a traditional response rate due to the recruitment method described below, we did have a 97% completion rate among those who registered for the study. There were 206 subjects: 76 males, 130 females, with an average age of 20. Additionally, the subjects were classified into one of 15 majors of study covering most of the colleges at the university. The most common major was communication at 17% followed by biology at 15% and engineering at 10%. The least common major was animal science with 1%. All subjects completed the survey, but some questions were omitted if the subject lacked experience in a communication environment. All students reported experience with email and instant messaging, but only 158 and 171 reported having experience with blogs and social networks respectively. This sample is certainly a subset of the population being educated, college internet users, however generalization issues are noted in the discussion.

Apparatus

We created an online survey to record the subject's beliefs and experience with deception online. We opted against a survey service and built the survey in-house to ensure security, flexibility, and credibility. The survey consisted of 47 questions, but could be fewer if the subject lacked experience in a communication environment as previously noted. In addition to their beliefs and experience, subjects reported their gender, age, and major. All other questions were on recorded using a 7-point Likert-scale. First, both the scale and the questions were evaluated by a pilot study. In the pilot study, subjects were given open-ended questions to explain their beliefs and usage patterns. Additionally, we were able to reword our questions based on feedback and response patterns. The resulting close-ended questions helped ensure accurate results, while minimizing time required to complete the survey. Subjects could complete the survey over standard or secure http access. The survey data was stored in a MySQL database and the participation records were store in a protected text file. Hash codes protected the subject's identity in the database.

Procedure

Students were recruited by asking them to complete a brief survey concerning online deception. Unlike the Caspi and Gorsky (2006) study which used the channel under examination to recruit subjects, we used a selected audience. Specifically, we utilized class announcements to recruit participants in addition to posting a participation flyer near other experiments. At that time, the students received the survey's web address. After visiting the site and reading the informed consent, the subject entered their full name and NetID (email address) to register for the survey. An automatic email sent to the subject then provided a unique link to the survey. Upon following their link, the subject

was asked again for their NetID to confirm their identity. Once their identity was confirmed they began the survey.

The survey was broken into two bundles of questions. The first bundle of questions addressed perceived characteristics of the communication environment and the second examined beliefs and experiences. Each bundle contained four sets of questions, one for each communication environment. The first communication environment examined was email, followed by instant messaging, blogs, then social networks. Both bundles followed the same sequence and was identical for all participants.

The subject first completed items concerned with each communication environment's general characteristics, referred to as the General Characteristics Bundle of questions. These questions included: the subject's experience level, who they communicate with, reasons for their communication, the level of synchronicity, and evanescence. If the subject reported having never used the communication environment, they were only asked questions about their beliefs about a communication environment (i.e., synchronicity and evanescence). Appendix A illustrates the structure of the General Characteristics Bundle. This bundle represents the independent variables in the study.

After completing the General Characteristics Bundle, subjects were given the Beliefs and Experience Bundle. If the subject reported having experience in a communication environment, they were asked about both their beliefs and then experience with the communication environment. If the subject previously reported having never used the communication environment, they were only asked for their beliefs about the communication environment. The first two questions were general or belief questions. They asked how often people lie in the communication environment and to

what extent the lies are little white lies. Four experience questions then followed which examined the frequency and magnitude of lies. In the experience set, the first two asked when they were the subject of deception and the last two were for when the subject was the deceiver. Appendix B demonstrates the structure of the Beliefs and Experience Bundle and the dependent variables under examination.

Although these constructs have taken the form of individual questions, they are addressed for each of the environments. In other words, we are not simply asking if a participant is lied to in an individual environment, but instead examining if they are lied to in each of the environments.

If a question was skipped at any point during the survey, the subject was sent back to the page and required to answer the missing question. However, a subject could not change their answers once it was submitted. The subject was required to complete the entire survey during a single session, which took approximately 10 minutes.

Once the survey is completed, the subject was asked to select the class for which they would like to receive credit. If the subject did not take the survey for credit, they could simply close their browser.

Due to the exploratory nature of this study, we were unable to test the external validity measures. However, the dataset was proven reliable by confirming the predefined characteristics of each CMC environment. For instance, the subjects accurately reported the synchronicity order of the environments as described in the introduction. This finding was repeated in two scenarios which are further discussed in the results.

Once the survey ended, we examined the data for incomplete entries and other possible problems. We then recoded the majors into 15 categories and prepared the data

for SPSS. SPSS is a statistical data software package that provides the means for an array of test. We ran multiple analysis of variance (ANOVA) and Spearman tests to examine our hypotheses in addition to other interesting patterns.

Results

The first hypothesis predicted that those with higher experience would perceive less deception online. The only statistically significant finding was for social networks, which revealed a positive correlation of $r = 0.23$ (see Table 1), suggesting that those who reported having more experience with social networks believed there is more deception in social networks. This pattern is the opposite of the hypothesis. Therefore, we are unable to support the first hypothesis that higher levels of experience will lessen perceived deception.

The second hypothesis predicted that synchronous communication environments have lower perceived deception. Two communication environments returned statistically significant results, but only one supported the hypothesis. Email had a negative correlation of $r = -0.16$ (see Table 2), suggesting that those who believe email is asynchronous also believe deception is more prevalent. While email supports the hypothesis, there was a positive correlation of $r = 0.17$ between synchronous blogging and perceived deception (see Table 2). Subjects reporting that blogging is asynchronous did not believe there is a lot of deception in blogs. Therefore, we are unable to support the second hypothesis that the more synchronous a communication environment the lower perceived deception. Although the hypothesis was not supported using subjects perception of synchronicity, the data did confirm our assumptions about synchronicity among the communication environments (see Table 3). Using a paired T test confirmed

that IM was the most synchronous followed by email, social networks, and then blogs at the 0.01 significance level. In other words, the subjects accurately portrayed the synchronicity levels in each of the communication environment, which supports the reliability of the sample.

The third hypothesis predicted higher social distance with increased perceived deception. Only one communication environment returned statistically significant findings, but the correlation contrasts with the hypothesis. There was a positive correlation ($r = 0.18$) in social networks for communicating with people known personally and perceived levels of deception (see Table 4). In other words, people communicate with people that they know personally in social networks and believe their communication or profiles contain lies. Therefore, we are unable to support the third hypothesis that higher social distance will lead to increased perceived deception.

The final hypothesis predicted that lies of lower magnitude would be observed in communication environments with higher recordability. There were no statistically significant findings either for or against the hypothesis (see Table 5). Therefore, we are unable to support the fourth hypothesis that recordability influences the magnitude of lies. However, similar to the second hypothesis the predictions about the communication environment were accurate and statistically significant (see Table 6). Email has the greatest recordability followed by social networks, blogs, and then IM when using a paired T test at the 0.01 significance level. Once again, the subjects accurately portrayed the synchronicity levels in each of the communication environments, which supports the reliability of the sample.

Prevalence Paradox

After failing to support the above hypotheses, we tested the assumption that the prevalence paradox exists in the dataset. We performed a paired T test to examine the prevalence paradox in each of the communication environments. To compute the prevalence paradox we subtracted each subjects reported encounter or lied to frequency from their general belief frequency. If the result is a positive value, the subject believes there is more deception in the communication environment than they reportedly encounter. Therefore, a positive value illustrates the existence of the prevalence paradox. As seen in Table 7, the existence prevalence paradox was positive and statistically significant for all of the communication environments, suggesting that the paradox does exist in each of the four environments.

In addition to checking for the existence of the paradox, we also analyzed the regression coefficients holding the general frequency of lies as the dependent variable. For each of the environments, the “lied to” value had the greatest *b* value except in email where it is second to lying (see Table 8). Although this is not a causal relationship, it does suggest that one’s encounters with deception influence their general belief about deception. In other words, one’s encounters or lied to frequency has the greatest influence on their general beliefs about deception, but frequently encountering lies does not cause one to believe deception is more prevalent in the communication environment.

Discussion

This study examined the existence of the prevalence paradox in email, IM, blogs, and social networks. The results demonstrated that the prevalence paradox exists in each of the communication environments. Additionally, we identified three predicting factors to the prevalence paradox. Although a few instances supported our predictions, we were

unable to support the hypotheses. Below we discuss each of the factors and possible reasons for why the predictors failed in addition to their implications for understanding the prevalence paradox.

The prevalence paradox

The support for the prevalence paradox across the different communication environments was statistically supported. The paradox demonstrates that we encounter a disproportionate amount of deception online from what we believe to generally exist in the communication environment. In other words, we encounter a lower frequency of deception in email, IM, blogs, and social networks than what we believe exist in each of the environments overall. While original support for this finding was limited to newsgroups (Caspi & Gorsky, 2006), demonstrating its existence in other communication environments provides ground for further research of the phenomena. Although this study covers the major online communication environments, other CMC environments remain untested.

The study was also interested in what types of factors lead to the prevalence paradox, including the experience of the user and characteristics of the communication environment. Below each factor is discussed.

User Experience

The first hypothesis predicted that more experience with a medium would lead to less perceptions of deception in that medium. In general, the data did not support this hypothesis. There was, however, a significant correlation between experience and perceived deception in social networks. This positive correlation suggests that those with more experience in social networks believe deception is more prevalent. One would

expect that social networks had the strongest correlation because it is the youngest communication environment. Similar to Wether's (1996) hyperpersonal model, we over attribute the general characteristics of online communication environments. If we encounter a few lies in our first experience, we will believe the environment is plagued with deception. Since we do not encounter lies frequently, one would expect that users with experience believe there is less deception than those without experience. However, when comparing the general belief of deception frequency for subjects with no social network experience and those with experience we found a difference of 0.57 (see Table 9). This means that there is no statistically significant difference between those with and without experience on social networks regarding their general beliefs about the frequency of deception.

We also found that there was a strong correlation between communicating with someone known personally and being the target of deceit on online social networks (see Table 4). This positive relationship claims that the more someone interacts with their friends on social networks, the more they feel like the target of deception. Since one presumably interacts more frequently with friends, they have more opportunities to identify deception. In addition to having more opportunities, friends know more about one another and are more likely to identify a lie that is not identified by a stranger. This relates to the concept of the truth-bias and trusting strangers. As described by O'Sullivan (2000), we strategically present ourselves to others in different communication environments. This includes being able to afford little white lies on online social networks because it more difficult to check facts. Since we have a truth-bias towards strangers and are unable to check potential lies in online social networks, the correlation

between communicating with people known personally and being lied to should be positively correlated. Additionally, this aligns with the regression analysis suggesting that being the target of deceit influences our overall perception in a communication environment. Although being lied to does not cause us to perceive more general deception, it has an influence.

Finally, the questions concerning experience required a range of experience levels for accurate predictions concerning its influence on the beliefs about deception. The mean experience levels were 6.61, 6.39, 4.11, and 5.74 for email, IM, blogs, and social networks respectively. In other words, the subjects reported having lots of experience in all the channels except blogs. The third-person effect explains that subjects believe they are at less risk of persuasion than others (Paul, Salwen, & Dupagne, 2000). Further exaggerations take place when examining college students and self-reporting higher experience levels demonstrates the subjects' higher assumed intelligence. Since users self-reported their experience level, it is difficult to determine a user's actual experience level when compared to others. Along the same lines, the subject pool is required to use email frequently, which may artificially inflate their scores when compared to the general population. Lastly, the wording of the question might also influence the inflated experience levels. The question asked them to rate their experience level, not their level of expertise. While not all of us are experts, many of us use the communication environment frequently and thus have a high level of experience.

Beliefs about medium characteristics

Through the analysis of the second hypothesis, we find that people believe that there is little deception in blogging, even though it is asynchronous. This is the opposite

of both the hypothesis and the case with email. This positive relationship means that people do not believe being able to edit a blog post correlates with deception. In fact, they believe that since blogs are not synchronous, they contain less deception. To explain such a finding, one must examine the relationship between the blogger and the reader. The average rating of people that use blogs to communicate with people they know personally was 2.46. This means that people are using blogs to communicate with strangers. We can therefore claim that people are more likely to trust blog posts from those that they do not know personally. When communicating with strangers we often employ the truth-bias, but also retain fragile trust (Bos, Olson, Gergle, Olson, & Wright, 2002). If a blog reader encounters deception, they are likely to remember the deceit and retain distrust in for the blog. Since blogs are typically multiple blogs on a singular topic, one is likely to abandon a distrustful blog for a new one. Readers understand that accurate reporting and ethics provide the foundation for the blogging community. Walther's (1996) hyperpersonal model is again employed as blog readers because our trust is continually confirmed by blog posters, which leads to the over attribution of truthfulness. Therefore, in communicating with strangers through blogs, we are more likely to trust what we read.

In general, the social network results were the opposite of what was expected. According to the ease of lying heuristic, people should assume fewer lies on social networks because there are non-verbal cues in addition to verbal cues. The deviation from the heuristic could be due to an increase in the opportunity to lie. Unlike the other communication environments, information in social networks does not pertain to a singular topic. Since there are more opportunities to deceive, people may believe that deception is more prevalent. However, the subjects reported that most of these lies were

little white lies. Such lies could range from editing a profile picture to befriending someone who is not a friend, which are common practices in social networks. Therefore, we likely overlook little white lies because of their insignificance.

The lack of statically significant findings suggests additional factors are at play. One such possible factor is filtering noise. This means that users might not even acknowledge messages considered deceptive by others, but rather perceive them as noise in the communication environment. Benway (1998) defines the filtering of consistent noise as banner blindness. For example, users ignore potential spam emails, phishing, or bots, instead of classifying them as deceptive. Users with more experience will do this more naturally, but users with less experience still filter out the noise even though it takes more time. It is possible that subjects would report more deception if they were asked to visit the communication environments and count the instances of deception. However, in doing so, it would be impossible to test for the prevalence paradox, as all the subject would be the target of all lies.

Implications and limitations

The most obvious limitation is the restrictions in the sample population. While limiting the population to undergraduate students at Cornell University limits the abstraction of the results, it does provide insight into a sample immersed in the technology. This could influence the factors that contribute to the prevalence paradox. Although repeating this study on a national level may return different results for the aforementioned predictors, it would likely also support the prevalence paradox.

The support for the prevalence paradox does in fact have implications. Although the population was limited, the prevalence paradox does exist in each of the four CMC

environments. While this study fails to identify particular contributing factors to the prevalence paradox, it does support its application to CMC environments other than newsgroups. Future research could explore communication environments that are untested by the prevalence paradox. Doing so may reveal a better understanding for the characteristics that contribute to the formation of the prevalence paradox. Another possibility could force those with less experience in a channel to explore it for a period of time and see how their beliefs change. This experimental methodology would provide a stronger insight to the prevalence paradox.

Finally, this study demonstrates the importance of beliefs about deception. It raises questions about the classification and recall of deception as it relates to self-reporters. This study further supports the discrepancy between perceived and reported encounters with deception online, but the cause of the discrepancy is still unknown.

References

- Benway, J.P. (1998). Banner Blindness: The Irony of Attention Grabbing on the World Wide Web. In *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 42, 463-467.
- Bond, C.F., Omar, A., Pitre, U., Lashley, B.R., Skaggs, L.M., & Kirk, C.T. (1992). Fishy-looking liars: Deception judgment from expectancy violation. *Journal of Personality and Social Psychology*, 63(6), 969-977.
- Bos, N., Olson, J., Gergle, D., Olson, G., and Wright, Z. (2002). Effects of four computer-mediated communications channels on trust development. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves*, 135-140.
- Carlson, J.R., & Zmud, R.W. (1994). Channel expansion theory: A dynamic view of media and information richness perceptions. *Academy of Management Best Papers Proceedings*, 280-284.
- Caspi, A., & Gorsky, P (2006). Online deception: Prevalence, motivation, and emotion. *CyberPsychology & Behavior*, 9(1), 54-59.
- 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, 979-995.
- George, J.F., Marett, D., & Tilley, P. (2004). Deception detection under varying electronic media and warning conditions. *System Sciences*, 1(1), 9-17.
- Hancock, J.T., Thom-Santelli, J., & Ritchie, T. (2004). Deception and design: The impact of communication technology on lying behavior. In *Proceedings of CHI 2004*, 130-136.

O'Sullivan, P. B. (2000). What you don't know won't hurt Me: Impression management functions of communication channels in relationships. *Human Communication Research, 26*, 403-431.

Paul, B., Salwen, M. B. & Dupagne, M. (2000). The Third-Person Effect: A Meta-Analysis of the Perceptual Hypothesis. *Mass Communication and Society, 3* (1), 57-85.

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research, 23*, 3-43.

Vrij, A. (2000). *Detecting Lies and Deceit*. Chichester: John Wiley & Sons.

Appendix A

Bundle of questions about user experience and beliefs regarding a communication

space

Channel	Question
email	I have a lot of experience using email.
	I use email to communicate with people that I know personally.
	I use email mostly for work.
	<i>I believe that email takes place in real time and feedback is immediate.</i>
	<i>I can easily retrieve an old email.</i>
IM	I have a lot of experience...
	I use instant messaging to communicate with...
	I use instant messaging mostly...
	<i>I believe that instant messaging takes place...</i>
	<i>I can easily retrieve...</i>
blog	...
	...
	...
	...
	...
	...
social networks	...
	...

	...
	...
	...
	...

Appendix B

Bundle of questions concerning beliefs about deception in communication

environments

Channel	Preface	Question	
email	<i>In general, I believe that...</i>	<i>people lie a lot in email.</i>	
		<i>the lies people tell in email are mostly little white lies.</i>	
	In my experience, I believe that...	people lie to me a lot in email.	
		people tell me mostly little white lies in email.	
		I lie a lot in email.	
		I tell mostly little white lies in email.	
	IM	<i>In general, I believe that...</i>	<i>people lie a lot...</i>
			<i>the lies people tell...</i>
In my experience, I believe that...		people lie to me a lot...	
		people tell me mostly...	
		I lie a lot...	
		I tell mostly...	
blog	
	
	
	

		...
		...
		...
social networks
		...

		...
		...
		...
		...

Table 1

Spearman correlations for the first hypothesis examining experience and frequency of deception

	N	people lie a lot	people lie to me a lot	I lie a lot
I have a lot of experience using email	206	-0.083	-0.067	0.02
I have a lot of experience using instant messaging	206	-0.008	0.05	-0.009
I have a lot of experience using blogs	158	0.063	0.073	0.052
I have a lot of experience using online social networks	171	.234(**)	.237(**)	0.123
* Correlation is significant at the 0.05 level (2-tailed).				
** Correlation is significant at the 0.01 level (2-tailed).				

Table 2

Spearman correlations for the second hypothesis examining synchronicity and frequency of deception

	N	people lie a lot	people lie to me a lot	I lie a lot
I believe that email takes place in real time and feedback is immediate	206	-.158(*)	-.165(*)	-0.116
I believe that instant messaging takes place in real time and feedback is immediate	206	0.008	-0.093	-0.052
I believe that blogging takes place in real time and feedback is immediate	206	.171(*)	0.082	-0.008
I believe that online social networks take place in real time and feedback is immediate	206	-0.025	0.009	-0.008
* Correlation is significant at the 0.05 level (2-tailed).				
** Correlation is significant at the 0.01 level (2-tailed).				

Table 3

Paired T test examining the assumed synchronicity of the communication environments

	Mean	Std. Error Mean	t	Sig. (2- tailed)
I believe that email takes place in real time and feedback is immediate - I believe that instant messaging takes place in real time and feedback is immediate	-2.18	0.12	-18.13	$p < .001$
I believe that email takes place in real time and feedback is immediate - I believe that online social networks take place in real time and feedback is immediate	0.568	0.134	4.227	$p < .001$
I believe that blogging takes place in real time and feedback is immediate - I believe that online social networks take place in real time and feedback is immediate	-0.956	0.11	-8.714	$p < .001$

Table 4

Spearman correlations for the third hypothesis examining social distance and frequency of deception

	N	people lie a lot	people lie to me a lot	I lie a lot
I use email to communicate with people that I know personally	206	-0.028	-0.116	-0.012
I use instant messaging to communicate with people that I know personally	206	0.059	.165(*)	0.065
I use blogs to correspond with people I know personally	158	0.069	0.04	0.016
I use online social networks to communicate with people that I know personally	171	.176(*)	.210(**)	0.054
* Correlation is significant at the 0.05 level (2-tailed).				
** Correlation is significant at the 0.01 level (2-tailed).				

Table 5

Spearman correlations for the fourth hypothesis examining evanescence and magnitude of deception

	N	the lies people tell in email are mostly little white lies	people tell me mostly little white lies in email	I tell mostly little white lies in email
I can easily retrieve an old email	206	-0.063	-0.06	-0.054
I can easily retrieve an old instant message conversation	206	-0.059	-.151(*)	-0.072
I can easily retrieve an old blog post	158	0.031	0.053	.202(*)
I can easily retrieve old online social network information	171	0.09	.157(*)	0.078
* Correlation is significant at the 0.05 level (2-tailed).				
** Correlation is significant at the 0.01 level (2-tailed).				

Table 6

Paired T test examining the assumed evanescence of the communication environments

	Mean	Std. Error Mean	t	Sig. (2- tailed)
I can easily retrieve an old email - I can easily retrieve old online social network information	1.99	0.144	13.788	$p < .001$
I can easily retrieve an old blog post - I can easily retrieve old online social network information	0.602	0.173	3.48	0.001
I can easily retrieve an old instant message conversation - I can easily retrieve an old blog post	-1.549	0.192	-8.049	$p < .001$

Table 7

Paired T test examining the existence of the prevalence paradox in the dataset

	Mean	Std. Error Mean	t	Sig. (2- tailed)
people lie a lot in email - people lie to me a lot in email	.981	.097	10.102	$p < .001$
people lie a lot in instant messaging - people lie to me a lot in instant messaging	.840	.099	8.485	$p < .001$
people lie a lot in blog posts - I encounter many lies in blog posts	.228	.107	2.123	.035
people lie a lot in online social networks - I encounter many lies in online social networks	.444	.098	4.543	$p < .001$

Table 8

Simple regression coefficients holding the general frequency of lies as the dependent variable

	beta	t	sig.
people lie to me a lot in email	.259	3.931	$p < .001$
people lie to me a lot in instant messaging	.507	6.864	$p < .001$
I encounter many lies in blog posts	.586	8.475	$p < .001$
I encounter many lies in online social networks	.659	11.069	$p < .001$
Dependent Variable: people lie a lot in *.			

Table 9

Paired T test for equality of means with equal variances assumed in perceived frequency of lies in blogs and online social networks

	t	Sig. (2-tailed)	Mean Difference	Std. Error Difference
people lie a lot in blog posts	-0.209	0.835	-0.046	0.221
people lie a lot in online social networks	-1.862	0.064	-0.571	0.307