Argument Formation in the Reasoning Process: Toward a Generic Model of Deception Detection

Deqing Li and Eugene Santos, Jr.
Thayer School of Engineering
Dartmouth College
Hanover, N.H., U.S.A
{ Deqing.Li, Eugene.Santos.Jr }@Dartmouth.edu

Abstract

Research on deception detection has been mainly focused on two kinds of approaches. In one, people consider deception types and taxonomies, and use different counter strategies to detect and reverse deception. In the other, people search for verbal and non-verbal cues in the content of deceptive communication. However, general theories that study fundamental properties of deception which can be applied in computational models are still very rare. In this work, we propose a general model of deception detection guided by a fundamental principle in the formation of communicative deception. Experimental results using our model demonstrate that deception is distinguishable from unintentional misinformation.

Introduction

Conventional research on deception detection focuses on deception taxonomies and deception cues. Unfortunately, both of them neglect the fact that deception is rooted in the formation of arguments mainly because such formation is not directly observable. However, since the formation of arguments is where the implementation of deception starts, it is necessary to study it in depth.

The act of deceiving involves two processes: the formation of deceptive arguments (the reasoning) and the communication of deception. The communication part is intuitive to understand and has been the focus of recent research efforts in deception detection. The reasoning part is a necessary component of deception because deceiving has been found to require a heavier cognitive load than telling the truth (Greene et. Al, 1985). The reasoning process involves generating and selecting arguments while the communication process involves wording and phrasing of the arguments. Deception detection in the process of communication is not ideal because firstly, it is easy to hide deceptive cues using careful wording and phrasing, and secondly, wording and phrasing of communication are mediated by the framing of the other party's response (e.g. the answer to the question "Did you go to class today?" always starts with "Yes, I" or "No, I"). On the other hand, it is hard to hide the intent of deception by distorting arguments formed in the reasoning process because it requires higher-order deception that takes the other party's intent and even the other party's belief about the speaker's intent into consideration. Higher-order deception demands much more cognitive load than first-order deception in order to retrieve the memory about the other party's intent and leverage the original reasoning process behind it. Thus, the reasoning process provides more effective and reliable observations than the communication process. Moreover, it also guides and explains some observations in the communication process such as compellingness and level of detail of a story.

We will illustrate the formation of deceptive arguments in the next section, according to which, we propose three hypotheses of the fundamental differences between deception and non-deception. In Section 3, we describe our model of detection and the data simulation process. Experiment setting and results are

Proceedings of the EACL 2012 Workshop on Computational Approaches to Deception Detection, pages 63–71,
Avignon, France, April 23 - 27 2012.

discusses in Section 4, followed by conclusions and future work in Section 5.

1 Formation of Deceptive Argument

The reasoning process can be regarded as inference based on the conditional relationship between arguments by assuming that human reasoning is akin to informal logic. Since deceivers intentionally reach the conclusion that they target at, we propose that the act of deceiving is to reason by supposing the truth of deceivers' targeted arguments, but the truth of the targeted arguments is not actually believed by the deceivers. For example, if a person is asked to lie about his attitude on abortion, he might raise arguments such as "fetuses are human", "god will punish anyone who aborts children" and "children have the right to live". He did not raise these arguments because he believed in them but because they support the false conclusion that he is against abortion. It is thus natural to imagine that the conclusion comes into deceivers' minds before the arguments. According to Levi (1996), "*The addition of the supposition to the agent's state of full belief does not require jettisoning any convictions already fully believed. The result of this modification of the state of full belief by supposition is a new potential state of full belief containing the consequences of the supposition added and the initial state of full belief*", which means that the reasoning with a supposition is a regular reasoning with the addition of a piece of knowledge that has been assumed before the reasoning starts. It also follows that the reasoning with a supposition can be exactly the same as a regular reasoning in which the supposition in the former case is a true belief. That is to say, the reasoning in deception formation can be regarded to follow the same scheme as that in truth argumentation. However, even if deceiver and truth teller share the same reasoning scheme, their beliefs and processes of reasoning are different. In particular, if an opinion-based story is required from the speaker, truth tellers propagate beliefs from evidence, while deceivers adapt beliefs to suppositions. If an event-based story is required, truth tellers retrieve relevant memory which is based on past behavior and past behavior is based on past belief, which was propagated from past evidence, while deceivers suppose a part of the event and adapt his fantasy to the supposition. This fundamental difference in the reasoning of deceiver and truth teller is unavoidable due to the intentionality of deceivers. It provides reasoning a stable ground on which schemes of deception detection can be built.

As we have discussed, the product of reasoning from truth teller and deceiver may be exactly the same. However it is hardly true in the real world because they do not share the same belief system that supports their reasoning. If in any case they do share the same belief system, they would reach the same conclusion without any deception and there would be no need to deceive. In order to mimic truth teller's story, deceiver may manipulate his conclusion and distort other arguments to support the manipulated conclusion, but the supporting arguments are biased by his honest but untruthful belief system. **Therefore, discrepancies in arguments that deceivers are reluctant to believe but truth tellers embrace can be expected.** On the other hand, deception has been defined as "*a relationship between two statements*" (Shibles, 1988), according to which, deception is a contradiction between belief and expression. A deceiver may lie about the polarity of belief as well as the strength or extent of belief as long as his belief expression deviates from his honest reasoning. The more manipulation he did to mimic the truth, the farther he deviates from himself. **Therefore, discrepancies in arguments that are manipulated by deceivers can be expected.** The above two discrepancies in deception have been popularly embraced by existing researchers (Mehrabian, 1972; Wiener & Mehrabian, 1968; Johnson & Raye, 1981, Markus, 1977). Our focus is to explain and measure them in terms of human reasoning, and argue that these two discrepancies follow our proposal that deceptive reasoning is reasoning with presupposition, due to which the discrepancies are the fundamental difference between deception and truth that produces other observable patterns.

2 Hypotheses and Justification

We have argued that the basic discrepancy in deceptive reasoning exists in inconsistency and untruthfulness. Inconsistency means that the arguments in the story contradict with what the speaker would believe. Untruthfulness means that the arguments in the story contradict with what an honest person would believe in order to reach the conclusion. On the other hand, inconsistency indicates that an honest person

should behave as he always does, which requires some familiarity with the speaker, whereas untruthfulness indicates that an honest person should behave as a reasonable and convincing person, which requires some knowledge of the topic domain. Opinion change violates the former one but not the latter one as it changes the prior knowledge but still maintains truthfulness, and innovation violates the latter one but not the former one as innovation is convincing but not expectable. They do not violate both so they are not deceptions. However, these two elements are not the unique characteristics of deception because random manipulations without any purpose to deceive such as misinformation also show inconsistency and untruthfulness. Fortunately, deceivers can be distinguished by the manner they manipulate arguments. We propose the following hypotheses that can be expected in deceptive stories but not others.

Firstly, explicit manipulations in deception continuously propagate to other arguments which become implicit manipulations. The purpose, of course, is to spread the manipulation to the conclusion. The propagation spreads to surrounding arguments and the influence of manipulation decreases as the propagation spreads farther away, which random manipulations do not exhibit. If one overlooks the abnormality of the explicit manipulations, the story would seem to flow smoothly from the arguments to the conclusion because the connection between the arguments is not broken. Inconsistency is particularly important when individual difference should be considered.

Secondly, there is a correspondence between inconsistency and untruthfulness. Some inconsistencies were manipulated significantly because the deceiver wants to convince the listener of the argument and these arguments seem more reasonable to support the conclusion after manipulation. Therefore, the significant manipulations are often convincing, but there are also exceptions in which deceivers overly manipulate arguments that are usually ignored by truth tellers. We call these Type I incredibility: incredibility due to over-manipulation. The arguments that are not convincing usually can be found in the inconsistencies that were slightly manipulated or ignored by the deceiver because deceivers do not know that they are important supports to the conclusion but truth tellers never neglect these details. This is called Type II incredibility: incredibility due to ignorance. Type I and Type II incredibility are two examples of unconvincing arguments (According to DePaulo et. al (2003), liars tell less compelling tales than truth tellers), which can be quantitatively measured in the reasoning process. On the other hand, random manipulations do not show this correspondence between inconsistency and untruthfulness. Measuring untruthfulness is particularly effective in detecting deception from general population whom the detector is not familiar with.

Thirdly, deceptions are intentional, which means the deceiver assumes the conclusion before inferring the whole story. Or in other words, deceivers fit the world to their mind, which is a necessary component of intentionality according to Humberstone (1992). They are convincers who reach arguments from conclusions, while others reach conclusions from arguments. According to the satisfaction of intention (Mele, 1992), an intention is "satisfied" only if behavior in which it issues is guided by the intention-embedded plan. Thus, deceivers choose the best behavior (argument in this case) that is guided (inferred in this case) by his desire (conclusion in this case), but not any behavior that can fulfill his desire. In particular, deceivers will choose the state of the argument in the story that is most effective compared with other states of the argument in reaching the conclusion of the story (e.g. the best state of whether ‘an unborn baby is a life’ towards the conclusion of supporting abortion is no). In deception, the inconsistent arguments are usually effective to the conclusion, while in random manipulation the inconsistent arguments are not.

Inconsistency, untruthfulness, propagated manipulation and intentionality are the guiding concepts of our deception detection method, which is a general model independent of the domain knowledge.

3 Methodology

In this work, we will not only test the hypotheses proposed above, but also provide a computational model to identify the discrepancy in arguments that are manipulated by deceivers and the discrepancy in arguments that are not as convincing as truth tellers’.

3.1 Computational Model of Deception Detection

We propose a generic model to detect deception through the reasoning process without assuming human’s reasoning scheme. As shown in Figure

1, the model is composed of two networks: Correlation Network and Consensus Network. Correlation Network connects each agent with agents who correlate with him in a specific argument. Neighbors in the Correlation Network represent acquaintances who can anticipate each other's arguments. Consensus Network connects agents with similar conclusions. Neighbors in the Consensus Network represent people who agree with each other. We have pointed out that deception is deviation from one's own subjective beliefs, but not deviation from the objective reality or from the public. Thus Correlation Network is essential in predicting an agent's belief according to neighbors who can expect each other. This idea of measuring individual inconsistency has been discussed in our former work (Santos et. Al, 2010), which also provides details on the computation. The Consensus Network provides a sampled population of truth tellers who reach the same conclusion as the deceiver. If the deceiver told the truth, he should behave in no difference with the population. The untruthfulness of the deceiver can be evaluated by comparing the deceiver with the truth tellers. Functionality of the arguments can be revealed from the history data of the deceiver. By studying the history data, we can evaluate which arguments are effective to which from the perspective of the deceiver.

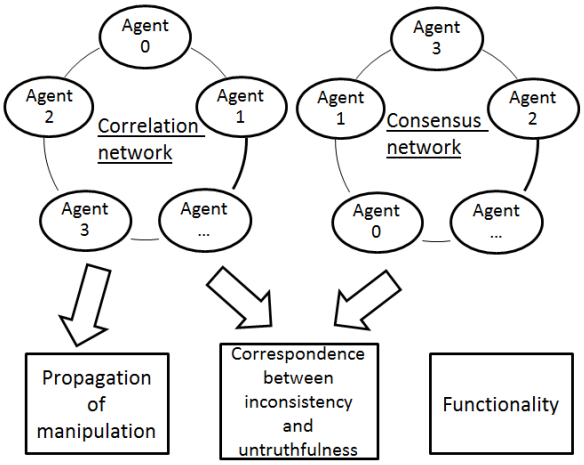

Figure 1: Architecture of the model of deception detection

3.2 Date Collection and Simulation

To test the hypotheses we proposed, we simulate the reasoning process of a deceiver according to our assumption that deceivers pre-suppose conclusions before reasoning. The deceiver we simulate is a plaintiff in a lawsuit of a rape case shown in a popular Hong Kong TV episode. The case is described as following. A female celebrity coded as *A* claims that she was raped by an Indian young man coded as *B*. *A* claims that she keeps away from *B* because both her and her mother do not like the physical odor of Indians. *A* claims that *B* once joined her birthday party without any invitation and fed *A* drugs. *B* then conveyed *A* home and raped *A*. After *A*'s boyfriend arrived, *A* called police. However, the truth is that *B* is a fan of *A* and joined *A*'s party at *A*'s invitation. *A* lied about her aversion to Indians because she used to prostitute to Indians. Besides, *B* is new to the party club, so it is unlikely for him to obtain drugs there. *A* used drugs and enticed *B* to have sex with her. This artificial scenario is a simplification of a possible legal case, which provides realistic explanations compared with simulation data that simulate deception arbitrarily without considering the intent of deceiver. We did not use real cases or lab surveys because they either do not have the ground truth of the speaker's truthfulness or lack sufficient information about the reasoning of the deceiver. Data that do have both ground truth and sufficient information such as military combat scenarios are mostly focused on behavioral deception instead of communicative deception. In addition, real cases may contain noisy data in which the communication content is mediated by factors other than reasoning. For the purpose of evaluating hypotheses about deceptive reasoning it is ideal to use clean data that only contains the semantic meaning of arguments. The evaluation of the hypotheses guides the development of our detection model, which we will apply to real data eventually.

A's belief system is represented by a Bayesian Network (BN) (Pearl, 1988). BNs have been used to simulate human reasoning processes for various purposes and have been shown to be consistent with the behavior of human (Tenenbau et. Al, 2006). A BN is a graphical structure in which a node represents a propositional argument and the conditional probability between nodes represent the conditional relationship between arguments. For example, the reasoning that *B* drives *A* home because *B* knows *A*'s address can be encoded in the conditional probability P(*B_drive_A_home*|*B_know_A_s_adr*)=0.9. In order to eliminate the variation due to wording, the semantics of the arguments instead of the phrases are encoded in the nodes. We designed a BN representing *A*'s belief system and also a BN

representing the belief system of a true victim of the rape case according to the description of the scenario and some common sense. More specifically, we connect two arguments if their causal relationship is explicitly described by the deceiver or by the jury when they are analyzing the intent of the deceiver. The conditional probabilities between states of arguments are set as 0.7 to 0.99 according to the certainty of the speaker if they are explicitly described. As to the states that are not mentioned in the case, they are usually implied in or can be inferred from the scenario if their mutual exclusive states are described in the scenario, such as the probability of *A_hate_Indian* given that *B*'s relation with *A*'s mother is good and that *A* used to prostitute to Indians. Otherwise the mutual exclusive states are given the same or similar probabilities indicating that they are uncertain. To make sure that the discrepancies in deception are resulted from the manner of reasoning instead of from the inherent difference between the deceiver's belief system and the true victim's belief system, we minimize the difference between their belief systems. Specifically, we keep all their conditional probabilities the same by assuming that both are rational people with the same common sense. Only their prior probabilities of *A*'s experience as prostitute and whether *B* is new to the party or not are adjusted differently, because they are the essential truth in a true victim's perspective. That is to say, those who do not like Indians could not prostitute to them, and to obtain drugs from the party club, B has to be a regular guest. However, as a result of sharing a similar belief system with the true victim, the deceiver's story may become highly convincing. Although we expect it to be hard to detect the untruthfulness of the deceiver, the deceiver's simulation is not unrealistic because some deceivers are consistently found to be more credible than others based on the research by Bond and Depaulo (2008). It is highly likely that a randomized BN with a perturbed copy can also serve our purposes, but again, building belief systems based on the intent of deception will provide more realistic data, more convincing results and more intuitive explanations. The BN of the deceiver is depicted in Figure 2. Its conditional probability tables are shown in Appendix A.

The process of reasoning is represented by the process of inferencing, and the product of reasoning is represented by the inferred probabilities of the nodes. Computing posterior probabilities, P(A|E), is not feasible here since it does not consider the consistency over all variables. Consider the following example. Suppose 10 people join a lottery of which exactly one will win. By computing posterior probabilities, we obtain the result that no one will win because each of them wins with probability 0.1. To retain the validity of the probability of each variable as well as the consistency over all variables, we propose the following inference. We first perform a belief revision and obtain the most probable world, which is the complete inference with the highest joint probability. Then for each variable, we compute its posterior probability given that all other variables are set as evidence with the same assignment as in the most probable world. By inferring the lottery example in this way, in each of its inferred world a different person wins with equal probability. Specifically, the probability of a person winning given all others not winning is 1, and the probability of a person winning given all but one winning is 0. As we proposed earlier, the reasoning process of the deceiver presupposes her target arguments, that is, she was raped, by adding the argument as an extra piece of evidence. The inference results of *A* in both deceptive and honest cases and those of a true victim are shown in Table 1. The arguments *B_relation_with_A_s_mother=bad*, *B_drive_A_home=true*, *A_is_celebrity=true* and *A_s_boyfriend_catch_on_the_scene=true* are set as evidence as suggested by the scenario.

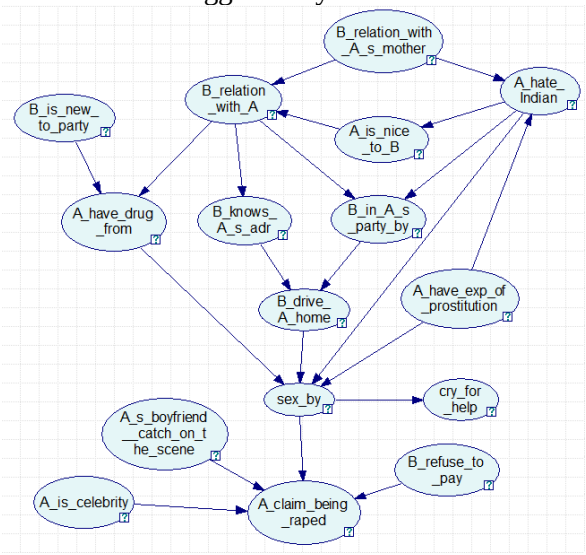

Figure 2: BN of the deceiver in the rape case

People express attitudes as binary beliefs in communication if not as beliefs with fuzzy confidence, but not as degree of belief

formulated by real-valued probabilities. To map degree of belief to binary beliefs, we need to know how much confidence is sufficient for a person to believe in an attitude. Or in other words, what is the probability threshold of something being true. Research has suggested that truth threshold varies by proposition and by individual, which means it is a subjective criterion (Ferreira, 2004). Since we use simulated data, we arbitrarily choose 0.66 as the threshold since it equally spaces the interval of an argument being true, unknown and false. Then the binary beliefs in the deceptive story and honest story of the deceiver and those in the true victim's story would be the same as Table 2. To verify the inferred beliefs, we compare Table 2 with the scenario. An argument is validated if it is in the same state as described in the scenario or in the unknown state given that it is ignored in the scenario. We verified that 13 out of the 16 arguments in the deceptive story corresponds with what the deceiver claims, all of the arguments in the honest story corresponds with what is the truth. Although it is hard to verify the true victim's story because we do not have its ground truth, we observe that all the arguments are reasonable and most are contrary to the deceiver's honest story except the evidence.

Arguments	**Decept.**	**Honest**	**True**
B_relation_with_As_mother=good	0	0	0
A_have_exp_of_prostitution=T	0.66	0.88	0.11
A_hate_Indian=T	0.74	0.07	0.89
A_is_nice_to_B=T	0.18	0.88	0.18
B_relation_with_A=rape	0.98	0.16	0.96
B_in_A_s_party_by=self	0.9	0.4	0.90
B_knows_A_s_adr=T	0.95	0.95	0.95
B_drive_A_home=T	1	1	1
B_is_new_to_party=T	0.76	0.82	0.16
A_have_drug_from=B	0.76	0.07	0.92
sex_by=rape	0.93	0.08	0.98
As_boyfriend_catch_on_the_scene=T	1	1	1
A_is_celebrity=T	1	1	1
B_refuse_to_pay=T	0.8	0.85	0.50
A_claim_being_raped=T	0.6	0.7	0.60
cry_for_help=T	0.8	0.2	0.80

Table 1: Inferred results of the deceiver's deceptive story, her honest story and a true victim's story

The computation of the discrepancies assumes acquaintance of the deceiver, which requires sufficient number of history data and neighbors of the deceiver. To achieve it, we simulate 19 agents by perturbing the deceiver's BN and another 10 agents by perturbing the true victim's BN. In total, we have 29 truth telling agents and 1 deceiving agent. We simulate 100 runs of training data by inferring the network of each agent 100 times with different evidence at each run, and convert them to binary beliefs. Training data is assumed to contain no deception. This approach of inconsistency detection is borrowed from our past work (Santos et. Al, 2010).

Arguments	**Decept.**	**Honest**	**True**
B_relation_with_As_mother	bad	bad	bad
A_have_exp_of_prostitution	unknn	T	F
A_hate_Indian	T	F	T
A_is_nice_to_B	F	T	F
B_relation_with_A	rape	fan	rape
B_in_A_s_party_by	self	unknn	self
B_knows_A_s_adr	T	T	T
B_drive_A_home	T	T	T
B_is_new_to_party	T	T	F
A_have_drug_from	B	self	B
sex_by	rape	entice	rape
As_boyfriend_catch_on_the_scene	T	T	T
A_is_celebrity	T	T	T
B_refuse_to_pay	T	T	unknn
A_claim_being_raped	unknn	T	unknn
cry_for_help	T	F	T

Table 2: Binary beliefs of the deceiver's deceptive story, honest story and a true victim's story

4 Experiment and results

To test the hypotheses, we compare the result of deceptive story with the result of misinformative story. A misinformative story is simulated by adding random error to the inferred results of the arguments.

- Propagation of manipulation

To calculate inconsistency we predict binary beliefs in the deceptive story using GroupLens (Resnick et. Al, 1994) based on stories of neighboring agents in the Correlation Network. We then compare the binary beliefs in the deceptive story with predicted binary beliefs to measure deviation of each argument due to inconsistency. We measure how many standard (std.) deviations the prediction error in deceptive story deviates from the prediction error in training data, and plot them according to their locations in the BN, as shown in Figure 3. The

width of the links represents the sensitivity of each variable to its neighbors.

We observe that the variables at the boundaries of the graph and not sensitive to neighbors (e.g. *B_is_new_to_party*) are ignored by the deceiver, while the variables in the center or sensitive to others (e.g. *A_hate_Indian*) are manipulated significantly. It demonstrates that manipulations propagate to closely related arguments. Unrelated arguments are probably considered as irrelevant or simply be ignored by the deceiver. On the other hand, if we compare deceptive story with honest story in Table 2, we obtain 9 arguments manipulated by the deceiver. Out of these 9 arguments, 8 are successfully identified as inconsistent by Figure 3 if we assume the suspicion threshold is 3 std. deviations.

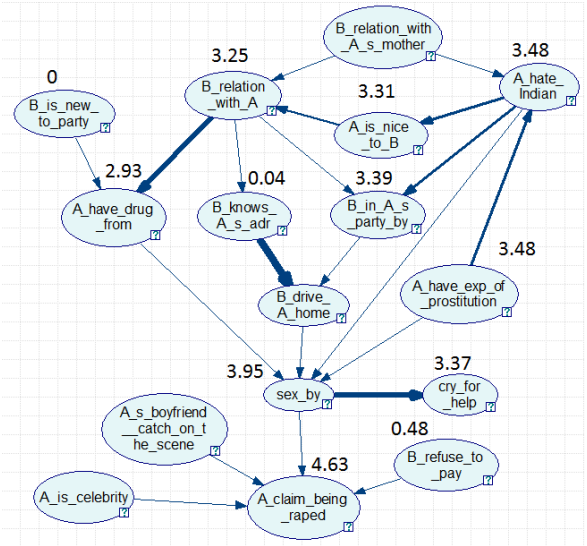

Figure 3: Inconsistency deviation of each variable

- Correspondence between inconsistency and untruthfulness

To compute untruthfulness, we calculate the deviation of the binary beliefs in the deceptive story from the population of truth teller's stories who agrees with the deceiver in the Consensus Network. We then compare the deviation due to inconsistency with respect to the deceiver herself and that due to untruthfulness with respect to truth tellers. The result is shown in Table 3.

The correlation between the deviation due to inconsistency and that due to untruthfulness is -0.5186, which means that untruthfulness has a large negative correlation with inconsistency. It credits our hypothesis that significant manipulations are often convincing and unconvincing arguments usually can be found in slightly manipulated or ignored arguments. The only exception in the result is the argument *B_knows_As_address*, which is not manipulated but convincing. It is probably because the evidence *B_drive_A_home* enforced it to remain honest. Type I incredibility does not occur in this case, but type II incredibility appears in the argument *B_is_new_to_party* and *B_refuse_to_pay*. The deceiver ignored these arguments, which results in the incredibility of the story. The correlation between inconsistency and untruthfulness in misinformative stories ranges between 0.3128 and 0.9823, which demonstrates that the negative correction cannot be found in misinformative stories. If we compare the deceptive story and the true story in Table 2, we find out that 3 arguments in the deceptive story are unconvincing. By observing the untruthfulness in Table 3, we find out that 2 of the 3 arguments are out of at least 1.44 std. deviations of the sample of true stories and all of them are out of at least 0.95 std. deviations. The small deviations indicate a high credibility of the deceiver, which is caused by the similarity between the belief systems of the deceiver and the true victim.

Belief	**Incon.**	**Untru.**
B_relation_with_As_mother=good	N/A	N/A
A_have_exp_of_prostitution=T	3.48	0.95
A_hate_Indian=T	3.48	0.28
A_is_nice_to_B=T	3.31	0.28
B_relation_with_A=rape	3.25	0
B_in_A_s_party_by=self	3.39	0.28
B_knows_A_s_adr=T	0.04	0
B_drive_A_home=T	N/A	N/A
B_is_new_to_party=T	0	1.59
A_have_drug_from=B	2.93	0
sex_by=rape	3.95	0
As_boyfriend_catch_on_the_scene =T	N/A	N/A
A_is_celebrity=T	N/A	N/A
B_refuse_to_pay=T	0.48	1.44
A_claim_being_raped=T	4.63	0.41
cry_for_help=T	3.37	0.41

Table 3: Comparison of inconsistency and untruthfulness of the deceiver

- Functionality

Functionality means that the manipulated arguments are effective in reaching the goal and at the same time satisfies the evidence. In other words, we can expect the manipulated arguments from the goal and the evidence. The calculation

of functionality is as following. For each inconsistent argument, we measure its correlation with other arguments in the past using training data. We then predict each argument's binary belief based on the value of the conclusion and the evidence. If the predicted belief corresponds with the belief in the deceptive story, the variable is functional. We compare the results of deceptive story with those of misinformative story. In Table 4, all but one manipulated arguments in the deceptive story complies with the value expected by the conclusion and evidence, but none of the inconsistent arguments in misinformative stories does. Although the result shown in Table 5 comes from a random sample of misinformative story, we observed that most of the samples show the same functionality rate. Therefore, the functionality rate of deceptive story is 6/7, while the functionality rate of misinformative story is around 0/3.

Arguments	Pred.	Decept.
A_have_exp_of_prostitution=T	0.24	0.5
A_hate_Indian=T	0.85	1
A_is_nice_to_B=T	0.07	0
B_relation_with_A=rape	0.99	1
B_in_A_s_party_by=self	1	1
A_claim_being_raped=T	0.58	0.5
cry_for_help=T	0.86	1

Table 4: Functionality of the deceiver's story

Arguments	Pred.	Misinfo.
B_in_A_s_party_by=self	0.45	0
B_knows_A_s_adr=T	0.90	0.5
A_claim_being_raped=T	0.94	0.5

Table 5: Functionality of a mininformative story

5 Conclusion and future work

We proposed in this work two fundamental discrepancies in deceptive communications: discrepancies in arguments that deceivers are reluctant to believe but truth tellers embrace and discrepancies in arguments that are manipulated by deceivers. The proposal follows the following three assumptions: The act of deceiving is composed of deceptive argument formation and argument communication; Deception is formed in the reasoning process rather than the communication process; Reasoning is interaction between arguments, and deceptive reasoning is reasoning with presupposition. Then we proposed three hypotheses in order to distinguish deception from unintentional misinformation: manipulations propagate smoothly through closely related arguments, inconsistency and untruthfulness are negatively correlated, and deceptive arguments are usually functional to deceiver's goal and evidence. To evaluate and to measure these hypotheses from communication content, we designed a generic model of deception detection. In the model, agents are correlated with others to expect each other's consistency in beliefs and consenting agents are compared with each other to evaluate the truthfulness of beliefs. Our experimental results credit the hypotheses. The main contribution of this work is not to follow or reject the path that linguistic cues have laid out, but to suggest a new direction in which deeper information about the intent of deceivers is carefully mined and analyzed based on their cognitive process.

In the future, we will further develop the model by designing and implementing detection methods based on the hypotheses. Currently we use simulated data based on an artificial story, which is closer to a real legal case that provides concrete information about the reasoning of deceivers with minimum noise. In the future, we will apply the model to survey data that is commonly used in the area. Various natural language processing techniques can be utilized in the retrieval of the reasoning process. Specifically, Latent dirichlet allocation (Blei et. Al, 2002) can be used to categorize the sentences into topics (or arguments), sentiment analysis (Liu. 2010) can be used to extract the polarity of each argument, and various BN constructors such as PC algorithm (Spirtes et. Al, 1993) can be used to construct the belief systems. On the other hand, linguistic cues have been observed in past research (DePaulo et. al, 2003), but has not been defined or explained quantitatively. The study of the pattern of deceptive reasoning can ultimately provide guidance and explanations to existing observations in deception cueing.

Acknowledgments

This work was supported in part by grants from AFOSR, ONR, and DHS.

References

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet allocation. Journal of Machine Learning Research, Lafferty, John. Ed., 3 (4–5): 993–1022.

Bella M. DePaulo, James J. Lindsay, Brian E. Malone, Laura Muhlenbruck, Kelly Charlton, and

Harris Cooper. 2003. Cues to deception. Psychological Bulletin, 129(1): 74-118.

Ulisses Ferreira. 2004. On the Foundations of Computing Science. Lecture Notes in Computer Science, 3002:46-65.

John O. Greene, H. Dan O'hair, Micheal J. Cody, and Catherine Yen. 1985. Planning and Control of Behavior during Deception. Human Communication Research, 11:335-64.

I. L. Humberstone. 1992. Direction of Fit. Mind, 101(401): 59-84.

Marcia K. Johnson and Carol L. Raye. 1981. Reality Monitoring. Psychological Bulletin, 88:67–85.

Isaac Levi. 1996. For the Sake of the Argument. Cambridge University Press. New York, NY, USA.

Bing Liu. 2010. Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing Issue, 1st ed., Taylor and Francis Group, Eds. CRC Press, 1-38.

Hazel Markus. 1977. Self-schemata and Processing Information about the Self. Journal of Personality and Social Psychology, 35:63–78.

Albert Mehrabian. 1972. Nonverbal Communication. Aldine Atherton, Chicago, USA.

Alfred R. Mele. 1992. Springs of Action: Understanding Intentional Behavior. Oxford University Press. New York, NY, USA.

Judea Pearl. 1988. Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann Publishers, San Francisco, CA, USA.

Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. 1994. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. Proc. of the Conference on Computer Supported Cooperative Work, 175-186. ACM Press, Chapel Hill, NC, USA.

Eugene Santos, Jr. and Deqing Li. 2010. Deception Detection in Multi-Agent Systems. IEEE Transactions on Systems, Man, and Cybernetics: Part A, 40(2):224-235.

Warren Shibles. 1988. A Revision of the Definition of Lying as an Untruth Told with Intent to Deceive. Argumentation, 2:99-115.

Peter Spirtes, Clark N. Glymour, and Richard Scheines, 1993. Causation, Prediction, and Search. Springer-Verlag, New York, NY, USA.

Morton Wiener and Albert Mehrabian. 1968. Language within Language: Immediacy, a Channel in Verbal Communication. Meredith Corporation, New York, NY, USA.