The Use of Second Life for Deception Detection Research

Stephen Kunath, Kevin McCabe
Center for the Study of Neuroeconomics
George Mason University
4400 University Drive, MSN 1B2
Fairfax. VA 22030, USA
skunath@gmu.edu, kmccabe@gmu.edu

Abstract

Detecting deception in natural language is a problem amenable to economic analysis. Economics typically assumes that individuals are self-interested, which leads them to perform actions in accord with their own goals. The field of experimental economics emerged to construct environments wherein human subjects make decisions so as to test economic hypotheses. Experimental economists recently have developed virtual worlds to better situate experiment subjects in more realistic environments. Virtual word experiments represent an exciting new area for deception research as they offer insight into individuals both acting out and communicating in accord with their intentions. This paper describes the use of virtual world experiments for economic research incorporating the detection of deceptive individuals.

1 Introduction

The fields of linguistics and economics typically assume different analytical techniques. Most peoples experience in an undergraduate economics class leaves them with the notion that the subject of economics contains only graphs. People then reasonably conclude that graphical depictions of supply and demand have little, if any, relation to speech acts. A result of the presumed limited applicability of economic modeling is that many do not realize that the graphs suggest how real life exchange actually operates. At the same time, as anyone should know, real life exchange requires some amount of communication between buyer and seller. Communication can take many forms from grunts and clicks, to the use of a human language, to an automated exchange of data between information systems. Fundamentally though human language will always play a necessary role in facilitating exchange between groups or individuals either directly or by helping different parties to negotiate the actual rules of exchange themself.

Suggesting that language and communication play some role in person to person exchange is correct but insufficiently descriptive. Language certainly plays a significant role in exchange as to effect a meeting of the minds between individuals there must be some common agreement concerning what is being exchanged and why. Potential obstacles can emerge during the communication phase of any round of negotiating exchange. An initial problem relates to the communicative sophistication of the parties involved in an exchange. One speaker could be a lawyer with an extensive background in negotiation while the other party might be a well educated layperson without specialized training. Differences between the individuals' verbal facility can lead to confusion or a wariness to engage in exchange for fear of manipulation. Another problem arises when two parties have different knowledge of things in the real world understood as an asymmetry of information between individuals. For example, a baseball card collector might know what a particular card is worth and believe the person he is trading with is unsure about the actual value and so the collector might offer a lower value than the market would suggest. Individuals possessing better information can then

Proceedings of NAACL-HLT 2016, pages 32–39,
San Diego, California, June 12-17, 2016. ©2016 Association for Computational Linguistics

reap an economic reward by using private information to their advantage.

The examples of problems found in exchange described above relate to communicative facility and information symmetry. When combined together one can reasonably expect an advantage to more well informed and verbally proficient exchange partners. Throughout history individuals have attempted to reduce the frequency of abuse in exchange by standardizing contract language and practices to reduce the advantages provided by enhanced verbal facility. To reduce information asymmetry consumer review services like CarFax collect and provide information to individuals about the actual history of a car, for example. Information asymmetry and superior communicative abilities, however, remain potential sources of discrepancies between individuals when the subject of exchange is an infrequently traded commodity. Unscrupulous parties can make use of the presence of asymmetries and verbal proficiency differences to deceive other people in an exchange.

2 Background

The ability to detect when other persons are being deceptive has always been a highly valued skill. Anecdotes abound in literature and personal histories of moments when one person grows suspect that another party was being deceptive. Detecting deception remains a perpetual challenge though, as it results from the fact that at any given time, one cannot immediately verify a claim that an another individual might make. Researchers have made progress in developing techniques to detect deception by studying what deception looks like. Identifying the characteristics of deceptive communication involves looking at numerous examples of dialogue and selecting those statements that contain deceptions. The modern study of deception has adopted a data driven approach which has resulted in many insights, but at the same time it has left open the question of how to obtain useful example data for research purposes. Different fields investigate deception using wholly different techniques aiming towards different goals. Some might be more interested in the contexts where deceptions can occur, while others might have more interest in the language differences between truthful and deceptive communications. One possible obstacle to progress in understanding and detecting deception could be the sheer number of different fields investigating the question such as psychology, computational linguistics, sociolinguistics, and economics. Each field uses its own data, methods, and models for the problem and a result has been less of an interdisciplinary approach than might otherwise be hoped for.

Much of the deception detection research investigates techniques for constructing datasets containing useful examples of deceptive communication. An obvious challenge constructing any deception dataset involves correctly labelling instances of deception and truth telling. Researchers have previously constructed datasets from online resources (Gokhman et al., 2012) as well as having human subjects intentionally engage in deceptive behavior (Almela et al., 2012). A necessary first step in the development of models of deceptive communication involves collecting and constructing appropriate corpora. It is crucial then that a deception corpus including instances of deceptive communication must also incorporate information concerning the actual behavior and goals of individuals as made manifest in their communication and also in their actions. That is, an ideal deception corpus would include both what a person said along with their motivation for saying it and whether what they said was true. Unfortunately, we cannot have complete knowledge of the current state of an individual's perspective on the world and as such we will never be capable of constructing an ideal corpus.

Absent an ideal corpus, insights from the field of economics present some interesting possibilities on the study of deception. For example, many people find it perfectly reasonable to assume that deception is a common part of many everyday transactions. Since economics assumes people are self-interested, one possible extreme assumption suggests people will be deceptive whenever they feel it serves their self-interest. Experience, however, suggests that people are not deceptive at all times. Anecdotal evidence instead suggests that individuals are more deceptive when the consequences of their deception are not severe or the probability of being caught is low. In point of fact, some economists have broken up deceptive behavior along the consequences that it might produce (Gneezy, 2005). The key distinction between different types of lies, as Gneezy suggests, involves understanding who might benefit from the deception. Deception then can be understood as benefiting or harming no one, the deceiver, the deceived, or some combination thereof. Here the economic conclusion of a possible economic outcome to the deceiver can then drive the communicative behavior to facilitate achieving the benefit. Seeing that there is an economic consequence to deception leads to the possibility that further research into deception could benefit by integrating insights from economics. Experimental economists have attempted to model institution formation in a number of different types of experiments that could be suggest a paradigm for further deception research (Kosfeld et al., 2009).

3 Experimental Environments

In recent years experimental economists have developed ever more sophisticated environments to test their models. One area that is particularly interesting for its potential with helping with research in deception involves the construction and use of virtual world experiments. One recent study constructed a virtual world to investigate the question of whether the discovery of natural resources can negatively impact a country (Al-Ubaydli et al., 2014). In their paper Al-Ubaydli et al. built an environment for their experiment subjects using Second Life. The factors the researchers varied were the presence of a resource in a particular country and how the subjects were able to communicate. While the economic question raised by the paper is interesting in its own right, the concern here is the part of the experiment that relates to controlling the subjects' ability to communicate. Al-Ubaydli et al. developed their model with a particular emphasis on controlling the ability of individuals to communicate based on the notion that communication is an essential component of institutional development and economic exchange.

Here it is worth considering that the term 'institution' carries a special meaning in economics. Generally it is meant to indicate the formal and informal rules of a society including property rights and dispute resolution (North, 1991). Economists suggest that for a society or group of individuals to succeed they must construct and maintain formal and informal institutions. If no one, for example, enforces property rights, then one would expect that people would alter their behavior to avoid issues arising from property questions like theft. Economists agree that good institutions are essential for economic flourishing, but then the next question is how to construct effective institutions.

A reasonable first guess is that good institutions involving humans do not appear randomly. Good institutions necessarily require trust between the different members of a society as manifested in cooperation and coordination. But trust, too, does not emerge spontaneously. Ostrom investigated the ingredients and steps for successful institutions to emerge. In Ostrom's view there must be both communication channels between individuals and some way to monitor compliance with rules (Ostrom, 2000). Communication is an essential aspect of monitoring since humans lack omniscience and cannot therefore know when someone is not living up to their side of an agreement. Unable to monitor all events, individuals must construct contracts whereby agreements are formed that require performances by each party and penalties for nonperformance. In this agreement and performance component, the possibility remains that either party can deceive at any stage. That is, deception can occur when a party willingly agrees to something they never intend to do. Individuals can agree to perform their duties and then not do them, but they might realize that absent effective monitoring they can simply report that they completed their tasks. Deception can therefore occur during the negotiating involved in constructing an agreement or it can happen after an agreement takes place where an individual deceives as to what they have done.

The ability to verify that an individual is telling the truth constrains the deceptive behavior of individuals especially in the presence of negative consequences for deceiving. In the real world it is difficult for individuals to monitor everything related to a particular agreement, but virtual worlds offer the opportunity to monitor the actions of individuals and compare them to whatever claims they make in their communication. With this in mind, some of the tools from experimental economics seem particularly well suited to the problem of deception research as experimental economists can construct environments and incentives for subjects to deceive to enhance their own benefits. Virtual world experiments can be used for experiments that also allow rich communication between players with the added benefit of being able to monitor all player actions. Thereby the data produced by virtual world experiments provide both the actual messages communicated by individuals and also the insight into what actions an individual actually took. Appropriately constructed virtual world experiments then can offer researchers an incredible new tool to peer into deceptive behaviors and their manifestation in communication.

4 Example Experiment

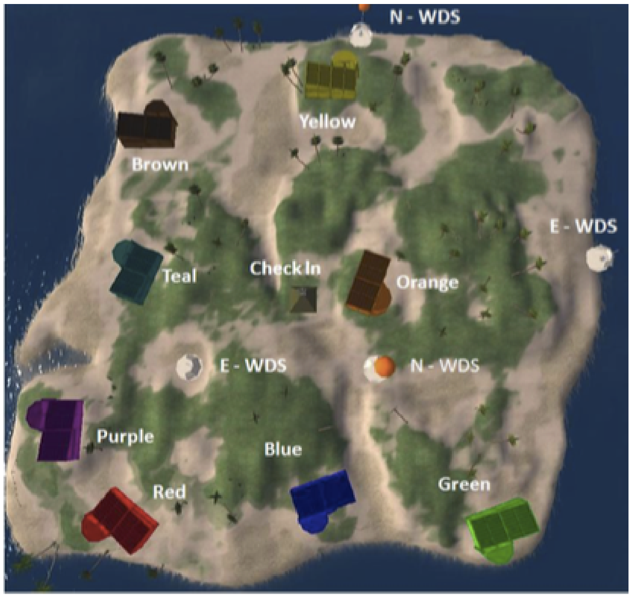

Figure 1: An overhead image of the Second Life Island

Coordinating activities between a group of people is always difficult but can be especially challenging when people have different incentives. This problem is further compounded when individuals are not capable of precisely monitoring the activity of their collaborators. To that end we have begun constructing virtual world experiments that focus on collective action problems with the hope of getting useful data on how people use communication to coordinate towards socially optimal solutions. Our interest concerns experimental designs where the incentives for the group and individual subjects are in conflict. Creating experiments where subjects must make choices between fulfilling their own incentives or supporting the outcome of the group allows for the possibility of deceptive behavior. Our experiments are instrumented so we can record what a subject communicates and also all of their actions.

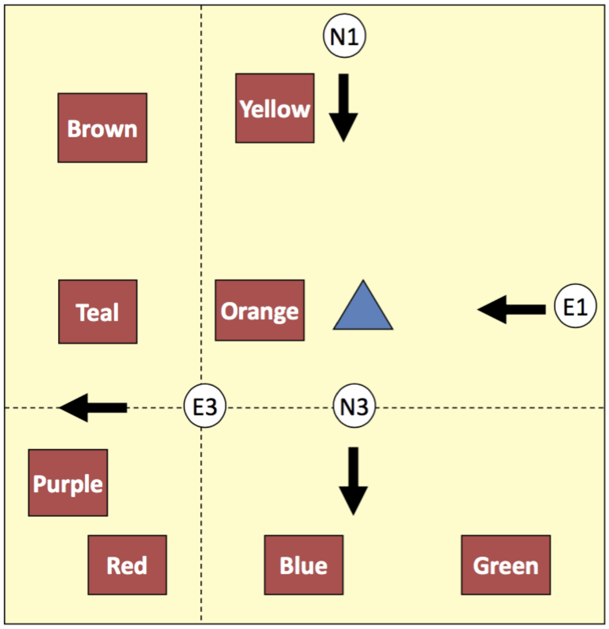

Figure 2: A depiction of the subjects' houses, squares, and weather defense stations, circles. Storms come from either the north or east edge. The arrows depict the areas of the island that would be protected by operating a weather defense station in the presence of a storm coming from the appropriate direction.

Our lab conducted one experiment where eight subjects are placed together on an island inside of a virtual world constructed in Second Life. We refer to the experiment as "Hurricane Island." Each subject is assigned a color and is referred to by that color throughout the experiment. The island contains eight separate houses and each subject is assigned to a particular house. Figure 1 provides an overhead view of the island. The experiment subjects earn money while they stay in their homes. In our experiment the subjects could earn 40 cents per minute while they were inside their homes. However, for the subjects to earn the full amount their house must not become damaged. If a house does become 50% damaged, for example, the subject could only earn up to 20 cents a minute. The island is subject to external events similar to natural disasters, in this case hurricanes, that can damage houses with some frequency throughout an experiment session. Hurricanes appeared on the island approximately every five minutes. Subjects could defend either their house individually or portions of the island by operating what was called a weather defense station. Figure 2 shows the locations of the subjects' houses as well as the weather defense stations. In Figure 2 storms travel in only one of two possible directions, North to South, or East to West. Operating a weather defense station only provides protection for the part of the island behind the location of the station. So, for example, in Figure 2 operating the N3 weather defense station will only be effective for storms coming from the North and will only prevent damage from occurring to the Purple, Red, Blue, and Green houses.

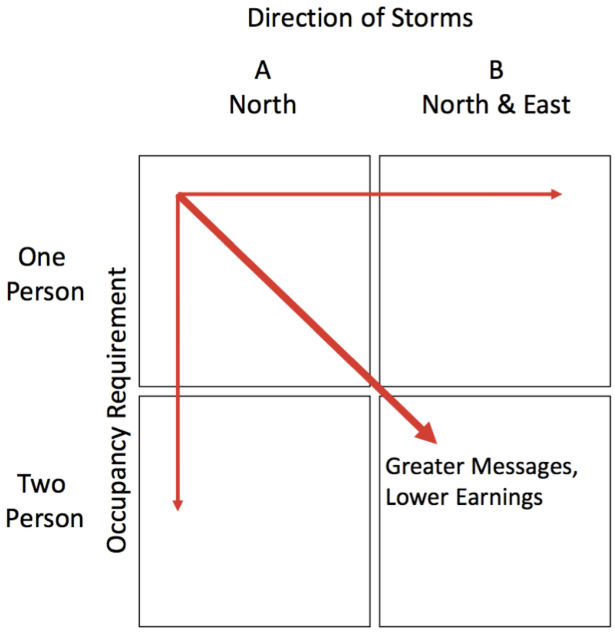

Figure 3: A depiction of the 4 treatment types. The occupancy requirement axis refers to the number of subjects that must occupy a weather defense station to make it operational. The horizontal axis represents whether storms come just from the North or come from both the North and East.

Since subjects are being paid to occupy their houses, doing any other activity such as defending their house individually, repairing their house after it is damaged, or operating a weather defense station represents a loss in earnings. Obviously, each subject could simply decide to defend their own home, but this would mean that the group would not be capable of maximizing the groups total income. The maximum social outcome can only occur when subjects coordinate and have individuals protect the entire island from the effects of the hurricanes by going to the appropriate weather defense stations placed on the island. Our research goal was to investigate institution formation and so we constructed 4 separate treatments for the experiment. We ran a total of 16 sessions across the four treatments and each session lasted approximately 90 minutes. The different treatments involved altering the number of originating directions of the storms and the occupancy requirements for the weather defense stations. Figure 3 provides a graphical overview of the different treatments. In Figure 3, coordination becomes more complex when moving treatments in the diagonal direction towards the lower right cell. The basic treatment, as seen in the top left cell of Figure 3, involved all storms coming from the same direction, North, and requiring only one person in a weather defense station to make it operational. More complex treatments involved having storms that could originate from one of two different directions and also requiring that two subjects be inside a weather defense station to make it operational. Moving from only one person being necessary to operate a weather defense station to two persons places a heavier burden on coordination as now at least two people will not be generating income in their houses and they must also agree on what weather defense station to meet at. Storms originating from two different directions also increases the burden of coordinating as it would require having the appropriate number of people in two separate weather defense stations until the subjects determine the direction of the storm.

An additional design concern related to the placement of the weather defense stations on the island. Each defense station protected the parts of the island behind the station itself. Four weather defense stations were placed in different positions such that one would have some defense stations that would only protect a limited part of the island. It is important to note that the weather defense stations were different distances from each subject's house. Any time they spent defending their house, the island, repairing their home, or traveling to a weather defense station entailed that they were not producing anything and thereby their possible payout would be reduced. Since some weather defense stations were closer to some houses it remained possible that some individuals or groups could decide to only protect their area of the island. The fact that distances were different for each subject and those distances affected their payout could dispose players to form groups to protect only limited parts of the island or it could dispose a player to simply defend their own home.

Throughout the experiment subjects were able to communicate with each other with the chat system provided by Second Life. For the subjects to achieve the optimal social outcome some amount of planning and cooperation is necessary between all the subjects involved in the experiment. When the players in our experiment use the chat system every subject is able to read the message so there is no private communication. One ideal solution solution to achieve the social maximum would entail that subjects find some way of rotating the responsibility of operating the weather defense stations throughout the experiment session. Analyzing the chat data in the earlier parts of the experiment sessions we noticed that subjects frequently attempt to cooperate by identifying a rotation for operating weather defense stations. There is no way, however, for each subject to effectively monitor the behavior of other subjects at all times. and this provides subjects with the possibility of not fulfilling their responsibility. This allows individuals to deceive others by agreeing to do something they have no intention of doing or simply stating that they did something that others cannot verify. In the context of this experiment the subjects could only really determine that another subject had not fulfilled their part of a coordination program by experiencing damage to their houses. At the same time, since the experiment is in a virtual world we configured appropriate monitoring to record the activities of subjects and then compared those actions to whatever messages had been communicated. Comparing the log of subject messages to the actions subjects actually performed allowed us to identify different types of non-cooperative behavior including deception.

Due to the nature of the experiment and the amount of time, groups of subjects used the early minutes of the session to organize their operation of the weather defense stations. A common strategy

Speaker	Message
Yellow	where was teal and orange
Yellow	im getting damaged
Teal	i hv covered
Purple	me too
Red	im getting damaged?
Yellow	teal and orange you did not do it right
Purple	wher s orange?
Yellow	teal and orange were both supposed to go to n1
Yellow	everyone needs to read there messages
Teal	i am at n1
Teal	but no one came
Yellow	orange fault
Teal	orange are you there
Teal	we all loose
Teal	coz of you
Red	earth to orange lol
Yellow	orange hasnt sent a message the whole time
Orange	hey, look at ur manuals again and do the math
Orange	you stand a chance of making a lot more by staying home and defending

Table 1: Example of session communications

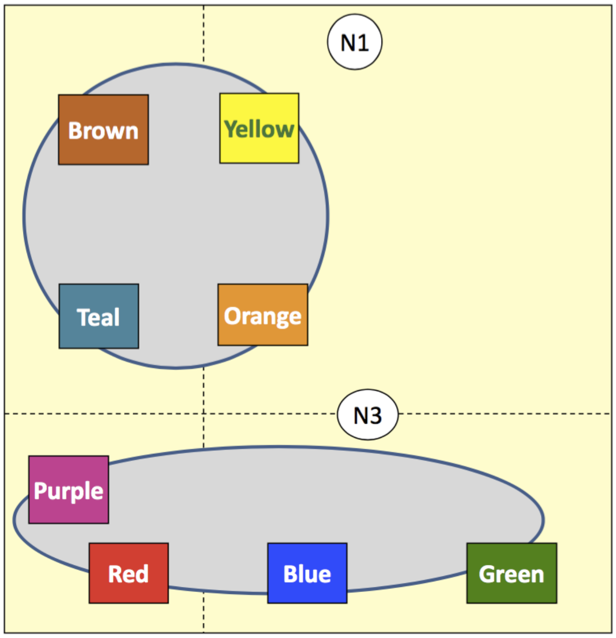

Figure 4: Game theory predictions for storms originating from the north. The players in the top part of the map have an incentive to defend at the N1 weather defense station and thereby protect the whole island. Players in the lower part of the map have an incentive to operate the N3 weather defense station and only protect their part of the island.

that was employed was to have subjects rotate who would take turns operating the weather defense stations. While the strategy was simple, it still meant that enforcement was a challenge as individual subjects could not verify that someone was operating the weather defense station until it was too late. Table 1 provides an example of communication from an experimental session. In the experiment session where the chat took place two subjects are required to operate the weather defense station but only one person has shown up. The orange player has not fulfilled their obligation to operate the weather defense station and is not being communicative with the other subjects for an extended period of time. Eventually the orange player suggests that the reason for their absence is due to a belief that they can have a larger payout if they simply defend their homes individually which is, in fact, incorrect. Since the experiment was designed with different distances to weather defense stations it is reasonable to assume that some subjects would determine that only defending a part of the island is in their best interest.

If one analyzes the layout of the experiment using game theory, it becomes clear that different groups of subjects have different incentives to defend only parts of the island. Figure 4 shows the game theory predictions for storms originating from the North. The specific prediction for North originating storms is that group in the upper half of the map has an incentive to operate the N1 weather defense station. When the subjects in the upper part of the map operate the N1 defense station they are both protecting themselves and also creating a public good for the players in the lower part of the map. One would then predict that players in the lower portion of the map would only be interested in operating the N3 weather defense station which would leave the players in the upper part of the map without any defense. As the treatments become more complex with two subjects required to operate a weather defense station and two possible storm directions the benefits derived from non-cooperative play increase. Creating incentives to not cooperate thereby indirectly encourages the possibility of deceptive communication as subjects would be interested in having other people provide weather defense for the island while they free ride on the benefit.

5 Conclusion

This paper has provided an example showing how the tools and techniques of experimental economics can be used to construct situations where individuals have an incentive to engage in deceptive communication with other subjects. Collecting both the communication data and the actual actions of the individual subjects a virtual world experimental paradigm can create new corpora for research into deception in natural language. In order to assist further research in this area we would like to release several example experiments and tools we used to construct and operate this experiment to the community so that researchers can formulate and construct their own experiments useful for research into deceptive behavior. Our belief is that by leveraging economic theory and techniques developed in experimental economics researchers can construct different experimental scenarios where subjects have incentives that induce behaviors in individuals and create situations where deception should occur. At the same time, by making use of virtual world technology researchers can be effectively omniscient and compare the communication of individuals to their actions in the experiment. To help others interested in constructing a virtual world experiment we have begun to produce documentation and examples. The documentation for this effort along with working code will be available at `http://gmucsn.github.io/VirtualWorldExperimentTutorial`

References

[Al-Ubaydli et al.2014] Omar Al-Ubaydli, Kevin McCabe, and Peter Twieg. 2014. Can More Be Less? An Experimental Test of the Resource Curse. *Journal of Experimental Political Science*, 1(01):39–58, March.

[Almela et al.2012] Ángela Almela, Rafael Valencia-Garc 'i a, and Pascual Cantos. 2012. Proceedings of the Workshop on Computational Approaches to Deception Detection. pages 15–22. Association for Computational Linguistics.

[Gneezy2005] Uri Gneezy. 2005. Deception: The Role of Consequences. *The American Economic Review*, 95(1):384–394.

[Gokhman et al.2012] Stephanie Gokhman, Jeff Hancock, Poornima Prabhu, Myle Ott, and Claire Cardie. 2012. Proceedings of the Workshop on Computational Approaches to Deception Detection. pages 23–30. Association for Computational Linguistics.

[Kosfeld et al.2009] Michael Kosfeld, Akira Okada, and Arno Riedl. 2009. Institution Formation in Public Goods Games. *The American Economic Review*, 99(4):1335–1355.

[North1991] Douglass C North. 1991. Institutions. *The Journal of Economic Perspectives*, 5(1):97–112.

[Ostrom2000] Elinor Ostrom. 2000. Collective Action and the Evolution of Social Norms. *The Journal of Economic Perspectives*, 14(3):137–158.