Liar Buyer Fraud, and How to Curb It

Markus Jakobsson*
Zapfraud Inc.
Email: markus@zapfraud-inc.com

Hossein Siadati*
Department of Computer Science and Eng.
NYU Polytechnic School of Engineering
Email: hossein@nyu.edu

Mayank Dhiman*
Department of Computer Science and Eng.
UC San Diego
Email: mdhiman@eng.ucsd.edu

Abstract*—We describe a common but poorly known type of fraud – so-called *liar buyer* fraud – and explain why traditional anti-fraud technology has failed to curb this problem. We then introduce a counter-intuitive technique based on user interface modification to address liar-buyer fraud, and report result of experiments supporting that our technique has the potential of dramatically reducing fraud losses. We used a combination of role playing and questionnaires to determine the behavior and opinions of about 1700 subjects, and found that our proposed technique results in a statistically significant reduction of fraud rates for both men and women in an experimental setting. Our approach has not yet been tested on real e-commerce traffic, but appears sufficiently promising to do that. Our findings also support that men are more willing to lie and defraud than women are; but maybe more interestingly, our analysis shows that the technique we introduce make men *as honest as women.

I. Introduction

As online commerce has skyrocketed, phishing and malware has gone from obscurity to ubiquity. The technical community has responded firmly, deploying measures such as DMARC and anti-virus filters. However, not all types of fraud enjoy the same attention of the technical community as they do among fraudsters. One such type of fraud is referred to as *liar buyer*. Liar buyer fraud accounts for a significant portion of the losses within some sectors [22], [43], and yet, remains almost entirely undefended against.

While many types of fraud is carried out mostly by career criminals, liar buyer fraud is almost exclusively the result of temporarily poor judgment of otherwise honest people. In a typical liar buyer instance, a consumer orders and receives some merchandise, and then reports it not delivered in order to get a refund. Commonly, the liar buyers are not repeat fraudsters, and many of them are believed to act in response to losing a similar amount to another instance of fraud – then contesting the charges but not being ruled in favor of.

According to industry estimates, liar buyer fraud accounts for between 25 [2] and 50 percent [22] of the direct fraud losses of affected organizations. While the exact amounts of liar buyer losses are unknown, more is known about other, related types of fraud. It is, for example, believed that 10% of all insurance claims and 15% of U.S. tax filings are fraudulent [43].

One thing that makes liar buyer fraud different from other types of online fraud is that traditional countermeasures do not work against it. For example, fraudsters attempting to use stolen credentials are often identified due to an unusual IP address or a lack of cookies and other machine identifiers. Those techniques are not useful to identify liar buyers, since typical liar buyers use their regular computers to commit the fraud. Similarly, many "common" types of fraud are blocked based on detection of a large number of transactions from one and the same subnet, a transaction anomaly, or even an unusual way of entering a password. Again, none of these techniques are helpful to block or detect liar buyer fraud. There are analytics in place to block repeat offenders – these, however, do not help against first-time offenders, since they are simply based on judging how anomalous repeated incidents involving the same user would be.

It is interesting also to note that even when a liar buyer is identified with a reasonable certainty after the fact (e.g., after having filed an unusual number of complaints of not having received merchandise), the fraudster is commonly not confronted with the goal of getting the money back. This is because the cost of collecting the amounts owed often exceeds these amounts, and because companies fear the risk of making mistaken accusations. While delivery confirmations may seem to be potentially helpful, they are, in practice not helpful at all. There is no verification of identity of the signer to begin with, and for some services (like the USPS), the only data that is saved is the fact that a confirmation was received – and the zip code for which the delivery was made. The signature or the name of the signer is not stored, nor is the delivery address. An orthogonal approach to deal with this problem is to improve delivery security. We have not studied the extent to which this is a useful approach. However, our approach, which considers the psychological aspects of fraud prevention, is likely also to apply to settings where there is no physical delivery of goods.

We developed and tested an approach that promises a substantial reduction of liar buyer fraud. The main underlying principle of our approach is to convey knowledge of identifying information to users – *before they commit to a complaint*. Specifically, by clarifying to users that a contested purchase was made using a recognized computer and by geographically pinpointing its location, potential liar buyers become dramatically less likely to complete the fraudulent complaint. This may seem counter-intuitive since, after all, the user is saying that he *did not receive* an item – and *not* that he did not order it.

*Work done while at PayPal.

USEC '15, 8 February 2015, San Diego, CA, USA

http://dx.doi.org/10.14722/usec.2015.23010

We used a role-playing approach to test our techniques. We recruited large numbers of subjects and incited them to perform liar buyer fraud in the context of the role playing scenario. We exposed different sets of subjects to different information as they were in the process of completing a complaint, and measured the statistical impact of these different treatments. Among other things, and as mentioned above, we found that identifying a user's machine as recognized *and* geographically localizing the machine caused a substantial reduction of (simulated) fraud; however, doing only one of these had *no* measurable effect. Moreover, we found that the correctness of the location information is important. More specifically, we found that incorrect location information reduced the benefits of our countermeasures, but that it did not *aggravate* the extent to which users would perform fraudulent actions beyond what subjects are willing to do in the absence of our countermeasures. We also found that while showing a map impacts user actions, there was no such impact to be seen from displaying IP addresses. While in hindsight, this may not seem very surprising, it calls into question the common use of IP addresses in many types of user notifications.

Outline: We begin by describing the related work (section II). In section III, we outline our solution and describe the experiment we used to test our proposed solution. We analyze our findings in section IV and conclude and describe future work in section V.

II. Related Work

Liar buyer fraud has not, to our knowledge, been addressed in the security literature. There are, however, large bodies of work adressing both *lies* and *fraud*.

One important question is what makes regular people lie – or not. There is work that shows that the lying behavior of an individual depends on the lying behavior of a group. More specifically, it has been shown [5], that students were much more likely to cheat if they knew that other students were cheating as well. There is also work on fairness perceptions that may help explain some aspect of what gives rise to liar buyer fraud. Wirtz et al. [43] showed that service disruptions and disappointments gives rise to dishonesty, and that the amount of opportunistic claims depend among other things on the size of the service provider, as well as the customer relationship ties it has. As a consequence, larger companies are more likely to suffer liar buyer fraud, and one-time customers are more likely to be abusive than long-term customers.

Researchers have also conducted various experiments to find out which methods may be used to *detect* lying. Polygraph testing, for example, is a well-known technique used to detect liars based on blood pressure, respiration, and skin conductivity [31]. Linguistic cues can also be used to detect lies. Newman et al. [28] built a classifier to distinguish lie from truth based on text features such as text-complexity, number of self-references, and usage of the negative words; this was later further researched by Hancock et al. [14]. Other cues such as increased cognitive load has also been used to detect liars, who are asked to tell the story in the reverse order [37].

Another line of research that is closer to our effort addresses how to *mitigate* lying. A well-studied context is classroom cheating, with research showing that moral obligations and personal codes of honor can significantly decrease lying behavior [3], [24], [36]. It has also been shown that risk perception affects cheating: For example, students are less likely to cheat in high risk situations i.e., where they think that the possibility of getting caught is high [8], [9], [16]. In fact, surveillance is observed to have significant impact in reducing cheating behavior in general [17], and it has been shown that closed circuit television and associated signage (such as "Security Camera" and "Shoplifters will be Prosecuted") reduces shoplifting [7]. It has even been shown that the mere *presence* of an observer – or just an image evoking an observer – reduces cheating [25].

In the digital realm of mitigation of lying, Hancock et al. [15] showed that the *recordability* of a medium affects the lying rates – people lie less when it is clear that their actions are automatically documented. This confirms the findings in real-world scenarios.

Turning to online fraud, this is also a topic that has received considerable attention during the last few years, after emerging as an academic discipline in the early 2000s. Fraud countermeasures can be broken down into client-side techniques and back-end techniques, with some techniques straddling the fence between the two. Client-side techniques typically either have a filter component (such as client-side phishing detection [1]) or a user-interface component (such as the lock icon, the colored address bar or expired/unsigned certificate messages for HTTPS, and other user-interface related security issues [19], [42].) Back-end techniques commonly are filters – whether methods based on machine learning, and aimed at detecting anomalies (e.g., [30]) or semantic methods such as DMARC, email spam filters and corporate firewalls [32]. As mentioned before, these techniques are not helpful to address liar buyer fraud, in spite of being useful to detect or block a wide array of other types of fraud.

Typical user interface methods with security benefits aim at helping users distinguish legitimate messages/webpages from fraudulent or dangerous ones – for example, by identifying secure connections by coloring URLs, or by including locks and personalized images [10]. Whereas the methods we describe rely on changes to the user interface, the purpose is not to help the user distinguish between good and unsafe messages or sites, but rather, to promote user honesty. The closest related work we are familiar with is the work by Warkentin et al. [41], showing that connections between online and real-world identities curb online deception; and Rule [34], who investigates the importance of good dispute resolution mechanisms in the context of user satisfaction.

Another relevant line of work is surveying and experimentation. Surveying is not suitable to assess the strength of potential countermeasures. The reason is that if one were to ask users what they would do in a hypothetical context, this often does not produce accurate results – especially where morality or deceit is involved. Simply stated, what people say and what they do are poorly aligned in such contexts. Interviewing self-confessed liar buyers is also not practical, other than as a first step towards understanding the problem. Moreover, while the willingness to lie is not hypothetical for these subjects, the use of countermeasures still is. Therefore, it is hard to assess the likely reaction these subjects would have to different countermeasures. As a result, while surveying is commonly used in

many disciplines, it is poorly suited for our context. Another experimental approach used in fraud experiments is so-called naturalistic experimentation [11], wherein users are *unwittingly* participating in an experiment and their natural reactions are measured. This approach offers advantages for studying some types of fraud, and is, for example, useful for carrying out phishing experiments (see, e.g., [18]). However, the approach is not well suited to study liar buyer fraud, as it would require to incite very large numbers of users to potentially commit liar buyer fraud by exposing them to stressful situations. This would be both unethical and impractical. A third experimental approach used by fraud researchers is role-playing [13], and this appears to be the most practical methodology for testing liar buyer countermeasures before deploying them. On one hand, one may argue that role-playing, as opposed to real-life experience, takes away the raison d'être for all the emotions related to liar buyer fraud – greed, a wish for revenge, and fear of legal consequences. By this argument, the results of any role-playing fraud study of our kind would be questionable. On the other hand, there is ample evidence that people transfer real-world behavior to representative but fictional contexts [27], [29], [38], [4], suggesting that if the measured impact is not identical to what would be observed for a real deployment, there would at least be significant similarities.

Our experiment was designed around the Amazon Mechanical Turk platform, and takes advantage of insights from an array of recent work using this platform [20], [21], [23], [33].

III. Hypotheses and Experiment

A. Interviews

At the beginning of our efforts, we interviewed a collection of self-confessed liar buyers – a total of five people, mostly friends and family of one of the co-authors. The goal was to identify the reasons why these otherwise honest people would engage in fraud, and what emotions prompted their actions. All five of them had suffered the economic consequences of fraud or mismanagement[1] and had filed complaints, but were ruled against. This made them feel betrayed and angry, and left with only one option to get justice. None of them had performed more than one fraudulent transaction (but one of them had contemplated it) and for all the users, the amounts in the fraudulent transactions were reported to be approximately the same as the lost amounts. One of the interviewees pointed out that if he had not been the victim of fraud – and the dispute process following this – then he would not have realized how easy it would be to *commit* fraud. We note that these interviews were not conducted systematically and are not part of our results; rather, they were used to develop the story-line for the fraud scenario in the role-playing experiment that we designed.

We also interviewed two customer service representatives at a large service provider concerned with liar buyer fraud, and discussed their view of the problem. Both these interviewees said that they commonly would be fairly certain when a dispute is a liar buyer case, but they had no practical tools to address such cases.

[1]It was not always clear to the interviewees – or the interviewer – whether they had suffered fraud or a system failure.

B. Initial explorations

Following the interviews, an MTurk based role-playing experiment with two conditions was carried out. For the control group (N=318), we asked the subjects whether they would commit liar-buyer fraud to get even with a person who defrauded them in the past. For the test group (N=304), we used the same scenario, except that we added that the webpage where they would file the complaint shows a photo of their front door, with the disputed item delivered. We asked the subjects in both conditions to commit to or abort their compliant. We found that only 24% of subjects in the test group said that they would ask for refund (thereby committing liar-buyer fraud), in comparison with 64% of the subjects in the control group. A Pearson chi-squared (χ^2) test shows that the reduction of the amount of the fraud is statistically significant (χ=99.85, p-value$<$0.001).

While the result of this experiment is promising, it is based on an impractical assumption – namely, that the delivery service would photograph the delivered merchandise. Our main experiment addresses this problem.

C. Hypotheses

Based on the interviews and a general understanding of the nature of the problem, we put forth a collection of hypotheses of what may be impact liar buyer rates. Based on a simplified and preliminary version of the role-playing experiment described in greater detail below, we weeded out those that did not seem to have any potential. Among the hypotheses we put forth were the following:

1) Fraud rates may be impacted by disclosure that the user machine was recognized. We kept the description general, as opposed to specifying "cookies recognized" or "computer browser version and plugin combination recognized", etc. This was done for two important reasons. First of all, too much technical detail is confounding and potentially worrisome to typical users; and second, giving specific information of how devices are recognized may help would-be fraudsters make their devices harder to recognize.

2) Fraud rates may be impacted by disclosure of user location. We decided to try different representations of location – including IP address, a zip code, a map – and to try various degrees of precision, matching the inaccuracy of existing IP-to-geolocation services. We decided to drop the zip code, as its use did not seem to have any impact. However, we kept IP address (in spite of that also showing no promise) since that is commonly displayed to users by many service providers

3) Fraud rates may be impacted by delivery statements. We hypothesized that fraud rates may be affected by a statement from the delivery person that the delivery has been made; potentially combined with photographic evidence. Our preliminary experiments suggested that this hypothesis was correct, but still we dropped it in the follow-up experiment since typical service providers do not have access to such information, and we needed to focus on testing the most promising hypotheses in order to be assured of sufficiently large sample sizes for each treatment.

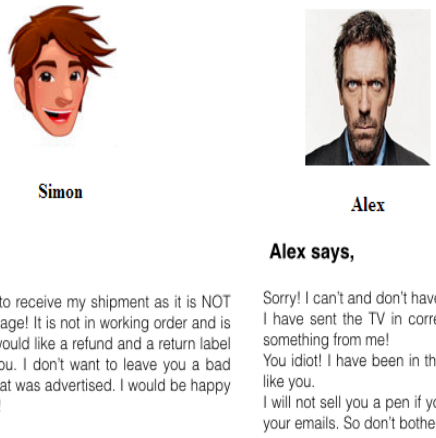

Fig. 1. A snapshot of the role-playing user interface. In this specific step, the user has sent a message (on the left) and received a rude reply from the seller (on the right). The response from the seller is intended to induce anger.

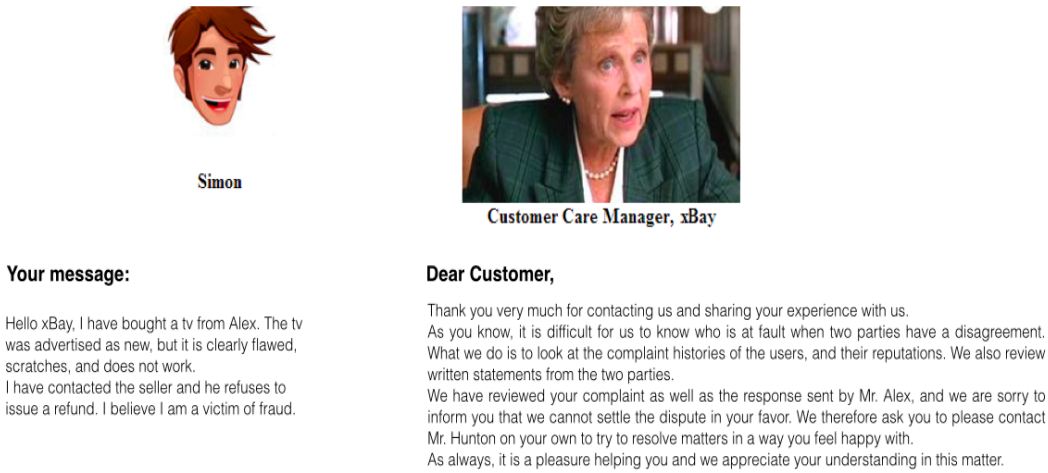

Fig. 2. A snapshot of the role-playing user-interface. The user has sent a message (on the left) and received a customer service response (on the right). The response is intended to disappoint and frustrate the user.

4) Fraud rates may be impacted by return options. We hypothesized that fraud rates may go down if would-be liar buyers were offered an easy way to return items instead of claiming that they did not arrive. This was motivated by a belief that some liar buyers felt that they simply could not afford the merchandise (potentially due to changed situations after the order was placed, or buyer's remorse) and that an easy way to return the merchandise would offer an alternative way of getting refunded, without having to lie. This hypothesis also turned out have significant promise, but still, we decided to exclude it from the main experiment to keep the story line of the experiment structurally simple. A careful validation of this hypothesis is an interesting topic for future work.

5) Fraud rates may be impacted by forcing the user to promise. It is well known that moral obligations reduce the rate of lying. Credit card companies successfully use the approach by reminding payers of their moral obligations as they sign a receipt – typically, the text on the receipt includes mention of the signer's intention to pay his or her bills. Surprisingly, we did not see any impact of this approach in our preliminary testing, and so, decided to not pursue it in the main experiment.

6) Fraud rates may be reduced by paying special attention to angry users. We have come to believe that much of the liar buyer fraud is motivated by anger and a feeling of not being treated fairly. We wanted to test the importance of kindness from customer representatives when these respond to initial complaints filed by users before the users proceed to filing a dispute. Since this may require a prioritization of efforts based on perceived anger, a tool based on natural language processing could be used to identify particularly angry complaints. The preliminary experiments showed some potential impact of this approach, but not as great as other methods, and we decided to exclude the testing of this hypothesis from the main experiment to keep things simple.

D. Using Amazon MTurk Multi-Round Experiments

We use multi-round subject interaction Whereas this is not directly supported on Amazon Mechanical Turk, it is possible to use built-in communication constructs to create such a structure.

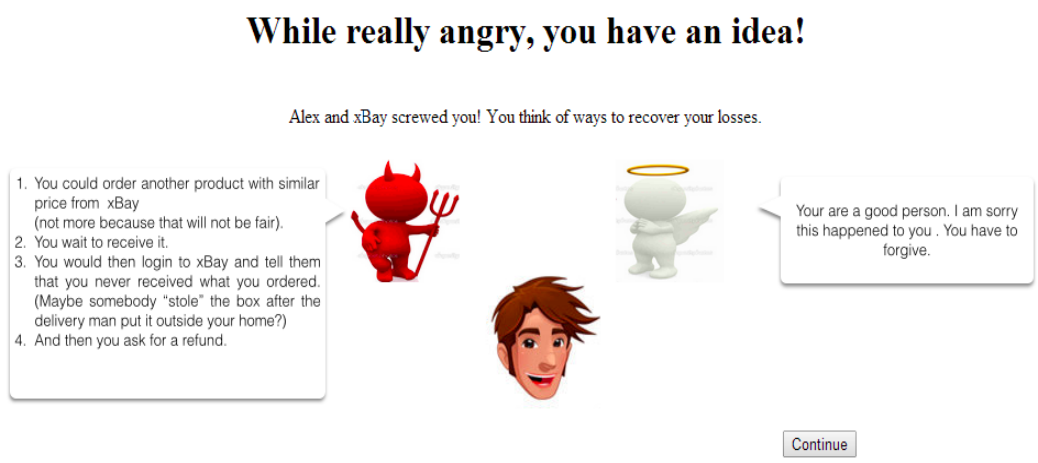

Fig. 3. This figure shows the "temptation" phase, which is used to introduce the idea of performing liar buyer fraud to get even. The angel on the right side tells the user not to fall for the temptation – this was added to avoid that the subjects feel that they are only given one option.

The first round is based on a publicly advertised task, and is used to build a large database of potential subjects and to collect demographic information. The second round is announced privately among those who participated in the first round. One advantage of this approach is that it allows us to filter out participants based on specific criteria, e.g., the correctness of the reported location. (Also, one could filter out based on demographics, to compensate for existing biases, although we did not do that in this study.) Another advantage of the two-round approach is that it "hides" well-paying tasks from low-quality workers by only disclosing them to workers who pass the first round. This is helpful to minimize the involvement of workers who will answer in an arbitrary manner. A collection of other techniques can be used to filter out low-quality workers. One such technique is to include questions or tasks intended solely to detect carelessness. Another is to include experimental "branches" where it is obvious to the subjects what branch will be the least demanding – and place the "experiment payload" on the other branch. We did not have to use either of methods, since part of task we specified was to write reasonably demanding letters. We reviewed the letters in an attempt to identify obvious cheaters (but did not find any such.)

Round 1: Recruiting Subjects. We recruited 2364 Amazon Mechanical Turk workers from the United States to perform a simple task in which each subject was asked to choose what five avatars out of a collection of a hundred most closely represented him or her, and indicate the city and state where he or she was located, paying each subject 10 cent for their participation. We filtered out any participants for whom there was a large discrepancy between the stated location and the geolocation associated with the observed IP address. Using the messaging service provided by Amazon Mechanical Turk, we invited subjects to participate in a second experimental round. A total of 855 subjects participated in the second round, which for each subject consisted of a common task and one out of six different treatments, described next.

Round 2, Part 1: Victimization. We ran a between-subjects experiment with 6 different conditions. All subjects were exposed to the same role-playing storyline, in which the subject first experiences being defrauded, and later gets an opportunity to get even by acting as a liar buyer.

More specifically, the role playing game starts by the user ordering a TV. However, when it arrives, it turns out to be defect – and it seems like the seller must have known that it was broken when he shipped it. We make the subject perform tasks to make him identify with the victim, and tasks to make him angry – these are described next:

Making Subjects Identify with the Buyer. We use two methods to make the subjects identify with the buyer in the role playing game. First, the story is told in a way that the subject *is* the buyer – and is represented by an avatar selected[2] by the subject. Second, one of the tasks of the subject is to write messages on behalf of the fictional buyer. After the subject writes a complaint to the seller, he or she would receive a rude response, supposedly written by the unethical seller. See figure 1 for a screenshot showing the subject (on the left) being insulted by the seller (on the right).

Making Subjects Angry. It is well known that people do not maintain a good separation between contexts when it comes to their emotions. This is why upset people are commonly letting their anger spill on to innocent people. We used this to our advantage to "manufacture" the most appropriate emotional context for our experiment – we wanted to make our subjects feel angry and agitated, matching the typical[3] emotions of liar buyers. We did this by letting subjects be insulted (as shown in figure 1), and then appeal to but receive no help from "authorities" (figure 2.)

Round 2, Part 2: Getting Even. After having been defrauded, insulted and ruled against, the subject is offered a "solution" in the form of a temptation to get even (see figure 3.)

[2]Each subject got to choose an avatar to represent himself or herself, from a list of the 25 most popular avatars, as indicated by the selections in round 1.

[3]This is based on our interviews with users who have committed liar buyer fraud, where the fraudsters reported acting out of a feeling of anger, stoked by a sense that they needed to take the situation in their own hands to get justice.

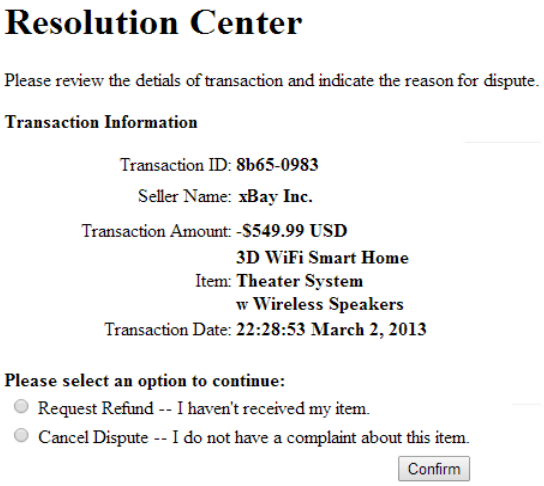

Fig. 4. This image shows a snapshot of the "control" treatment, which mimics real dispute webpages. At the bottom, the user is given the opportunity to pursue the complaint or to cancel the dispute. We measure the rates –and relative rates– of pursuing complaints for the different treatments.

The subject is finally given an opportunity to get even – by ordering a similarly priced item (as the defect item), receive it, and report that it was not received in order to get a refund. In one version of the exploratory experiments, we used the same product (i.e., a TV) for the complaint page, but we determined that users confused it with the damaged item that they received before. To avoid this misunderstanding, we used a "3D smart home theater system" instead, with a similar price. We also explicitly mentioned that the seller is xBay, not the person who defrauded him/her in the first place.

The subjects are exposed to one out of six treatments on the page where he or she has to commit to his actions (i.e., choosing to be a liar buyer, or decide not to.) The treatments can be described in the following way:

1) Control: This treatment corresponds to information which is usually shown in common resolution centers, and is used as our reference treatment, allowing a subject either to confirm or to abandon his complaint. A screenshot corresponding to this treatment is shown in figure 4.

2) IP: This treatment adds the display of the subject's IP address to the control treatment. (We used the subject's real IP address, as observed during the experiment.)

3) IP+map: Subjects in this treatment would be shown their IP address and a map locating them. We generated the map from the user-provided location information in the subject-recruiting part of the experiment.

4) Recognized computer: In this treatment, the subjects are shown material like that in the control treatment, except with an assertion that the computer is *recognized*.

5) IP+map+recognized computer: This treatment combines the features of the above treatments – a screenshot is shown in figure 5.

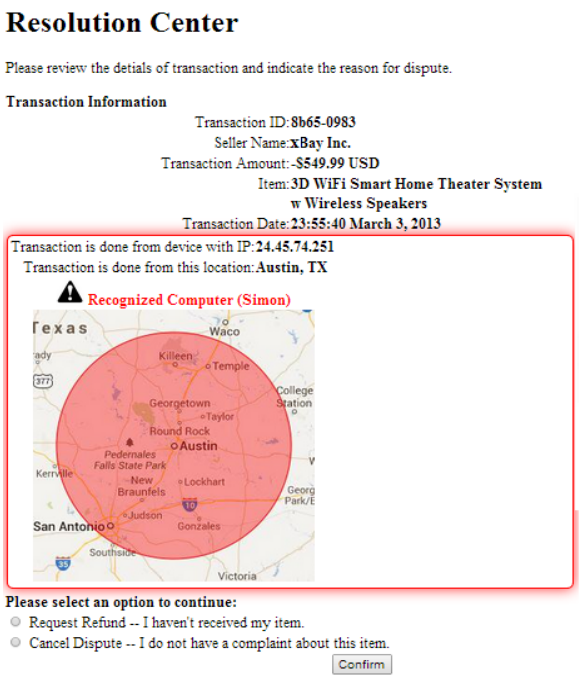

Fig. 5. This image shows a snapshot of the "IP+map+recognized computer" treatment. The user is told that her computer is recognized, and her location is shown on a map. At the bottom, the users gets to choose to pursue or abandon the complaint, just like in the other treatments.

6) IP+wrong map+recognized computer: This treatment is like the treatment above, except that we intentionally used a map showing an incorrect location.

The goal of the experiment was to quantify the differences in fraud rates between users of different treatments.

IV. Experimental Findings

A. Analysis and Findings

For each treatment, the subjects see information related to the complaint, such as the merchandize, the seller and the amount. Different treatments add different additional information, e.g., IP address; a notice stating "Recognized computer" and a map – whether accurate or not. For each treatment, the user is asked to perform a selection – requesting a refund or cancel the dispute.

Each subject only sees one of the treatments, and so, we do not know how he or she would have responded had we presented him or her with another treatment. However, we compare the statistics (of requesting refund vs. cancelling the dispute) for large number of users associated with different treatments. In the context of the experiment, the former corresponds to committing liar buyer fraud and the latter corresponds to being honest. Comparing the statistics associated with different treatments allows us to determine what impact the different user interfaces have on the honesty of the users in our experiment.

As we mentioned before, we are not asserting that the behavior would have been *identical* in a real-life scenario. For example, in our experiment, one might argue that subjects

do not have to fear the consequences of being dishonest, and therefore, would be willing to cheat – while in a real-world scenario, one might argue that people would cheat less. On another hand, one might argue that subjects in our experiment have nothing to gain from being dishonest – and therefore, would cheat less than in a real-world setting. In the end, we do not know[4] what force would be strongest, and simply see our experimental results as indicative of a solution that is likely to have a real-world impact.

We invited subjects who completed the avatar selection task to participate in the main experiment via private messages in the Amazon Mechanical Turk platform. A total of 855 subjects (36% out of the total of 2364 subjects in the first round) responded and participated in the task. Participants were randomly given one of the six different described treatments. Figure 6 shows the percentage of the liars in each treatment. 20% of the participants (25 out of 127) in the "Control" treatment chose the "Request Refund–I have not received my item" which corresponds to lying. In the "IP" treatment, 22% (29 out of 129) chose to lie, and 22% (28 out of 128) chose to lie in "IP+Map" treatment. Significantly different, only 10.5% of subjects (13 out of 124) in the "IP+map+recognized computer" treatment decided to lie. In the "recognized computer" treatment, on the other hand, 23% of participants (32 out of 138) chose to lie, and 21% of subjects (44 out of 209) in the "IP+wrong map+recognized computer" lied.

A Pearson chi-squared (χ^2) test between each pair of the treatments shows that there is a significant difference between the behavior of the control group and the subjects in the treatment referred to as "IP+map+recognized computer" (χ^2=4.13, p-value < 0.05). This shows that the modified user interface which contains IP, a correct map, and signage of "recognized computer" reduces the percentage of liars by 50% in comparison with user interfaces currently in use. We should emphasize that the exact reduction in *real* settings will only be known when our suggested modification is deployed for a real dispute scenario.

We also studied the effect of the map accuracy by running another treatment, which we call "IP+wrong map+recognized computer'. This was very similar to the "IP+map+recognized computer" treatment, except that we showed a random map instead of the correct map. 209 subjects were given this treatment and 21.0% of these chose to lie. A Pearson chi-squared (χ^2) test on the treatments "IP+map+recognized computer" and "IP+wrong map+recognized computer" shows a statistically significant difference (χ^2=5.66, p-value=0.017). Indeed, the effect of wrong map cancels the effect of the modification in the user interface. This shows the importance of the accuracy in the report of the location of the user and reflection in user interface.

B. Exit Survey Analysis

At the end of the experiment, we asked subjects to complete a survey. We asked their gender (46% were female, 54% male);

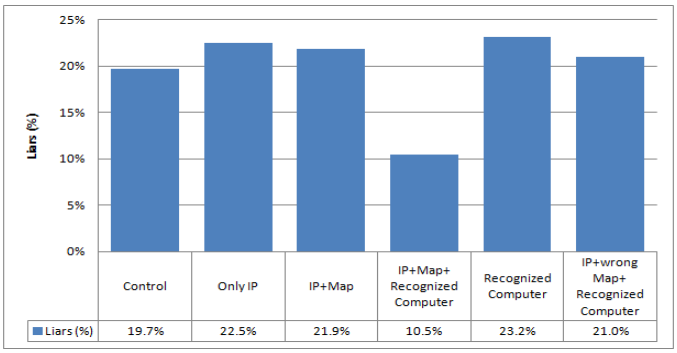

Fig. 6. This diagram shows the percentage of *liars* for different versions of user interfaces in resolution center. The percentage of the liar is reduced 45% in "IP+map+recognized computer" in comparison with the currently used versions ("Control" treatment)

whether it is fair to trick people to get even; *whom* it is fair to trick to get even; and whether the subject has had an experience similar to the one in the role-playing story. We then correlated the responses to their behavior in the role-playing experiment – and, more specifically, to whether they committed liar buyer fraud in the experiment. The findings were very interesting.

- **Male dishonesty.** First of all, we confirmed the common belief that men are more willing to commit fraud [26], [40]. Among all participants, 15% of the female subjects committed liar buyer fraud, while 22% of male subjects did (χ^2=7.4, p-value<0.001) Of those users who said that they had been tricked in a real-life online situation, 8% of the women (a total of 142 having been tricked) said they had gotten even, while 17% of the men (a total of 155 having been tricked) said they had. Among those who had *not* been tricked in a real-life online situation, 28% of the women said that they would *consider* getting even, while 36% of the men said they would. These observations show that men are much more willing than women are to lie and commit liar buyer fraud.
- **Swing liars.** Interestingly, when we compared the rates of committing liar buyer fraud in the treatment we refer to as 'IP+map+recognized computer" treatment of our experiment, we found that 12% of the female subjects and 10% of the male subjects committed fraud (χ^2=0.012, p-value=0.91), meaning that in the presence of the appropriate counter-measures, women and men commit liar buyer fraud to an equal extent.
  More specifically, our technique introduces a deterrence that applies to a greater extent to men, and as a result, reduces the (higher) fraud rates among men to the same as level as the it reduces the fraud rates of women to.
  One interpretation is that, in some emotional situations, a certain percentage of people is *willing* to commit liar-buyer fraud, independently of the user interface; another percentage is *unwilling* to do so, also independently of the user interface, while the actions of some percentage depend on the user interface. We may call the users of the latter group "swing liars" – like swing *voters*, the swing liars can be influenced

[4]Although we *do* know that the actual real-world consequences of liar-buyer fraud are not very severe, and in typical situations, none at all. This, as mentioned before, is because service providers are fearful of creating negative publicity by accusing potential cheaters. Our technique is designed with this in mind, and does not accuse anybody.

until the very last moment. In our experiment, 3% of the female subjects and 12% of the male subjects were swing liars.

- **The impact of experience.**
  We observed that among those subjects who responded that they had been tricked (29% of the participants), 40% committed liar buyer fraud in our experiment, while only 15% of those who reported not having been tricked did (χ^2=20, p-value<0.001) . This supports that having had a bad experience makes people willing to defraud others. Moreover, among those who were tricked *and* got even (28 subjects), 47% committed liar buyer fraud in our experiment.

C. Insights gained

Our findings identify a promising method to address the liar buyer problem. Analogously to how shop lifting is curbed by integrating (and making visible) security elements in stores, we find that we can address fraud by conveying information to users who are potentially in the process of committing fraud. Our techniques can most likely be improved in follow-up research, and our general principles may find applications to address other related fraud as well. The strong gender differences we found are of interest; especially as our proposed technique seems to make men as honest as women.

D. A Note About Map Accuracy

Our experiment obtained accurate location information simply by asking users for their location in round 1 of the experiment. We verified that the IP addresses were reasonably consistent between the first and second rounds; and used the location information from the first round in the second round. This was finally verified in a brief survey at the end of round 2.

In addition to relying on the user-provided location, and performing sanity checks on this, we estimated to what extent the user-provided location matched location information generated by free IP-to-geolocation tools, to establish a lower estimate of how accurately a service could obtain the location. For this purpose, we used three IP to geolocation conversion services: hostIP, IPInfoDB and freegeoIP. We queried all three of the above mentioned geolocation services with the test subject's IP address and used a vote-based approach to determine the user location. This information was scored by asking the user how accurate the shown location was, and comparing the shown location with the estimated location. Based on these self-reported answers, we calculated the accuracies of the IP to geolocation services – these are shown in figure 7.

Note that our algorithm which combines the geolocation from the three different services provide much better accuracy than any one of them individually. However, our approach, which achieves a geolocation accuracy of almost 89%, is still not as accurate as those provided by geolocation based services proposed in various papers [39], [35]. We used user-reported locations as a way to simplify the experiment design, since geolocation quality is not a focus of our effort.

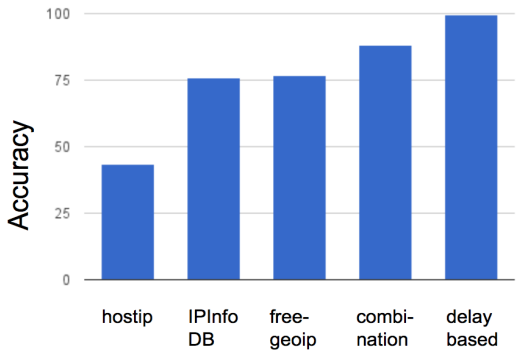

Fig. 7. This image shows a comparison of accuracies of various free geolocation services. This is of relevance since the accuracy of the location information conveyed to the user directly impacts the fraud rates.

E. Digital Identity: Confirming Privacy Beliefs

Our design is centered around our belief that the more a user believes that a service provider knows about him, the less he would be willing to lie. To verify that typical users perceived a difference in the extent to which their identity was known in the different conditions of our experiment, we performed an additional experiment, solely focused on establishing the connection between the identity indicators and the believed ability for a service provider to match a session to a user's real-life identity.

We recruited 300 MTurk users, each one paid $0.12, and divided them into three different groups.

- For group A – corresponding to the IP-only condition of the main experiment – we asked "A friend of yours downloads pirated movies from a website that later gets raided. As a result, authorities learn your friend's IP address. Do you think this gives enough evidence that your friend downloaded the movies?"
- For group B – corresponding to the IP+Map condition – we modified the above question to read "... authorities learn your friend's IP address, and information that determines your friend's approximate location ..."
- Finally, for group C – corresponding to the condition IP+Map+Computer – the question was modified to read "... authorities recognize your friend's computer and learn your friend's IP address and information that determines your friend's approximate location ...".

Each subject was asked to select an answer using a multiple-choice option, with the options (1) Yes, (2) No, (3) Maybe. The subjects' answer to the question is shown in Table I. We ran a Mann-Whitney's U test to evaluate the difference in the response between each pair of groups. We observe that combined information of IP, Map, and Computer identifiers increased sense of a service provider being able to match a user's digital actions to his physical person in comparison with only IP (W = 4165, p-value < 0.05); there was no significant difference between the other conditions.

This result reconfirms our finding from the role-playing experiment where the combination of all identifiers promoted

Group	Identifiers	Yes	No	Maybe
A	IP	17	63	20
B	IP+Map	12	56	32
C	IP+Map+Computer	24	46	30

TABLE I. THE RESPONSES IN THE DIGITAL IDENTITY EXPERIMENT. "YES" MEANS THAT THE SUBJECT BELIEVES THAT THE SERVICE PROVIDER CAN ASSOCIATE AN IDENTIFIER WITH A PERSON.

users honesty, while each of identifiers individually or any subset of them did not promote users' honesty. This also supporting our beliefs that an increased amount of disclosed information leads to an increased sense of a service provider being able to match a user's digital actions to his physical person.

F. Generalizations and Limitations

One limitation of our experiment is that the conclusions we draw are specific to settings where potential liar buyers are motivated by revenge. Although we do not see any reasons why our methods would not work for more general settings, we have also not explored the extent to which our methods would generalize to such settings.

Another important thing to note is that our experiment only establishes that our proposed method results in a significant reduction in fraud rates – it does *not* establish what the rates are. The reason for this is that subjects in our experiment are not motivated by actual greed or fear (of punishment), as real users contemplating liar buyer fraud would be. Therefore, our experiments are only telling us that the proposed method is worthy of being A/B tested on real populations.

V. Conclusions and Future Work

We have tested a collection of possible user-interface enhancements aimed at reducing liar buyer fraud. We have found that showing users in the process of filing a dispute that (1) their computer is recognized, and (2) that their location is known dramatically reduces the willingness to file false claims. We believe the reason for the reduction is that the would-be liars can *visualize* their lack of anonymity at a time when they are deciding whether to perform a fraudulent action. Interestingly, we also showed that users were not affected by knowing that their computer was recognized, but without their location being pin-pointed, or the other way around. We also determined that a reasonably accurate map was necessary – but that an inaccurate map does not seem to increase the willingness to lie.

We want to acknowledge that we do not fully understand the exact reasons why our approach works, and that this warrants more in-depth studies to explain. We also lay no claims to having found the approach that deters fraud the most, nor have we have had the opportunity to generalize our findings to different scenarios. In our view, the approach we have described is a first, and very promising, step toward a deeper body on knowledge about how to deter fraud, and we hope that our work will serve to inspire follow-up efforts.

To our knowledge, our result is both the first to address liar buyer fraud and the first to mitigate lying by disclosing information about the user's computer and location. We hope that our findings will entice others to study the problem and to improve on our result. Our experiment indicates that the suggested approach is helpful in reducing fraud, but it is hard to know – in a real-world scenario – what the actual reduction would be. Our experiment just involved hypothetical consequences – there was no actual monetary benefit associated with successful cheating, nor any actual punishment associated with a failure. While our experiments indicate that fraud would be reduced using our approach, only real-world deployments will be able to determine the exact benefits.

There are many reasons why liar buyer fraud is not well publicized. One reason is that while informed consumers can be cautious and thereby reduce their exposure to abuse such as phishing and virus-based attacks, there is not much that typical consumers can do to reduce the losses due to liar buyer fraud. (In this sense, liar buyer fraud is more similar to click fraud – which, although better known than liar buyer fraud, is also not part of the consumer vocabulary.) In fact, it is quite conceivable that increased awareness of the liar buyer problem among the public may cause losses to *increase* [6], [12]. Another reason why liar buyer fraud is not first page news is that it has the semblance of a victimless crime, much like tax evasion. This is based on a common misunderstanding of who bears the burden of such losses, though. Quite commonly, it is a peer consumer who loses money – in fact, it is well known in the industry that such losses often *cause* liar buyer fraud [2], in addition to commonly being *caused* by it. Other times, the losses are absorbed by an organization and passed on to consumers in the guise of higher service charges. One final reason why media does not report on liar buyer fraud is the absence of a good countermeasure – and the associated absence of an advocate wishing to promote awareness. This, no doubt, has stifled the development of methods to curb liar buyer fraud, and we hope that our results may serve as a first step to overcome this impasse by starting an earnest discussion of the problem in the technical community.

That said, there is a wealth of interesting questions to be addressed. We did not carefully test whether being offered an easy return option at the time of dispute would increase honesty, but have preliminary indications that this may be so. We did not pursue testing the impact on the fraud rates of statements from delivery persons, since this is not currently a practically available option. Moreover, it is possible that it may reduce fraud rates if it were possible to automatically detect situations with increased likelihoods to lead to fraudulent disputes, and to give special attention to address the emotional context of the users. Furthermore, it is interesting to study whether the proposed solutions could have any negative brand effect, i.e., whether users (and in particular, honest users) may feel that the mechanisms we use to reduce fraud affect their comfort. Will users, for example, feel that they are under surveillance? If this is so, one resulting question is what user interfaces have the beneficial effects we have discussed, but which do not cause brand damage. Alternatively, is it possible to detect very likely liar buyer disputes *before* the users commit to their actions, and provide only these users with a user experience that decreases the likelihood of fraudulent claims? Another question worth studying why a user is affected by the fraud deterrence user interface: is it because of an increase of guilt, fear of getting caught, or is related to the increased feeling of being observed?

Finally, it is a vital question to determine exactly how the effects we see transfer to real-world settings, and whether they are persistent. It is altogether possible that the effect would wear off over time.

REFERENCES

[1] S. Abu-Nimeh, D. Nappa, X. Wang, and S. Nair. A comparison of machine learning techniques for phishing detection. In *Proceedings of eCrime '07*, pages 60–69, New York, NY, USA, 2007. ACM.

[2] M. Barrett. Personal communication, June, 2013.

[3] L. Beck and I. Ajzen. Predicting dishonest actions using the theory of planned behavior. *Journal of research in personality*, 25(3):285–301, 1991.

[4] G. Bente, S. Rüggenberg, N. C. Krämer, and F. Eschenburg. Avatar-mediated networking: Increasing social presence and interpersonal trust in net-based collaborations. *Human communication research*, 34(2):287–318, 2008.

[5] D. N. Bunn, S. B. Caudill, and D. M. Gropper. Crime in the classroom: An economic analysis of undergraduate student cheating behavior. *The Journal of Economic Education*, 23(3):pp. 197–207, 1992.

[6] D. N. Bunn, S. B. Caudill, and D. M. Gropper. Crime in the classroom: An economic analysis of undergraduate student cheating behavior. *Journal of Economic Education*, pages 197–207, 1992.

[7] C. A. Cardone. *Opportunity makes the thief: Analysis of the physical cues that influence shoplifter perceptions of the retail interior and the decision to steal*. PhD thesis, University of Florida, 2006.

[8] K. J. Corcoran and J. B. Rotter. Morality-conscience guilt scale as a predictor of ethical behavior in a cheating situation among college females. *The Journal of general psychology*, 114(2):117–123, 1987.

[9] M. K. Covey, S. Saladin, and P. J. Killen. Self-monitoring, surveillance, and incentive effects on cheating. *The Journal of Social Psychology*, 129(5):673–679, 1989.

[10] R. Dhamija and J. D. Tygar. The battle against phishing: Dynamic security skins. In *Proceedings of the 2005 symposium on Usable privacy and security*, pages 77–88. ACM, 2005.

[11] P. Finn and M. Jakobsson. Designing and conducting phishing experiments. In *In IEEE Technology and Society Magazine, Special Issue on Usability and Security*, 2007.

[12] F. Gino, S. Ayal, and D. Ariely. Contagion and differentiation in unethical behavior the effect of one bad apple on the barrel. *Psychological Science*, 20(3):393–398, 2009.

[13] J. D. Greenwood. Role-playing as an experimental strategy in social psychology. *European Journal of Social Psychology*, 13(3):235–254, 1983.

[14] J. T. Hancock, L. E. Curry, S. Goorha, and M. Woodworth. On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes*, 45(1):1–23, 2007.

[15] J. T. Hancock, J. Thom-Santelli, and T. Ritchie. Deception and design: The impact of communication technology on lying behavior. In *Proceedings of the CHI '04*, pages 129–134, New York, NY, USA, 2004. ACM.

[16] J. P. Houston. Cheating behavior, anticipated success-failure, confidence, and test importance. *Journal of Educational Psychology*, 69(1):55, 1977.

[17] J. P. Houston. College classroom cheating, threat, sex and prior performance. *College Student Journal*, 17(3):229–235, 1983.

[18] M. Jakobsson and J. Ratkiewicz. Designing Ethical Phishing Experiments: A Study of (ROT13) rOnl Query Features. In *Proceedings of the 15th International Conference on World Wide Web*, WWW '06, pages 513–522, New York, NY, USA, 2006. ACM.

[19] M. Jakobsson, A. Tsow, A. Shah, E. Blevis, and Y.-K. Lim. What instills trust? a qualitative study of phishing. In S. Dietrich and R. Dhamija, editors, *Financial Cryptography and Data Security*, volume 4886 of *Lecture Notes in Computer Science*, pages 356–361. 2007.

[20] P. G. Kelley. Conducting usable privacy & security studies with amazon's mechanical turk. In *Symposium on Usable Privacy and Security (SOUPS)(Redmond, WA*. Citeseer, 2010.

[21] A. Kittur, E. H. Chi, and B. Suh. Crowdsourcing user studies with mechanical turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 453–456. ACM, 2008.

[22] M. Lashmar. 50% of online payment fraud is likely "liar buyers". http://l3payments.com/blog/suspected-50-of-online-fraud-is, February 2013. Accessed: 2014-07-25.

[23] W. Mason and S. Suri. Conducting behavioral research on amazonś mechanical turk. *Behavior research methods*, 44(1):1–23, 2012.

[24] K. M. May and B. H. Loyd. Academic dishonesty: The honor system and students' attitudes. *Journal of College Student Development*, 34(2):125–129, Mar 1993.

[25] N. Mazar, O. Amir, and D. Ariely. The dishonesty of honest people: A theory of self-concept maintenance. *Journal of marketing research*, 45(6):633–644, 2008.

[26] D. L. McCabe and L. K. Trevino. Academic dishonesty: Honor codes and other contextual influences. *Journal of Higher Education*, pages 522–538, 1993.

[27] C. I. Nass and C. Yen. *The man who lied to his laptop: what machines teach us about human relationships*. Current, 2010.

[28] M. L. Newman, J. W. Pennebaker, D. S. Berry, and J. M. Richards. Lying words: Predicting deception from linguistic styles. *Personality and social psychology bulletin*, 29(5):665–675, 2003.

[29] K. L. Nowak and F. Biocca. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 12(5):481–494, 2003.

[30] Y. Pan and X. Ding. Anomaly based web phishing page detection. In *Computer Security Applications Conference, 2006. ACSAC'06. 22nd Annual*, pages 381–392. IEEE, 2006.

[31] D. C. Raskin. Polygraph techniques for the detection of deception. 1989.

[32] R. F. Rights. Use offense to inform defense. find flaws before the bad guys do. 2013.

[33] J. Ross, L. Irani, M. Silberman, A. Zaldivar, and B. Tomlinson. Who are the crowdworkers?: shifting demographics in mechanical turk. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, pages 2863–2872. ACM, 2010.

[34] C. Rule. Quantifying the economic benefits of effective redress: Large e-commerce data sets and the costbenefit case for investing in dispute resolution. *U. Ark. Little Rock L. Rev.*, 34:767–833, 2012.

[35] B. Schneier. Pinpointing a computer to within 690 meters, April 2011.

[36] C. P. Smith, E. R. Ryan, and D. R. Diggins. Moral decision making: Cheating on examinations1. *Journal of Personality*, 40(4):640–660, 1972.

[37] A. Vrij, S. A. Mann, R. P. Fisher, S. Leal, R. Milne, and R. Bull. Increasing cognitive load to facilitate lie detection: the benefit of recalling an event in reverse order. *Law and human behavior*, 32(3):253, 2008.

[38] P. Wallace and J. Maryott. The impact of avatar self-representation on collaboration in virtual worlds. *Innovate: Journal of Online Education*, 5(5):n5, 2009.

[39] Y. Wang, D. Burgener, M. Flores, A. Kuzmanovic, and C. Huang. Towards street-level client-independent IP geolocation. In *Proceedings of NSDI'11*, NSDI'11, pages 27–27, Berkeley, CA, USA, 2011. USENIX Association.

[40] D. A. Ward. Self-esteem and dishonest behavior revisited. *The Journal of social psychology*, 126(6):709–713, 1986.

[41] D. Warkentin, M. Woodworth, J. T. Hancock, and N. Cormier. Warrants and deception in computer mediated communication. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 9–12. ACM, 2010.

[42] A. Whitten and J. D. Tygar. Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0. In *Proceedings of the 8th Conference on USENIX Security Symposium - Volume 8*, SSYM'99, pages 14–14, Berkeley, CA, USA, 1999. USENIX Association.

[43] J. Wirtz and J. R. McColl-Kennedy. Opportunistic customer claiming during service recovery. *Journal of the Academy of Marketing Science*, 38(5):654–675, 2010.