

Deceit and Deception: A Large User Study of Phishing

Alex Tsow and Markus Jakobsson
{atsow, markus}@indiana.edu

Abstract

This study is a large scale investigation of trust manipulation tactics used by phishing web sites and email messages. The experiment focuses on media authenticity evaluations, rather than content credibility with the assumption that its authors are known. It tests the effect of features ranging from URL plausibility to trust endorsement graphics on a population of 398 subjects. The experiment presents these trust indicators in a variety of stimuli since reactions will vary according to context. In addition to testing specific features, the test gauges the potential of a phishing tactic that spoofs third party program administrators rather than a brand itself. The results show that indeed graphic design can change authenticity evaluations *and* that their impact varies with context. We expected that authenticity inspiring design changes would have the opposite effect when paired with an unreasonable request, however our data suggest that *narrative strength* – rather than underlying legitimacy – limits the impact of graphic design on trust and that these authenticity-inspiring design features improve trust in both genuine and forged media.

Keywords: *phishing, authenticity, unsubscribe attack, third-party impersonation, narrative, endorsement logo, padlock, domain name alignment*

1 Introduction

Phishing – a short term scam using fraudulent email messages to solicit victims for disclosure of sensitive information through fraudulent, but authentic-looking, web sites – breaks two important assumptions made by much of the previous online trust research: that users can identify the agents with whom they interact and that it is expensive for web sites to shut down and redeploy. This paper contributes to an emerging body of research [7, 8, 21, 14, 13] that re-examines user behavior in under these degraded assumptions. Sending unsolicited email and hosting web sites for short periods of time has become an exceedingly inexpensive activity due to the prevalence of botnets [4] and the resulting commoditization of compromised hosts. Phishing scams last only on the order of days [1] and dupe most of their victims in the first 24 hours [13], so their infrastructure is not subject to the costs imposed by many e-commerce trust systems. For instance, third party endorsements can be revoked from misbehaving online retailers; this hurts retailers because they can't simply shut down and re-open with a brand new identity. On the other hand, phishers may freely display endorsement logos with confidence that any retaliation to their abuse will occur long after the fraudulent host has vanished. They use cheap and anonymous botnet nodes for short term hosting and are unconcerned with long term consequences that face an online retailer.

We examine various features in both email messages and web sites and how the impact the perception of authenticity. While this assesses trust manipulation tactics from the phisher's toolchest, it also informs legitimate institutions about how to credibly communicate with online clients. The experiment's primary goal is to gather statistically significant evidence for the following hypotheses: **a)** Graphic design, third-party endorsements, and small-print legal notices are significant trust factors. **b)** In email messages, the impact of these factors depend on its semantic content. **c)** Overemphasizing security concerns can diminish a web site's appearance of authenticity. **d)** Semantic alignment between domain name and web page content improve the appearance of authenticity. **e)** Third-party program administration exposes a potentially devastating new phishing attack.

The experiments show that sophisticated graphic layout, third party endorsements, and small print footers improve trust in email. They further suggest that this improvement is independent of the underlying authenticity – that they raise confidence in phishing and authentic messages – *and* that the improvement is constrained by the *narrative strength* of the message. This was an unexpected result that emerged from close examination of the different contexts in which the design features were effective. We loosely characterize narrative strength as how well the message text itself connects to the reader. Factors that increase strength include: telling the reader something surprising, connecting the message to well publicized events, and informing the reader about an imminent change in her routine. However, excessive message length can shift attention away from the narrative to the design features.

The narrative content of a web login page is highly limited and in a phishing attack its expectations are established by the email message that leads the user to the page. Testing shows that URLs that are semantically well aligned with page content improve authenticity evaluations. Moreover, deceptively constructed URLs – such as `www-wellsfargo.com` – that align well with page content can still elicit high authenticity ratings in spite of their unusual appearances. Third party endorsement graphics were also found to improve authenticity ratings when paired with deceptive domains. SSL appears to have some impact on authenticity perceptions, but appears to lag well behind URL alignment. Finally, the test shows that overemphasizing concerns about security and associated vulnerabilities can significantly reduce authenticity perception.

2 Related Research

Phishing studies. There have been a few recent studies that focus on the mechanics of phishing. Dhamija, Tygar, and Hearst conduct a detailed user study on a group of 22 subjects [7] that assesses their strategies for recognizing phishing web pages. They identify 5 broad classes deployed by their subjects: a) web page content only analysis, b) the previous method plus address bar analysis URL, c) the previous method plus use of `https`, d) the previous method + browser bar padlock, and e) all previous methods plus SSL certificate analysis.

Jakobsson et al. qualitatively analyze reaction to various trust indicators in email and web pages [15]. Subjects interacted with the stimuli using actual web browsers and email clients, and were asked to *think-aloud* their reactions as they made authenticity judgments. The each survey was followed by an exit interview where subjects summarized the stimuli features that provoked confidence or mistrust. Their subject observations and interview suggested several patterns, some of which informed this experiment’s stimulus design: a) Layout sophistication can improve authenticity perception. b) Legal disclaimers, copyright notices, and small print improved confidence. c) Too much emphasis on security can backfire. d) People use URLs extensively in their trust evaluation. e) Third party endorsements vary significantly by brand: no subject expressed more confidence as a result of “BBBOnline” and “TRUSTe” certifications, while most subjects cited “Verisign secured” by name as improving trust.

Other studies have examined success rates of phishing email messages. Jagatic et al. performed an experiment which sent (benign) phishing messages to a large population of university students [13]. The researchers manipulated the messages to appear to come from friends of the subjects using data they had mined from online social networks. This highly contextualized message produced an 80% victimization rate while their control group (which did not appear to be from a friend) received a 16% disclosure rate. These messages were not subject to spam filtering.

Jakobsson and Ratkiewicz sent “phishing messages” posing as potential buyers to members of an online marketplace [14]. The “reply now” links were modified to look “phishy” (e.g., by using escape characters and naked IP addresses in the URL) yet still lead the user to the authentic login page for the message reply system. Sellers responded using the phishy looking links at a rates between 4% and 14%; these users were considered to be likely victims of a real phishing attack. Some of the messages utilized extra context by greeting the user by name; this did *not* have a statistically significant impact on response rates.

Wu, Miller and Garfinkel perform a phishing study to assess the effectiveness of security toolbars [21]. One of their important contributions is their task design for subjects. Each subject has to respond to 20

email messages directed at a fictional “John Smith” character – where five of the messages simulate phishing attacks. This methodology was designed to keep security from becoming the subject’s primary goal; users are principally concerned with accomplishing tasks. They found that when the email content looked good (e.g. no spelling errors and low-key requests) that spoof rates still ranged from 33% to 45% with the use of toolbars. The same study does a follow-up on warning methods to find that blocking pop-up warnings are more effective in alerting users to suspicious sites; however 70% of them still fell victim to at least one of the attacks.

Downs, Holbrook and Cranor [8] take a similar approach to studying how risk familiarity informs phishing defense strategies employed by end-users. The researchers furnish 20 subjects with fictional identities to role-play email and web interactions while they are closely observed, interviewed and recorded. They find that when users are aware of a particular scam tactic – e.g., advance fee scams – they are reasonably adept at protecting themselves, however they find that their subjects do not possess a sufficiently accurate mental model of Internet computer interaction to adequately protect themselves from more sophisticated or previously unseen scams.

Trust factors in non-phishing websites. Several studies have examined factors that improve credibility in web sites that do not masquerade as other entities. A study by Edelman [10] examines the trustworthiness of web sites that have *earned* third-party trust certifications. The study finds that *TRUSTe* certified web sites are twice as likely to be “untrustworthy” (according to the author’s SiteAdvisor metric) than randomly selected websites. *BBBOnline* certified web sites are more trustworthy than randomly chosen web sites according to the report, however to date only 631 certificates have been issued. This result supports the subject ambivalence observed in [15]; there is no actual gain in trustworthiness from *TRUSTe*, while the *BBBOnline* brand is not sufficiently widespread to have an impact.

Photographs of human faces have also been speculated to improve trust in online media. Riegelsberger and Sasse have conducted studies on the impact of faces on trust in online shopping web sites [18, 19]. The first study [18] uses the qualitative *think-aloud* observation methodology determine how different kinds of photographs change the trust perceptions of a web page. Their subjects had wide variances in their reactions: some preferred a minimal interface with no photographs, some only approved of photographs that furthered the brand image, others thought that pictures of staff were a deliberate attempt to manipulate trust. They joined McCarthy to follow up this experiment with a large scale study on 115 subjects [19]. Subjects played a game where they chose an amount of money to risk on each web site; they received a reward based on their success in discriminating “good” vendors from “bad” vendors. There were 8 different photos paired with 12 different web sites for a selection of 96 variants. They found no simple rules describing what photos to put on an e-commerce site to increase trust. Generally, they found that photos decreased subject ability to distinguish between “good” and “bad” vendors; photos tended to bolster trust for “bad” vendors but had indeterminate effects for “good” vendors.

Fogg et al. have done large user studies on factors that improve the web site credibility [11]. They collected the comments and ratings of over 2500 people for 100 websites spanning 10 different categories. They found that the largest factor in credibility was *design look* followed by *information design/structure*. All of the other factors can be observed inside the browser’s content window as well. These results foreshadow the effectiveness of content spoofing (i.e. phishing) as a plausible scam.

Camp, McGrath, and Genkina [2] study the impact of betrayal causes upon trust for Internet users. They find that users are more forgiving of information loss when it can be attributed incompetence rather than deception. Their experiment also concludes that Internet users do *not* discriminate between specific hosts when assigning trust – an answer to one of three trust conjectures posed by [3]. This work suggests the opposite conclusion, since changing a web page’s URL can significantly alter authenticity perceptions.

3 Experimental Design

This experiment tests the effects of several media features – sometimes in multiple contexts – on an individual’s evaluation of its phishing likelihood. This experiment shows six email screenshots followed by six

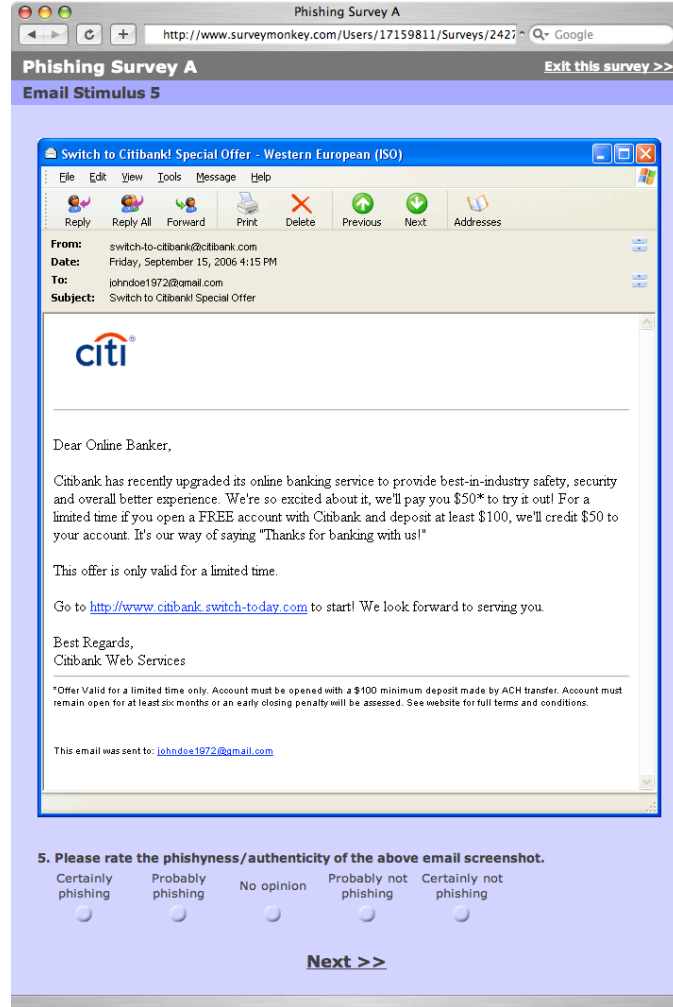


Figure 1: Subjects evaluate authenticity based on screenshots using a five point scale. The survey required a judgment before proceeding to the next stimulus.

web page screenshots to subjects, and asks them to rate its authenticity or phishiness on a five point scale: *Certainly phishing*, *Probably phishing*, *No opinion*, *Probably not phishing*, and *Certainly not phishing* (see Figure 1 for an example). The experiment was administered through *SurveyMonkey.com* [20], an online web survey service. Subjects were required to rate each screenshot before advancing to the next stimulus. The survey provided the following instructions to subjects:

Phishing is a form of Internet fraud which spoofs emails and web pages to trick people into disclosing sensitive information. When an email or web page fraudulently represents itself, we classify it as phishing.

This survey displays a sequence of email and website screenshots. Assume that your name is John Doe, and that your email address is johndoe1972@gmail.com. Please rate each screenshot's authenticity using the five point scale: Certainly phishing, Probably phishing, No opinion, Probably not phishing, Certainly not phishing.

Subjects were recruited from an introductory non-CS major class on computer usage and literacy. Through an agreement with the course's instructor, subjects were offered the opportunity to earn extra

credit through completion of this online survey or submission of a 250 word essay on the role of user studies in the software development life-cycle. Participation was completely voluntary and no penalties were given for early withdrawal or refusal to participate. The experiment was co-deployed with another study, the *Sustainability and Information Technologies Survey* which collected demographic data (in addition to data of its own concern) for both sets of questions. Out of a class size exceeding 600 students, 435 began this study. All but twelve subjects were between 17 and 22 years old with a gender split of 40.0% male to 57.9% female (2.0% did not respond). The socio-economic background of subjects was rather well off with 29.7% having a family income exceeding \$100k per year, 39.1% between \$50k and \$100k per year, 21.6% between \$30k and \$50k per year, 5.7% between \$20k and \$30k per year, 3.2 below \$20k per year, and 0.7% declining to answer.

Subjects were randomly chosen to evaluate one of two sets of stimuli. For ten of the twelve questions, the screenshots in the two versions differ by a set of test features. Our primary analysis compares the impact of the feature change between the two versions. In the other two question sets, subjects evaluate the authenticity of messages and web pages under third party administration – a potential vector for social engineering. We further designed the test to simulate a roughly equal number of authentic and phishing stimuli to avoid effective use of a trivial rating strategy. If there were significantly more phishing stimuli than authentic stimuli, subjects could employ an “always phishing” strategy that would correctly evaluate most of the stimuli without exercising due consideration.

Since the stimuli are only screenshots, their inauthentic features were designed to be evident upon examination (rather than mouse-over or source analysis). For instance, incorrect domains are apparent in email hyperlinks; they are not disguised by an inconsistent `href` attribute. The domains we chose to simulate inauthentic URLs are not in use today, however some are owned by their respective companies, others are owned by unrelated companies, while the rest appear to be unregistered according to the `whois` database. Our use of these domains as representations of inauthentic URLs is still valid because none of these URLs exist with the content we present. We outline the stimuli, their relevant features, and what we hope to learn below:

Authentic payment notification - plain vs. fancy layout

These two email messages use actual payment notification text from Chase Bank. The text personalizes its greeting and references a recent payment transaction; there are no hyperlinks. One version uses the original layout – a one color header containing the company logo followed by the message text – while the other version uses an enhanced layout – a header that includes a continuous tone shiny logo and a photograph of a satisfied customer, a smooth gradient footer graphic that spans the page with a gentle concave arch, hyperlinks to “Privacy” and “Terms of Use,” and a copyright notice; these graphics were adapted from the web page at <http://www.bankone.com>.

Strong phishing message - plain vs. fancy layout

We constructed the phishing message text to sound as plausible as possible. Opening with a personalized greeting, the message explains that former Bank One customers will need to register for Chase’s “ePIN” program – a replacement for ATM PINs that is also bundled with a new “eDebit” online service. It implicitly threatens service discontinuation by supporting “legacy 4 digit PINs for the rest of the calendar year.” The bogus registration hyperlink uses the made-up URL, <https://www.chase.ePIN-simplicity.com>. The message closes with a bogus phone number to call for assistance. The two versions of this message use the same plain and fancy layout schema described above with one exception: the fancy layout adds shiny letters proclaiming “Bank One is now Chase,” between the header and message text.

Authentic promotion - effect of small print footer

This is an authentic message from AT&T Universal card that promotes their paperless billing system. It personalizes the greeting and includes the last four digit of the account number. There are multiple

company logos, a blue outline around text, an “Email Security Zone” box, and a small print footer filled with trademark, copyright, and contact notices, as well as various informational and administrative hyperlinks. The principal login hyperlink conceals its destination. The test pair consists of the original message and a modified version that excludes the small print footer.

Weak phishing message - effect of endorsement logo

This is a phishing message promising \$50 for opening an account with Citibank. There is a simple company logo in the header, while a footer contains legal disclaimers about the offer. There is no personalization and the lone hyperlink is a made-up domain (actually owned by an unrelated organization), <http://www.citibank.switch-today.com>. Figure 1 contains its screenshot. The two versions differ only by the presence of a center-aligned “Verisign Verified” endorsement logo that follows the footer’s legal disclaimers.

Authentic message - effect of endorsement logo







The test determines the impact of a “Verisign Verified” logo added to the footer of an authentic message. The notice begins with a personalized greeting and informs the client about changes in PayPal’s logo insertion policy. The message body is considerably longer than all of the other messages except for the Netflix stimulus (described below). The primary message contains no hyperlinks, but their small print footer furnishes a hyperlink to unsubscribe from their newsletter. One interesting feature of the message is a boldfaced statement: “If you do not wish to have PayPal automatically inserted in your listings, you must update your preferences by 9/25.” Though genuine, this message parallels the account shutdown threats brandished by many phishing messages. The header contains a monochrome company logo and a two-tone horizontal separation bar.

Login page - strong and weak URL alignment

This browser’s content window displays an exact copy of the AT&T Universal Card login page. Like most web login pages, it displays a high level of layout sophistication: photographs of happy clients, navigation bars, product pictures, side bar, promotional windows, and small print legal disclaimers. They also display a “Verisign Secure Site” endorsement logo. The two versions of this stimulus differ by their address bar contents: version *a* uses <https://www.accountonline.com/View?docId=Index&siteId=AC&langId=EN> – and consequently displays a browser frame padlock – while version *b* uses the unencrypted URL, <http://www.attuniversalcard.com/> (owned by AT&T but not in use).

Login page - strong and weak content alignment

This next set takes the alternative approach to aligning the address bar URL with content: change the content. Both pages use the unregistered URL, <http://www.citicardmembers.com/>. Version *a* displays a precise copy of the authentic Citi Credit Cards login page in its content window, while version *b*’s content window displays modified logos and links for better alignment with the URL:

Location	Original	Modified
Header		
Header Menu	 citicards.com	 CitiCardMembers.com
Footer	 Citi.com	 CitiCardMembers.com

The page used a sophisticated layout with nearly all the same identifiable features of the AT&T Universal Card login: photographs of happy clients, navigation bars, product pictures, side bars, promotional windows, and “Verisign Verified” logo.

Login page - authentic and bogus (but plausible) URL

These two stimuli test the impact of changing a well aligned authentic URL, <http://www.paypal.com/ebay>, to a reasonably well aligned bogus URL, <http://www.ebaygroup.com/paypal> (domain owned by eBay, but not in use). The main content window is the same for both: the eBay decorated version of the PayPal login page which contains an eBay logo to the lower right of the primary PayPal logo. The page layout contains all the main features of the previous login pages, but includes a more thorough set of third party endorsement logos: “Verisign Verified,” “Reviewed by TRUST-e,” and “Privacy: BBB *OnLine*.” SSL is not used in either stimulus.

Login page - hard and soft emphasis on security

This pair tests if it is possible to undermine confidence in an authentic login page with excessive concern about security and online fraud. These two stimuli represent an extreme – but real world – case. Clients of the Indiana University Employees Federal Credit Union (IUCU) were targeted by a phishing attack in early August 2006. In response, the credit union altered their web page to include a large yellow banner that read:

WARNING: ALL IU CREDIT UNION MEMBERS PLEASE READ! Phishing scam in progress;
click here for important information

They further augmented the news section with a similar message: “Warning! Phishing Scam in progress – learn more.” Finally, a section named “Critical Fraud Alerts” contained the exact same warning as the one from the news section. The twin page eliminates all phishing warnings – including the banner – and changes the “Critical Fraud Alerts” section heading to read “Fraud Prevention Center.” Generally, the language was changed to sound “in control” rather than alarmist.

Bad URL - with and without SSL and endorsement logo

Can an endorsement logo and SSL padlock overcome a bad domain name? This next set tests these two features on a Wells Fargo phishing site based on the bogus domain, www-wellsfargo.com. The login page is similar in layout to the others, but does not feature photographs of people. The only continuous tone graphic is an image of a speeding horse-drawn carriage that evokes a wild-west money transfer service. One screenshot contains the original page content using an unencrypted connection, while the other stimulus uses SSL and adds a center-aligned “Verisign Verified” logo to the page’s footer.

High profile recall notice

At the time of testing (10/02/2006-10/12/2006), the press had been alive with recent reports [6] of laptop battery recalls from both Dell and Apple. Sony, the source for the batteries, ultimately issued a direct recall for the same batteries affecting several more brands on 10/23/2006.

This test set does not follow the controlled pair format of the previous stimuli. One stimulus is a screenshot of the official “Dell Battery Return Program” web page. Web page layout is substantially simpler than all other web stimuli. A four color header logo adorns the top of the page. They present the content as a letter to “Dell Customer” which explains the danger and how to determine eligibility for exchange. Notably, there is a single column of content, no photos, and no promotional content of any kind. They use the third-party domain, dellbatteryprogram.com. Use of this domain for the official page makes their replacement service ripe for phishing.

We constructed a phishing email message using the header, footer, and textual content from this web page. The phishing message omits the middle section on how to identify eligible batteries and instead requests that the receiver go to the bogus web page at <http://www.dellbatteryreplacements.com>.

Low profile class action suit

This last set of stimuli also follows the third-party email and web page form of the previous set. Both stimuli are authentic, but use altered dates to appear relevant at the time of testing. The email message is a lengthy notice which describes a class action lawsuit against Netflix, a settlement to the lawsuit, and options for claiming benefits. There is no greeting, signature, color, or graphics. The only hyperlinks direct the user to the authentic third-party URL, <http://www.netflixsettlement.com>. The web page has a similarly bare appearance, but with much less text. It behaves as a hyperlink gateway for more information under the URL, <http://www.netflix.com/settlement/> – the result of redirection from <http://www.netflixsettlement.com>.

4 Results and Analysis

Stimulus Description	Mean	Diff.	χ^2	p
Chase card payment statement (legit) - plain layout	3.40		11.89	0.018
Chase card payment statement (legit) - fancy layout	3.76	0.36		
Chase phish - fancy layout	3.19	0.02	6.31	0.177
Chase phish - plain layout	3.18			
AT&T Universal Card statement w/o legal notices	3.05		30.18	0.000
AT&T Universal Card statement w/ legal notices	3.66	0.62		
PayPal policy change + Verisign	3.19		5.75	0.219
PayPal policy change - no Verisign	3.30	0.11		
Citibank phish - no Verisign	2.40		9.62	0.047
Citibank phish + Verisign	2.69	0.29		
AT&T Universal Card login https://www.accountonline.com/View?docId=Index&siteId=AC&langId=EN http://www.attuniversalcard.com	2.76 3.25	0.49	15.46	0.004
Citibank (phish); URL = http://www.citicardmembers.com/ Copy of original site Logos modified to better match domain	3.11 3.43	0.32	7.06	0.133
PayPal website displaying eBay logo URL = http://www.ebaygroup.com/paypal/ URL = http://www.paypal.com/ebay/	3.35 3.70	0.35	8.83	0.065
Indiana University Credit Union home page De-emphasized security language; no mention of “attacks” Phishing attack banner + strong fraud warnings	3.69 3.37	0.32	11.45	0.022
Wells Fargo phishing page Reproduces original content; URL= http://www-wellsfargo.com/ Adds Verisign endorsement; uses SSL; URL= https://www-wellsfargo.com/	3.17 3.48	0.31	10.83	0.029
NetFlix class action settlement email (authentic)	2.72			
NetFlix class action settlement home page (authentic)	2.55			
Dell battery replacement email (phishing)	3.61			
Dell battery replacement web page (authentic)	3.54			

Figure 2: The first section of the table reports on the differences between email messages, the next section reports on the web pages, and the last section gives the average rating for the third party attacks.

Our experiment directly controls for the effect of several design features. There are some surprises in the direct results from these tests – including the stunning impact of a detailed small print footer on an otherwise well-conceived legitimate message – however the experiment reveals an unexpected, but in retrospect obvious, lesson about email messages: the “story” of the message is critical. Messages with strong and succinct narrative components rated highly *and* their ratings appear to be less susceptible changes in graphic design. On the other hand, authenticity perception changed significantly for messages that say little

(such as a service promotion) under document feature variances. Two of the five sets of “twins” did not change significantly, according to the χ^2 metric, when augmented with the very same features that produced significant changes in other messages. Subjects judged these messages principally on their narrative content.

The Chase phishing message uses their recent acquisition of Bank One as a pretext for imminent service change to ATM card authentication; the story further bundles this change with the addition of new service, *eDebit*, and implicitly threatens discontinuation of service by claiming to “support the legacy 4 digit PIN for the remainder of the year.” This consequence is much less direct than the standard “Your account will be suspended in 48 hours if ...” strategy used by many phishers. It works on a less urgent time scale and is pitched as a convenience to all clients, rather than an anomaly specialized to a specific client. There was no significant difference in subject evaluation between the plainly formatted version of this message with the monochrome corporate header and the one with several customized continuous tone graphics.

Yet these same graphics, minus the shiny “Bank One is now Chase” banner, produced a significant change ($p = 0.018$) in evaluations of Chase’s authentic payment notice. The payment notice thanks the receiver for a recent online payment, and then goes on to inform the user about features of their online account management interface. In terms of relevance, there is less potential impact on the user. Ignoring this message won’t expose the client to any changes (good or bad). Assuming that the name and recent online interaction are correct, the message communicates little that would surprise the average client. So in place of strong narrative components, the subject looks to formatting cues to further inform confidence. This message rated highly in simple form and was pushed higher by the improved graphics. We attribute its initially high rating to its exclusion of hyperlinks, informative nature, and well contextualized message.

Subject reactions to the presence of the “Verisign Verified” logo differed dramatically between the two test messages. One message, a change of policy notice from PayPal, experienced no statistical difference in subject evaluations of its endorsed and unendorsed forms. The policy change notice shares several narrative features with the Chase phishing message: both messages have a customized salutation, both inform the users about an institution wide change (in this case the particulars of their logo insertion policy), and both claim that inaction will result in a change of service. The PayPal message has no hyperlink in the message body, but *does* contain a link in its small-font gray footer to manage user preferences; we think that this forecasts a potential phishing strategy, described in Section 5.

The other message that tests the impact of the “Verisign Verified” logo is a phishing message that exploits the Citibank brand. The experiment shows a statistically significant change in subject evaluations ($p = 0.047$) due to this single feature change. Of all the messages, this message makes the weakest connections to the receiver. It begins with a generic salutation. Worse, the first sentence promotes the goodness of their online service but fails to involve the receiver in any way. Not until the second sentence does message’s relevance become evident to the reader – they are offering “\$50* to try it out!” These two messages had the lowest average ranking of all the email stimuli. Ignoring this message has no impact on the user except for failing to miss out on an offer of dubious value. Ultimately, the stimulus fails to engage the reader, and so subjects base more of their evaluation on non-narrative factors – such as the endorsement logo and the bogus URL.

The Dell battery replacement program message presents compelling story, but not directly. The incident received a high level of media coverage due to spectacular reports of exploding and burning laptop computers. The Dell message, which we manufactured, benefits from other sources spreading the story. Without this third party validation this message could have bordered on implausibility, but instead our subjects produced ratings that were statistically indistinguishable from the two most highly ranked email messages in the batch. This message contains a slightly nicer than average layout (multi colored header, footer graphic with links), but less personalization and an fraudulent (but semantically plausible) URL. The story was so powerful and present in the subject’s minds that they were willing to discount the suspicious link and generic greeting.

The other third party attack did not benefit from recent or high profile media coverage. We may have further lowered its rating by altering the dates to appear relevant at the time of testing. Subjects could have perceived the timeline as implausibly long (even for legal action), or may have been familiar with the case and known that the dates were incorrect. In addition to these changes, the original message is particularly poorly conceived. While it has strong narrative elements – which present lawsuit context, the elements of the settlement, and response options – the message is entirely too long. Message length and

detail create an incentive for users to quickly evaluate according to non-narrative features. The most visually obvious features are the inclusion of blue hyperlinks – the only non-black and white symbols – that point to www.netflixsettlement.com. While this *is* the legitimate domain, it should raise suspicion because it is an apparent “cousin domain” to the parent company’s web site. The lack of strong design features seals its poor evaluation. There is no company header – a feature present on every other stimulus – no opening salutation and no signature. The contact address appears to be an afterthought which does not even specify a division of the company, let alone an appropriate administrator.

The biggest surprise of the test appears in a pair of authentic AT&T Universal Card messages. In many ways, this is the polar opposite of the Netflix settlements message: it has strong design elements and a short, weak narrative. The message promotes their *Statements Online Only* program without bundling it with a recent action (as Chase does with their payment notification). Ignoring this message will not produce any change in the receiver’s service nor does enrolling in their program provide any obvious benefit to their client; in fact, enrollment could result in unintended late payments due to imperfect spam interdiction of electronic billing notices. What the message lacks in narrative appeal, it makes up for in design strength. It customizes the message to the receiver both in the opening salutation *and* in a “Email Security Zone” header box that displays client name and client account number suffix. The header also a spam awareness message and company logo. A blue outline which complements the company logo encloses the rest of the content. The corporate logo appears a second time within the blue content box and the letter opens and closes with personalized salutations: “Dear John Doe,” and “Sincerely, Julie A. Garry.” The two versions of this message differ by the presence of a detailed small print footer below the signature which contains hyperlinks to privacy and security policies, as well as hyperlinks bearing the universalcard.com domain. The footer uses a small gray font and presents text for adjusting “Email Preferences,” and a “Help / Contact Us” section containing their postal address and various trademark and copyright notices.

The one ambiguous feature of this email is a centrally located hyperlink, labeled “log in to Account Online”. It does not indicate the URL in the text. Phishers frequently employ this sort of hyperlinking strategy to conceal the bogus server’s URL. The footer may add confidence because its hyperlinks appear to reference URLs with legitimate and semantically-aligned domain names; none of the hyperlinks outside the header indicate a target domain. Alternatively, the contact, copyright and trademark notices themselves may improve confidence in the message. It is particularly interesting that even though the footer-less message displayed the last four digits of the credit card number, customized the greeting, and employed generally strong design elements that, except for the NetFlix settlement, it was still ranked lower than any other legitimate message. This supports the experimental results in [14] that indicate indifference to customized greetings in certain contexts. Yet adding the footer boosts the evaluations to the point where it is statistically indistinguishable from the other two most highly ranked messages – the Chase payment notification with fancy graphics and the Dell battery recall notice.

Web sites, particularly the login and information collection pages associated with phishing scams, do not present a story the way email messages do. Because of this, their credibility depends much more on document features and graphic design. Subjects assigned significantly different ratings to three of the five sets of twin stimuli. The results show that address bar alignment with page content, overwrought concerns about fraud, and third-party endorsements substantively change authenticity assessments.

The biggest rating difference among the web page stimuli was measured between the two versions of the “AT&T Universal Card Sign-on” page. The official version, which uses the address URL <https://www.accountonline.com/View?docId=Index&siteId=AC&langId=EN>, has the lowest average rating of the 10 web stimuli in the paired testing. With an average rating of 2.76 it rated lower than a simulated phishing website based on the suspiciously formed domain, www-wellsfargo.com (avg. rating 3.17). Subjects who saw the unused domain <http://www.attuniversalcard.com> in the address bar of the AT&T card login page rated it significantly higher ($p = 0.004$) than the authentic page. The page content strongly aligns with the URL text <http://www.attuniversalcard.com>: the phrase “AT&T Universal” appears no less than seven times on the login page, while the phrase “AT&T Universal Card” appears four times in the content window. Interestingly, the official page uses [https](https://) and displays an SSL padlock in the lower right-hand browser frame, while the [attuniversalcard.com](http://www.attuniversalcard.com) domain does *not* use SSL – and consequently does

not display the padlock on the browser frame. Subjects found the semantic alignment of the URL to be a much stronger indicator of authenticity than SSL utilization. In fairness, subjects could not examine the certificates nor would they have been subject to the “Unable to verify the identity of host” pop-up window in the case of a self-signed certificate. Nevertheless, other user studies have found that in practice subjects rarely consider these factors [7, 21].

Much to our surprise, the phishing simulation based on the URL <http://www-wellsfargo.com> rated significantly higher ($p = 0.00001$) than the official AT&T Universal Card login page. Subjects valued semantic alignment between content and host domain more than domain well-formedness. Syntactically there is nothing wrong with the domain, but replacing the dot with a dash is clearly an attempt at deception. Adding SSL and a “Verisign Verified” logo to the Wells-Fargo phishing page produced a significant improvement in authenticity ratings ($p = 0.029$). It’s worth noting that the authentic login page (not in the test) does *not* display a Verisign logo, but does use SSL. In spite of subject either failing to notice the dash-for-dot exchange or not thinking that it was suspicious, they did notice the presence of either SSL or the Verisign logo. Note that the AT&T login page also used SSL and displayed a Verisign endorsement, but neither of these features could overcome the mistrust of the accountonline.com domain.

The last statistically significant difference between twins ($p = 0.022$) shows that too much concern about security can reduce customer confidence. Subjects responded positively to use of less fearful language and rated the softer, more constructive content significantly higher than the page that displayed stark warnings. Note that correct domain and SSL were used on both stimuli. This is a case where a good faith effort to educate clients about phishing undermines confidence in the web site’s authenticity. Login pages are no place for fear-provoking messages.

One way phishers align page content with URLs is by choosing an apt domain name; the other way is to change the page content. While subjects gave a higher average rating to our modified Citibank cardmembers login page, the χ^2 test showed that the two distributions were not significantly different ($p = 0.133$).

The last test pair was nearly significant ($p = 0.65$), but does not confirm that the two ratings come from different distributions. This pair compared the effects of a plausible parent company domain and subsidiary subdirectory, <http://www.ebaygroup.com/paypal/>, with the authentic URL that reverses their positions, <http://www.paypal.com/ebay/>. Although the [ebaygroup.com](http://www.ebaygroup.com) domain is unbound, eBay has registered it. Nevertheless, this test shows a certain flexibility in user acceptance of domain alignment. No PayPal client has seen the PayPal page displayed under the <http://www.ebaygroup.com/paypal/> address, yet their authenticity ratings are not significantly different. This result furthers our conviction that semantic alignment between content and URL is a principal factor in authenticity evaluations.

The last two web stimuli are not twins; they are the Dell battery replacement page and the Netflix settlement page. Both pages are authentic, although the content of the Netflix page was altered to appear relevant at the time of testing. They received polar opposite ratings. The Dell battery page was statistically indistinguishable from the highest rated page (the authentic PayPal site), and the Netflix page rated dead last – significantly lower than the second lowest rating ($p = 1.20 \cdot 10^{-7}$). As mentioned before, the Dell battery program stimuli benefit from a high visibility news story. It’s noteworthy that the URL in the web page (authentic) is different from the URL in the email (phishing) – which subjects saw first. Subjects did not penalize the web page for this inconsistency. Subjects may have had difficulty constructing a phishing scenario based on the informational nature of the page; there is no request for personal information.

Similarly, the Netflix settlement page does not make any overtures for personal information. Even more surprising is that the URL, <http://www.netflix.com/settlement/> aligns well with the content. Subjects may have dismissed the page based on mistrust of the email stimulus which they viewed prior to (several screenshots before) the web stimulus. The Netflix page is notable for its brevity and unsophisticated layout. It is the only page without graphics or logos of any kind. There are no apparent links back to the primary Netflix page. With the exception of the blue underlined hyperlinks and gray margins, the page is black and white. We take from this rating that utilizing minimalist design is a poor strategy for unsolicited communications – even for important and serious matters such as law.

5 A Forecast of New Phishing Tactics

High volume phishers have access to large amounts of very relevant data. Assuming that our experiment approximates what they know, we offer some predictions on how future attacks could look. Future phishing attacks could use the following strategies:

1. Construct messages with weak narratives (bordering on innocuous), but use strong design elements – graphics, small print footer, endorsement logos – and identifying information to improve authenticity impressions.
2. Use softer bait. Messages that do not encapsulate an imminent request for information, such as the Dell battery bait rated highly in the test.
3. Use plausibly unfamiliar administration pages; e.g the Dell Battery Return Program web site provides a service that is not typically seen – like a login page – so visitor expectations are less concrete.
4. Leverage high profile news to produce messages with credible and strong narratives. Personalization will be less important in these cases.
5. Align domain names with page content. Although subjects were turned off by semantic mismatches between domain names, they were insensitive to malformed links, e.g. `http://www-wellsfargo.com`.

Recent phishing messages have used deceptive subdomains to circumvent the limitations imposed by domain name registrars.

```
http://www.paypal.com.cgi.bin.account.webscr.cmd.login.run.php.draciimasi.info/webscr.php?cmd>Login
```

The above URL comes from an actual phishing attack. Since subjects accepted the substitution of a dash for a dot in `http://www-wellsfargo.com`, they could easily accept a dot for a slash as above. Moreover, the preponderance of subdirectory names like `cgi`, `bin`, `webscr`, etc. further clouds the issue for the technically uninformed.

Example: the “Unsubscribe” spam attack

This attack combines first three strategies from above: weak narrative combined with good design, softer bait, and plausibly unfamiliar information collection pages. Some of the most highly rated email messages avoided hyperlinks in the main message text. The Chase account payment was completely devoid of hyperlinks and instead directed receivers to type `www.chase.com` into their address bar. Similarly, the PayPal message had no hyperlinks in its body, *however* it included a hyperlink in the footer to change preferences. The highly rated AT&T Universal Card promotion also contains links in its small print footer.

The phisher will send out promotional email that appears to come from the spoofed institution. The promotion would employ a weak narrative to shift user attention to a plethora of design features, e.g. graphical header, footer, small print, personalization, genuine (but unlinked) URLs in the body, etc. The body will generate the perception of authenticity by referring the receivers to the phone number on back of their credit card or by requiring users to manually type in the promotional URL. Among the design features is a small print footer with an *unsubscribe* hyperlink. This link will take the user to an unfamiliar web page (as many unsubscribe pages appear to live in the bowels of a web site) that requires “confirmation” of this change in the form of a username and password. The phisher could further enhance the illusion with a “confirm” button that causes a simulated username and password dialogue box to appear via JavaScript. The confirmation would be contingent on entering the credentials.

The bait message gets the user to click on the link indirectly: annoyance with the volume of unsolicited messages. Their web page would cause the user to invest effort by de-selecting several different promotional message categories. This invested effort will beget grudging compliance when confirmation turns into a query for authentication credentials.

The strong narrative attack

The strong narrative attack engages the receiver with plausible story, often bundling actions to well known news stories. The Chase phishing message that promotes “ePIN” to incoming Bank One customers is such an example; it leverages in the news of the Bank One acquisition. The Dell battery program stimuli gain most of their credibility from the story’s media coverage. The message maintains this credibility by deferring the request for personal information; standard attacks request an “account login” or “settings update” in the message body.

While scams that exploit strong narratives and current events are not new (e.g. many fraud cases capitalized on the September 11th and Hurricane Katrina tragedies [17, 16]), our research suggests that they are less influenced by design features. This finding is supported by the persistence of the Nigerian code 419 advance fee scams [22, 9]. One widespread form of this attack entices victims with a story of the death of a foreign dignitary and the need to move large amounts of money (allegedly to protect it from corrupt enemies); they offer the victim a cut for moving the money. After drawing the victim into this illusion, they request advance fees to enable the transfer of money. These messages break many design rules that promote trust: they use poor spelling and grammar, email messages are often plain text, return addresses are essentially anonymous using free email accounts. Yet these scams still account for large amounts of Internet fraud – exceeding \$3 billion in losses according to some estimates [12].

6 Conclusion

This study tested the impact of several document features on user authenticity perceptions for both email messages and web pages. The influence of these features was context dependent in email messages. We were surprised that this context was shaped more by a message’s narrative strength, rather than its underlying authenticity. Third party endorsements and glossy graphics proved to be effective authenticity stimulators when message content was short and unsurprising. The same document features failed to influence authenticity judgments in a significant way when applied to more involving messages. Most surprising was the huge increase in trust caused by a small print footer in a message that already exhibited strong personalization with its greeting and presentation of a four-digit account number suffix.

The data suggest a link between narrative strength and susceptibility to trust-improving document features, however the experiment was not designed to test this hypothesis. Future work should characterize more precisely what kind of messages can benefit from these features, and what kind of messages are resistant to their sway.

Since phishing web page content need not differ from the authentic pages, we focused three web page tests on the effects of semantic alignment between address bar URLs and page content. The first showed a clear statistical preference for a simulated web page whose domain name matched its content rather than the genuine page whose domain was only weakly aligned with the same content. The second test, which created better alignment with a bogus domain name by altering company logos, failed to register a statistically significant change in authenticity ratings. The third test compared an authentic page (and URL) with an authentic version of the same page content paired with a well aligned – but bogus – URL; the results which favored the genuine URL were just shy of statistical significance. In conclusion, we find that URL *can* change authenticity ratings. This result is one answer to the conjecture (in [3]) that “*Over time and with experience users will tend toward greater discernment among distinct remote computers.*” Domain names indicate interaction with specific hosts and our tests have shown that users are sensitive to this factor in their authenticity judgments.

This experiment also verified that it is possible to overuse well-intended notices about security and fraud. We observed a statistically significant *negative* effect of genuine, but heavy-handed, fraud warnings. Another test showed a statistically significant improvement in authenticity perception when using SSL and a third-party endorsement logo on a fraudulent web page showing a suspiciously formed, but semantically well-aligned, domain name.

The experiment simulated two sequences (one email and one web page) that appeared to be third-

parties charged with handling embarrassing incidents for their corporate clients. Though separated by many variables, one turned out to be among the most trusted stimuli in the test, while the other rank among the lowest. The poorly ranked one, though authentic, broke all the rules: poor publicity, long and rambling message, use of third-party domain names, and no graphics. The highly ranked one (whose bogus email message was concocted by the authors) benefited from a widely publicized recall message. The story overrode the message’s poor personalization, illegitimate URL, and relatively simple layout.

These factors offer a glimpse into what kinds of phishing strategies may be deployed in the future. We describe an *unsubscribe attack* which contains an innocuous message, many authenticity stimulating document features, and an unsubscribe link that demands login credentials prior to changing the email promotions policy. Our tests with third-party administration suggest that companies in the process of correcting an embarrassing incident are highly vulnerable to phishing attacks. Finally, our findings suggest some common pitfalls for legitimate e-commerce to avoid: overuse of fraud warnings, utilization of poorly aligned domain names, failure to use `https` for rendering login pages, and long or rambling email messages.

7 Acknowledgments

We would like to thank Charles Pope for making it possible to test such an ideal population. Also, thanks to Eli Blevis for his partnership in the human subjects review process and feedback on stimulus design. Thanks to Sukamol Srikwan and Youn-Kyung Lim for their statistical expertise and general guidance. Also, Dave Roedel and Sid Stamm deserve thanks for their technical expertise concerning experimental deployment.

References

- [1] APWG. Anti-phishing reports. <http://www.antiphishing.org/reports/>, Retrieved in December 2006.
- [2] L. J. Camp, C. McGrath, and A. Genkina. Security and morality: A tale of user deceit. In *Models of Trust for the Web (MTW’06)*, Edinburgh, Scotland, 22 May 2006.
- [3] L. J. Camp, H. Nissenbaum, and C. McGrath. Trust: A collision of paradigms. In *Financial Cryptography: 5th International Conference, FC 2001, Grand Cayman, British West Indies, February 19-22, 2001. Proceedings*, pages 91–105, Berlin Heidelberg, 2002. Springer-Verlag.
- [4] E. Cooke, F. Jahanian, and D. McPherson. The zombie roundup: Understanding, detecting, and disturbing botnets. In *Proceedings of the first Workshop on Steps to Reducing Unwanted Traffic on the Internet (STRUTI)*, pages 39–44, July 2005.
- [5] D. Darlin. Apple joins in a recall of batteries. *New York Times*, 04 December 2006. <http://select.nytimes.com/search/restricted/article?res=F00912FE3B5A0C768EDDA10894DE404482>.
- [6] D. Darlin. DELL WILL RECALL BATTERIES IN PC’S. *New York Times*, 15 August 2006. <http://select.nytimes.com/search/restricted/article?res=F10A1FF83C5A0C768DDDA10894DE404482>.
- [7] R. Dhamija, J. D. Tygar, and M. Hearst. Why phishing works. In *CHI ’06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 581–590, New York, NY, USA, 2006. ACM Press.
- [8] J. S. Downs, M. B. Holbrook, and L. F. Cranor. Decision strategies and susceptibility to phishing. In *SOUPS ’06: Proceedings of the second symposium on Usable privacy and security*, pages 79–90, New York, NY, USA, 2006. ACM Press.
- [9] M. A. Dyrud. “i brought you a good news”: An analysis of nigerian 419 letters. In *Proceedings of the 2005 Association for Business Communication Annual Convention*, 2005.

- [10] B. Edelman. Adverse selection in online “trust” certifications. In *The Fifth Workshop on the Economics of Information Security (WEIS 2006)*, Robinson College, University of Cambridge, England, 26-28 June 2006.
- [11] B. J. Fogg, C. Soohoo, D. R. Danielson, L. Marable, J. Stanford, and E. R. Tauber. How do users evaluate the credibility of web sites?: a study with over 2,500 participants. In *DUX '03: Proceedings of the 2003 conference on Designing for user experiences*, pages 1–15, New York, NY, USA, 2003. ACM Press.
- [12] U. A. G. Investigations. Advance fee fraud in 37 nations. http://www.ultrascan.nl/html/aff_37_countries.html, 25 March 2006. Retrieved December 2006.
- [13] T. Jagatic, N. Johnson, M. Jakobsson, and F. Menczer. Social phishing. *to appear in the Communications of the ACM*. Draft preprint available at <http://www.indiana.edu/~phishing/social-network-experiment/phishing-preprint.pdf>.
- [14] M. Jakobsson and J. Ratkiewicz. Designing ethical phishing experiments: a study of (rot13) ronl query features. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 513–522, New York, NY, USA, 2006. ACM Press.
- [15] M. Jakobsson, A. Tsow, A. Shah, E. Blevis, and Y.-K. Lim. What instills trust? A qualitative study of phishing. In *In submission*, 2006.
- [16] B. Krebs. Katrina phishing scams begin. *WashingtonPost.com: Security Fix*, 31 August 2005.
- [17] F. S. C. D. U. D. of Justice. Special report on possible fraud schemes. <http://www.usdoj.gov/criminal/fraud/WTCPent-SpecRpt.htm>, 27 September 2001. Retrieved December 2006.
- [18] J. Riegelsberger and M. A. Sasse. Face it - photos don’t make a web site trustworthy. In *CHI '02: CHI '02 extended abstracts on Human factors in computing systems*, pages 742–743, New York, NY, USA, 2002. ACM Press.
- [19] J. Riegelsberger, M. A. Sasse, and J. D. McCarthy. Shiny happy people building trust?: photos on e-commerce websites and consumer trust. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 121–128, New York, NY, USA, 2003. ACM Press.
- [20] SurveyMonkey.com. Surveymonkey.com - powerful tool for creating web surveys. online survey software made easy! <http://www.surveymonkey.com/>, Retrieved in December 2006.
- [21] M. Wu, R. C. Miller, and S. L. Garfinkel. Do security toolbars actually prevent phishing attacks? In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 601–610, New York, NY, USA, 2006. ACM Press.
- [22] M. Zuckoff. The perfect mark: How Massachusetts psychotherapist fell for a Nigerian e-mail scam. *The New Yorker*, 15 May 2006. http://www.newyorker.com/fact/content/articles/060515fa_fact.

A Stimulus and Hypothesis Index

This experiment shows a selection of email and web page screenshots to subjects, and asks them to rate its authenticity or phishiness on a five point scale: *Certainly phishing*, *Probably phishing*, *No opinion*, *Probably not phishing*, and *Certainly not phishing*. The experiment was administered through *SurveyMonkey.com*, an online web survey service. Subjects were required to rate each screenshot before advancing to the next stimulus. There were two sets of stimuli (an *A* and a *B* version). For ten of the twelve stimuli, the *A* and *B* versions differ only by a limited number of features. Stimulus exposure order was fixed for both groups, going from Email-1a to Email-6a and then Web-1a to Web-6a (similar for group *B*).

A.1 Email

All delivery times have been changed to occur during normal business hours and dates have been altered to appear relevant during the time of the experiment (October 2 - 12, 2006). Messages have been addressed to John Doe.

Email stimulus 1a

This is an authentic payment notification that Chase sends out to its customers after they have completed an online payment. The message uses a simple layout consisting of a white and blue header that displays the *CHASE* logo. The message body is terminated by a horizontal line, and after an additional line break there is a statement in small print This message was sent to: johndoe1972@gmail.com. There are no evident links on the page, although the domain **www.chase.com** appears as text (rather than a hyperlink) in the body. The opening salutation uses all capital letters for the name, JOHN DOE, and the closing salutation places only one line break after “Sincerely,” so that “Cardmember Services” appears on the following line.

Email stimulus 1b

This message uses the same header logo, footer, and font as the authentic notification stimulus 1a. It is a phishing message that we designed which prompts the addressee as a former Bank One customer to register for Chase’s “ePIN” that will be used for an upcoming “eDebit” program. It has a hyperlink **https://www.chase.ePIN-simplicity.com** that is isolated by preceding and succeeding line breaks. The message implicitly threatens account discontinuation by claim that legacy PINs will be supported through the rest of the year. The message provides a bogus toll free customer support number, 1-877-CHASEPC. Formatting of the opening salutation differs from 1a because only the first letter of each name is capitalized, i.e. “John Doe.” The closing salutation places an additional 3 line breaks between “Sincerely,” and “Client Services” **
** “JPMorgan Chase & Co.” This signature reflects the *From:* field, “JPMorgan Chase & Co. Client Services.” User identity will be stolen at the hyperlinked web site when the user authenticates himself and discloses private information for the registration process.

Email Stimulus 6a

This message uses the same text as phishing message in 1b, but uses a fancy header and footer that was featured at **www.bankone.com**. The header adds a “JPMorgan” logo to the right side of the large blue bar, and below the header are glossy blue letters proclaiming “Bank One is now Chase.” To the right of these letters is a large version of their octagonal corporate logo featuring a photograph of a grinning white middle-aged male in the center. The footer graphic has a gentle concave curve on the message side and uses a smooth blue gradient that changes saturation from left to right. Below it are centered links in small letters to “Privacy” and “Terms of Use”. On the next line is a centered copyright message, and the last line contains a left justified This message was sent to: johndoe1972@gmail.com, where the email address appears to be a clickable link.

Email Stimulus 6b

This uses the same layout as 6a, except that the “Bank One is now Chase,” message has been removed. The body text and line breaking are exactly the same 1a. For the purposes of this experiment, we consider this to be an authentic message.

Hypothesis

The primary comparison is between the plainly formatted message and the ornately formatted message. We expect that the ornately formatted messages will be trusted more than their plainly formatted counterpart. There are two different tests to see if the effect is context dependent. We hypothesize that the fancy graphics increase trust more when the message content is questionable, since the authentic text should already be highly rated even in the plain formatting. We will analyze a second comparison among messages of the same formatting; we expect that the authentic message will be more trusted than the phishing message given the same formatting. The following table summarizes these comparisons.

	Plain	Fancy
Authentic	1a	6b
Phishing	1b	6a

Because of our hypothesis that the fancy formatting would inspire more confidence, we placed the plain stimuli first (hence 1a and 1b), and the fancy stimuli last (6a and 6b). We did not want the fancy message to undermine the credibility of the plain one. We placed them as far apart as possible to minimize the effects of seeing two very different messages from the same institution.

Email Stimulus 2a

We constructed this phishing message using header and footer graphics as well as copy from the official Dell battery replacement website, <http://www.dellbatteryprogram.com>. This attack leverages high profile media coverage of a laptop battery recall where the affected items could explode or catch on fire [5, 6]. The header uses a four colored horizontal bar containing the Dell corporate logo. The opening salutation is not personalized: “Dear Dell Customer.” The bogus hyperlink, <http://www.dellbatteryreplacements.com> is padded from the message text by extra line breaks. The footer is a gray area that matches the width of the header graphic. Within the footer is a center-aligned copyright dates followed by another centered line with hyperlinks labeled, *Technical Support Policy*, *Privacy Policy*, *About Dell*, and *Contact Us*. The bogus website would steal information from the victim by offering a free replacement, but require a credit card number to hold a deposit should the customer fail to return the recalled item. This is an example of a *third party phishing attack*, because the authoritative website is not actually a part of the main corporate domain, www.dell.com, thus the program appears to be administered by a third party.

Email Stimulus 2b

This is an authentic message notifying Netflix customers about the terms of a class action legal settlement. We have modified the dates to appear relevant for this year – the important dates of action had already expired by the time of our experiment. The message has no opening salutation, closing salutation or signature. The message is very long compared to most email; it explains the legal context and presents the recipient with four follow up options. The message contains numerous hyperlinks to <http://www.netflixsettlement.com>. There is no graphical header in the content body, however the footer contains a limited amount of left justified small print indicating copyright dates, a corporate address and an SRC number. Below this information is the line This message was sent to [johndoe1972@gmail.com]. This could potentially be a third party attack since the authoritative domain is not part of the company’s main corporate domain, www.netflix.com.

Hypothesis

These messages represent real programs – although we constructed the Dell message for this experiment – yet the official web sites are not hosted on their respective company’s parent domain. These messages are not well suited for direct comparison because there are so many differences between them. However we expect that the media coverage of the Dell battery recall, its message layout and relative brevity will result in higher levels of trust when compared with Netflix message. Although the Netflix settlement message is authentic, we think that the relatively unknown nature of the business, lack of personalization and extensive message will cause subjects not to trust as much as the fraudulent Dell message. Since neither message’s domain matches the company’s domain, subjects will have to judge these messages on these other factors.

Email Stimulus 3b

This is an authentic message out by AT&T Universal Card that promotes its “Statements Online Only Program.” A blue border encloses main body message. The corporate logo, “John Doe”, and the last four digits of his account number are listed above the enclosed message. In the enclosed area, is another corporate logo, the main body of the promotion and an extended footer. The opening salutation is personalized, and the closing signature is signed by the name of an individual, rather than a corporate division. In the main body of the message, there is a hyperlink “log in to Account Online”; the subject is unable to tell where this text links by examining the screenshot. The footer contains small grey print on “Email Preferences,” “Help / Contact Us,” hyperlinks to “Privacy” and “Security,” copyright information, trademark information, mailing address, and a message reference number.

Email Stimulus 3a

This stimulus is exactly the same as 3b, except that the footer has been removed; the message ends after the closing signature. We do not consider this a phishing message for the purposes of this experiment.

Hypothesis

One of the surprising observations from [15] is that subjects cited the small print typically found at the bottom of a document – e.g. copyright, licensing notifications, and legal disclaimers – as a feature that improved their trust of web pages and email messages. Stimuli 3a and 3b test this observation; we expect that our qualitative results will be substantiated by this test.

Email Stimulus 4a

This is an authentic notice from PayPal that notifies the user about a change their automatic logo insertion policy. This message was originally sent out in 2002 – and it may not be representative of messages they send today. The message has a header featuring a PayPal logo followed by a thick blue bar. Below the header is the message body that uses a personalized opening salutation. The message explains the policy change, and makes the claim in bold letters toward the end

If you do not wish to have PayPal automatically inserted in your listings, you must update your preferences by 9/25.

This sort of veiled threat is often associated with phishing messages. They mitigate this phishiness by describing how to contact the appropriate account settings rather page than providing a hyperlink. The letter ends with a closing salutation signed by “The PayPal Team.” Below the signature is a footer in small gray text that includes information about PayPal account settings and a hyperlink for changing them. The footer includes copyright and trademark information as well. We have altered this message to include a “Verisign Secured” trust logo in the footer in the line below the account change information.

Email Stimulus 4b

This is exactly the same message as 4a, except that it does not include the “Verisign Secured” logo. Although the date and addressee have been changed, stimulus 4b is a completely authentic email message.

Email Stimulus 5a

This is a phishing message that offers a \$50 incentive to people switching their online banking account to Citibank. The message uses the Citi corporate logo in its header. The message begins with a generic salutation “Dear Online Banker.” Inside the body is a url that does not belong to the `citibank.com` domain: `http://www.citibank.switch-today.com`. The closing salutation is signed by “Citibank Web Services.” Below the signature is a footer with black small print summarizing partial terms and conditions of the offer. Below this summary is a delivery message, This email was

Email Stimulus 5b

This message is exactly the same as 5a, except that we’ve inserted a “Verisign Secured” logo between the terms and conditions footer area and the delivery message.

Hypothesis

The interviews in [15] indicate that “Verisign Secured” is the most trusted third party endorsement of the ones they presented. Comparisons of 4a to 4b and 5a to 5b are intended to test the sway of this indicator. We hypothesize that when a subject has already made a judgement about the authenticity of an email, any peripheral assertions about security (such as endorsement logos at the bottom of the page) only intensify the subject’s opinion. Thus we expect to see the logo have a negative impact on an obvious phishing message and a positive impact on an evidently authentic message.

A.2 Web pages

Web Stimulus 1a

This is a screenshot of the page resulting from entering the official link for the AT&T Universal Card, `www.universalc card.com`, into the browser’s address bar; it is the authoritative login. The domain redirects to the URL, `https://www.accountonline.com/View?docId=Index&siteId=AC&langId=EN`. The web page content the phrase “AT&T Universal Card” prominently and in multiple locations. Other trust signals include multiple content page padlocks, a “Verisign Secure Site” endorsement, and a bevy of legal disclaimers and notices in small print at the bottom of the page. An important feature of the page is use of `https`, where communication is encrypted between client and server.

Web Stimulus 1b

The content window of this screenshot is identical to that of stimulus 1a, however the URL has been changed to `http://www.attuniversalc card.com` – a domain that did not resolve to any IP address at the time of experimentation. This screenshot simulates a phishing web site.

Hypothesis

URL plausibility is one of the chief factors in web page trust evaluation. We hypothesize that the better a URL reflects the content of a page, the more trustworthy the page will be rated. We expect subjects to give the `http://www.attuniversalc card.com` (5b) stimulus a higher rating, because it more accurately reflects the corporate branding campaign and because there are no CGI variables embedded in the URL. We expect this effect to exceed any mistrust caused by the use of `http` in place of `https`.

Web Stimulus 2a

This stimulus displays an exact copy of the official Citi credit cards login page (<https://www.citibank.com/us/cards/index.jsp>) under the unused domain, <http://www.citicardmembers.com>. This simulates a phishing web site that exploits a plausible cousin domain name. The stimulus does not indicate use of <https>, although it retains multiple small padlocks in the content window.

Web Stimulus 2b

This page uses the same URL as 2a (and therefore simulates a phishing web page), but the content has been modified in four places to more accurately align with the [citicardmembers.com](http://www.citicardmembers.com) domain:

- We inserted “cardmembers” in a small red and blue letters below the Citi corporate logo (Figure 18). The font for “card” is in the original logo’s red, while the font for “members” uses the original logo’s blue. The switch in color is aligned with the right extremity of the original logo.
- We changed the [citicards.com](http://www.citicards.com) hyperlink on the top navigation menu to [CitiCardMembers.com](http://www.CitiCardMembers.com).
- We replaced a small version of the original logo with the “cardmembers” modification at the bottom left of the page.
- We change the [Citi.com](http://www.Citi.com) hyperlink at the bottom of the page to [CitiCardMembers.com](http://www.CitiCardMembers.com).

Hypothesis

We hypothesize that aligning a web page’s content and layout with its domain will produce higher trust ratings.

Web Stimulus 3b

This is an official PayPal login page. When users visit the PayPal from the eBay home page, the hyperlink directs them to a version of the login page that features the eBay logo beneath PayPal logo. The address bar in this screenshot reports <http://www.paypal.com/ebay>, the non-SSL version of this page.

Web Stimulus 3a

This simulated phishing web page has the same content window of 3b but displays a URL with a non-resolving domain name, <http://www.ebaygroup.com/paypal/>. This bogus URL approximates the parent company’s name with a cousin domain and uses the subsidiary’s name as a subdirectory. This association is reinforced by the eBay logo underneath the PayPal logo.

Hypothesis

We expect that 3a will receive a lower rating than the legitimate 3b, but that the difference will not be as high as great as that between 1a and 1b since both domains match the content window well.

Web Stimulus 4b

This is a screenshot of the Indiana University Credit Union home page as it existed August 8, 2006 (the date was modified to appear more current at the time of testing). The URL, <https://www.iucu.org>, is correct; this stimulus represents an authentic web page. This page was put up in reaction to a recent phishing attack targeting IUCU customers. It contains stark warning banner at the top of the page:

WARNING: ALL IU CREDIT UNION MEMBERS PLEASE READ! Phishing scam in progress;
click here for important information

The page contains two additional “Phishing scam in progress” warnings in the “What’s new @ IU” section and inside a section labeled “Critical Fraud Alerts!”

Web Stimulus 4a

This stimulus uses the IUCU home page as it existed on September 20, 2006 as a template; this simulates an authentic login page. By this time, the warning banner and the “phishing scam in progress” link in under “What’s new @ IU” had been removed. The “Critical Fraud Alerts!” heading continued to display the “phishing scam in progress” hyperlink, as well as other warnings about “Vishing” and a VISA phishing email. Tagline for the section was “Your account security is very important to us.” We softened the language to sound less threatening and more positive:

- We changed the “Critical Fraud Alerts!” section heading to “Fraud Prevention Center.”
- We changed the “Your account security is very important to us” tagline to “We’re committed to your account security.”
- We revised the sentence following the “Protect yourself from fraud” from “Click here for important security and blocked ATM/Visa Card updates” to “Click here for effective fraud prevention strategies.”
- We eliminated mention of all fraud warnings.
- We added a line, “Concerned about fraud? IU Credit Union helps ensure online account safety – learn more.”

Hypothesis

The qualitative observations of [15] suggested that an overemphasis on security could discredit a web page. Moreover, some participants thought that “phishing” sounded unprofessional and would not be used in communications from a financial institution. Stimuli 4a and 4b tests the effect of subtle and stark security language on web pages. We expect the subtle language to rate higher than the stark language.

Web Stimulus 5a

This is a simulated phishing web page that uses the poorly formed URL, <http://www-wellsfargo.com/>; a dash replaces the dot after **www**. The browser content window looks exactly the same as the authentic content served at <https://www.wellsfargo.com>.

Web Stimulus 5b

This simulated phishing web page also uses the poorly formed URL, <https://www-wellsfargo.com/>, but indicates an **https** connection instead. To further bolster the credibility of the SSL session, we have inserted a center-aligned “Verisign Verified” endorsement logo in the web page’s footer.

Hypothesis

Although interviews and observations in [15] indicate that Verisign endorsements are among the strongest third party recommendations, the results also indicate that people place more trust in their URL evaluation (when they remember to do it). Since this is the second to last stimulus in the test, we think that most people that can evaluated URLs will have remembered to do so by this point. We think that both web sites will be rated poorly with only moderate benefit (if any) resulting from the Verisign endorsement graphic and the **https** protocol.

Web Stimulus 6a

This is the official web page of the Dell battery recall program. It's URL is <http://www.dellbatteryprogram.com/>, a legitimate cousin domain of Dell's primary domain, dell.com. It has the same header, footer, and opening paragraph used in the simulated phishing message, email stimulus 2a. The page contains a list of affected battery part numbers and a black-and-white diagram showing where to read this number on the battery. The page is predominantly text and is devoid of the usual product promotions on the primary corporate web site. There are multiple links that assist with some aspect of battery identification and return procedures.

Web Stimulus 6b

This screenshot shows the official web page for the *Settlement of Frank Chavez V. Netflix, Inc. Class Action*. The address bar displays <http://www.netflix.com/Settlement> – the result of redirection from <http://www.netflixsettlement.com>. The page is plainly formatted: text only, no promotions, and fits on one 1024x768 screen. There are various hyperlinks that address potential customer questions and settlement details.

Hypothesis

Both pages display their information in a straightforward manner. The Dell web page has slightly more elaborate formatting due to its header graphic, language selection menu, and footer menu. We think that the Netflix page will rate higher because the URL indicates that it is part of the company's domain, while the Dell URL is a cousin domain that could have been plausibly purchased by a third party.

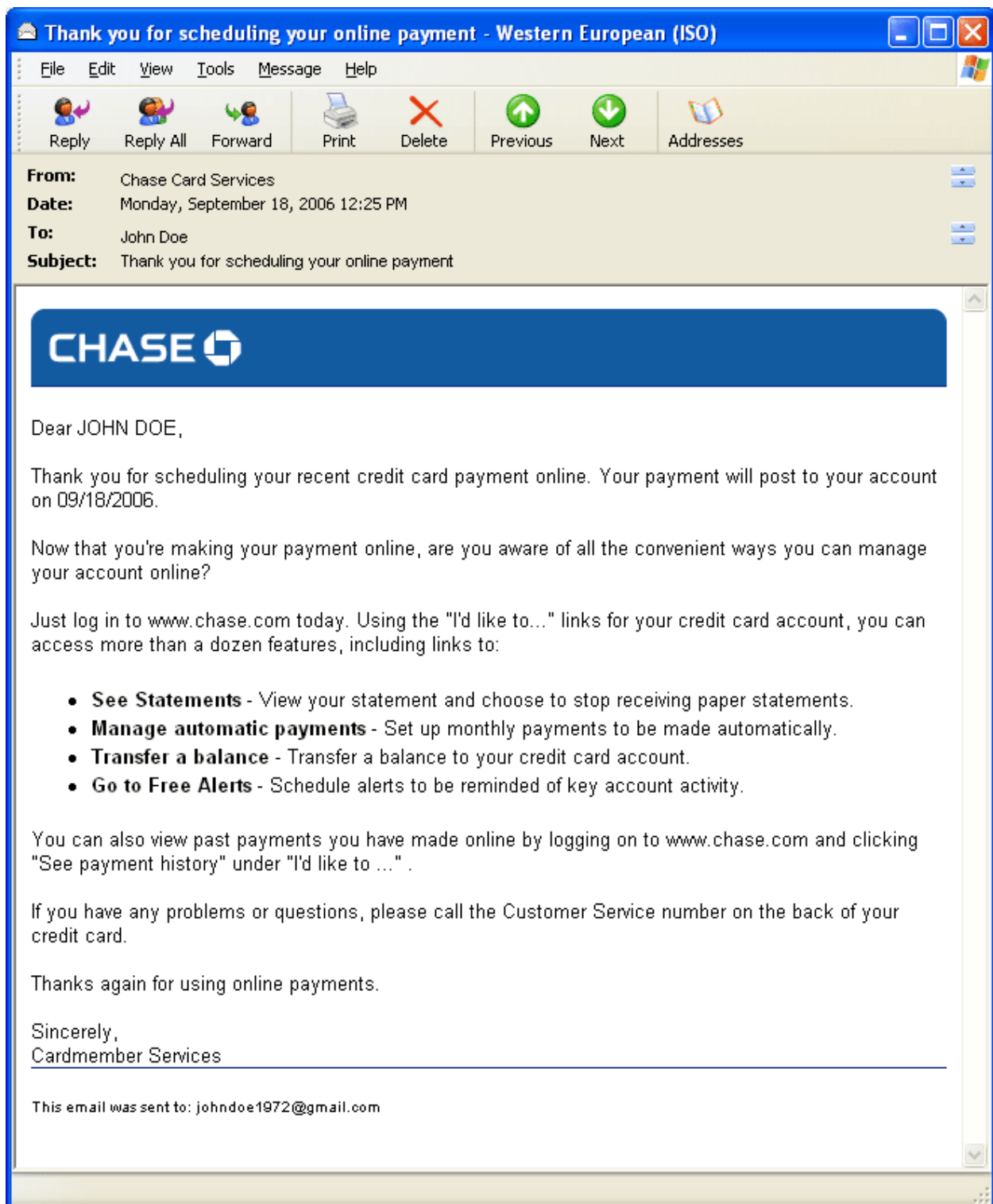


Figure 3: (Email Stimulus 1a; authentic) Authentic Chase payment notification (i.e. as sent). Rated 3.40.

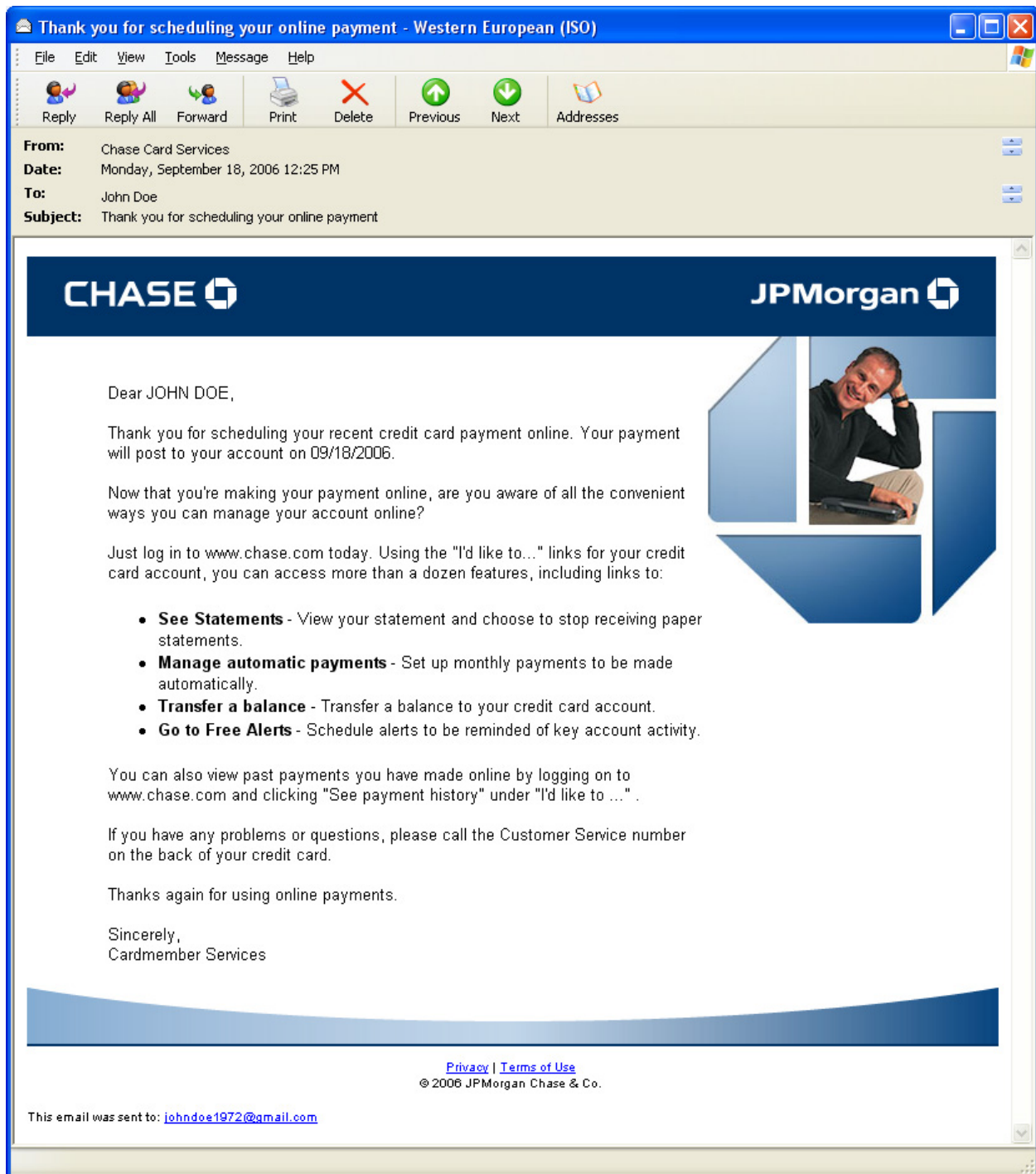


Figure 4: (Email Stimulus 6b; authentic) Plausible payment notification with fancy layout. Rated 3.71.

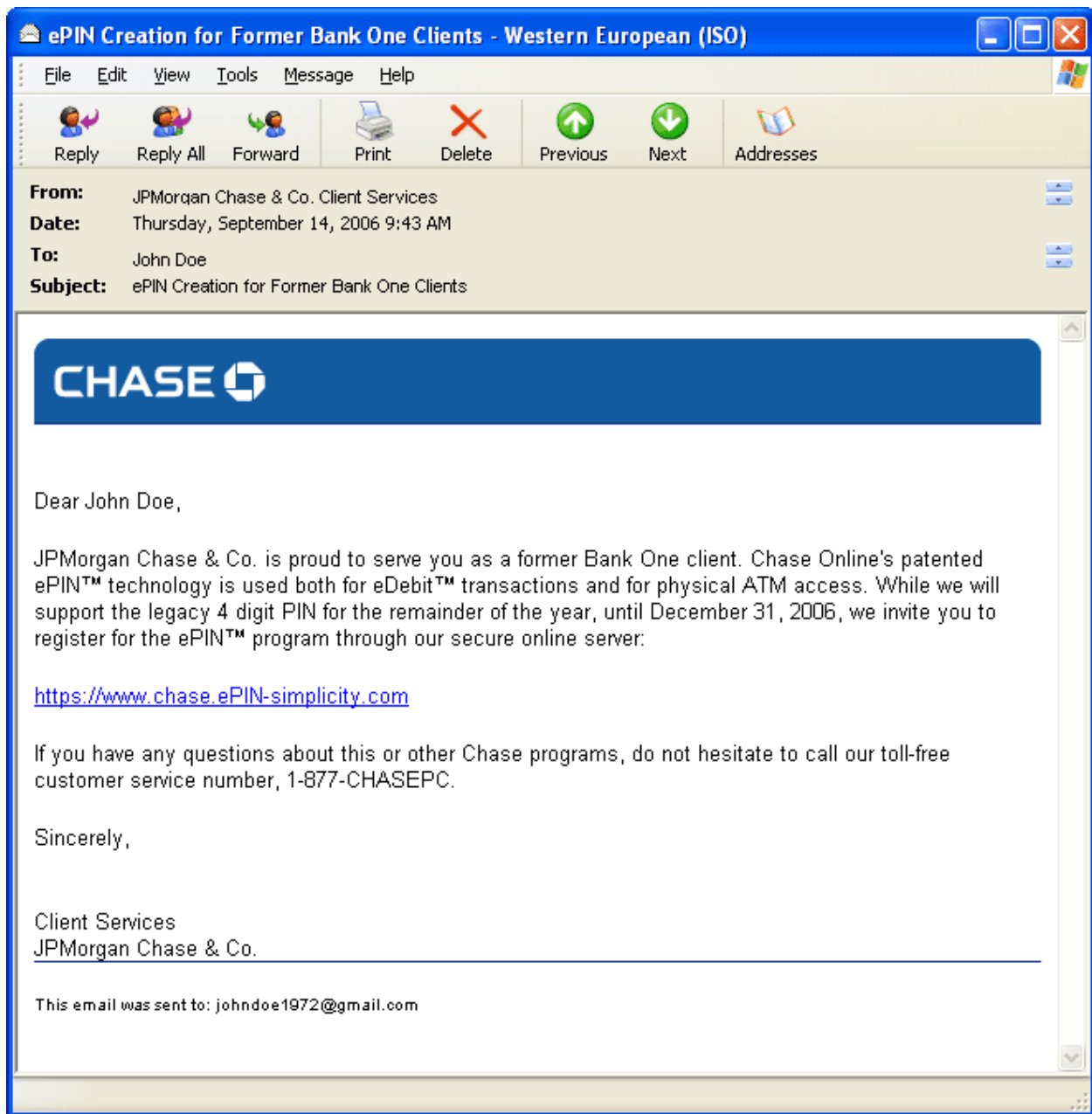


Figure 5: (Email Stimulus 1b; phishing) Phishing message with standard Chase layout. Rated 3.15.

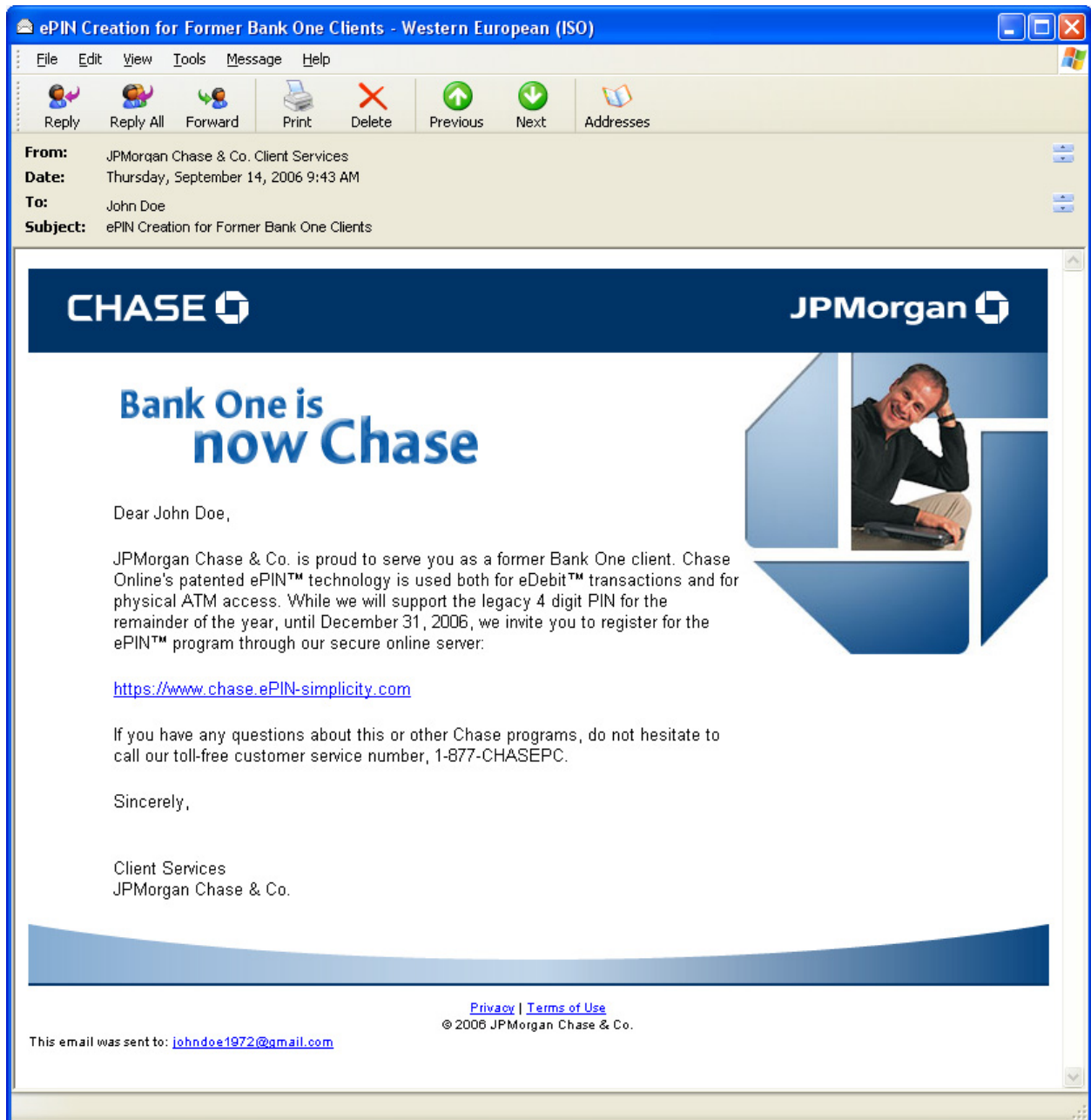


Figure 6: (Email Stimulus 6a; phishing) Phishing message with fancy layout. Rated 3.19.

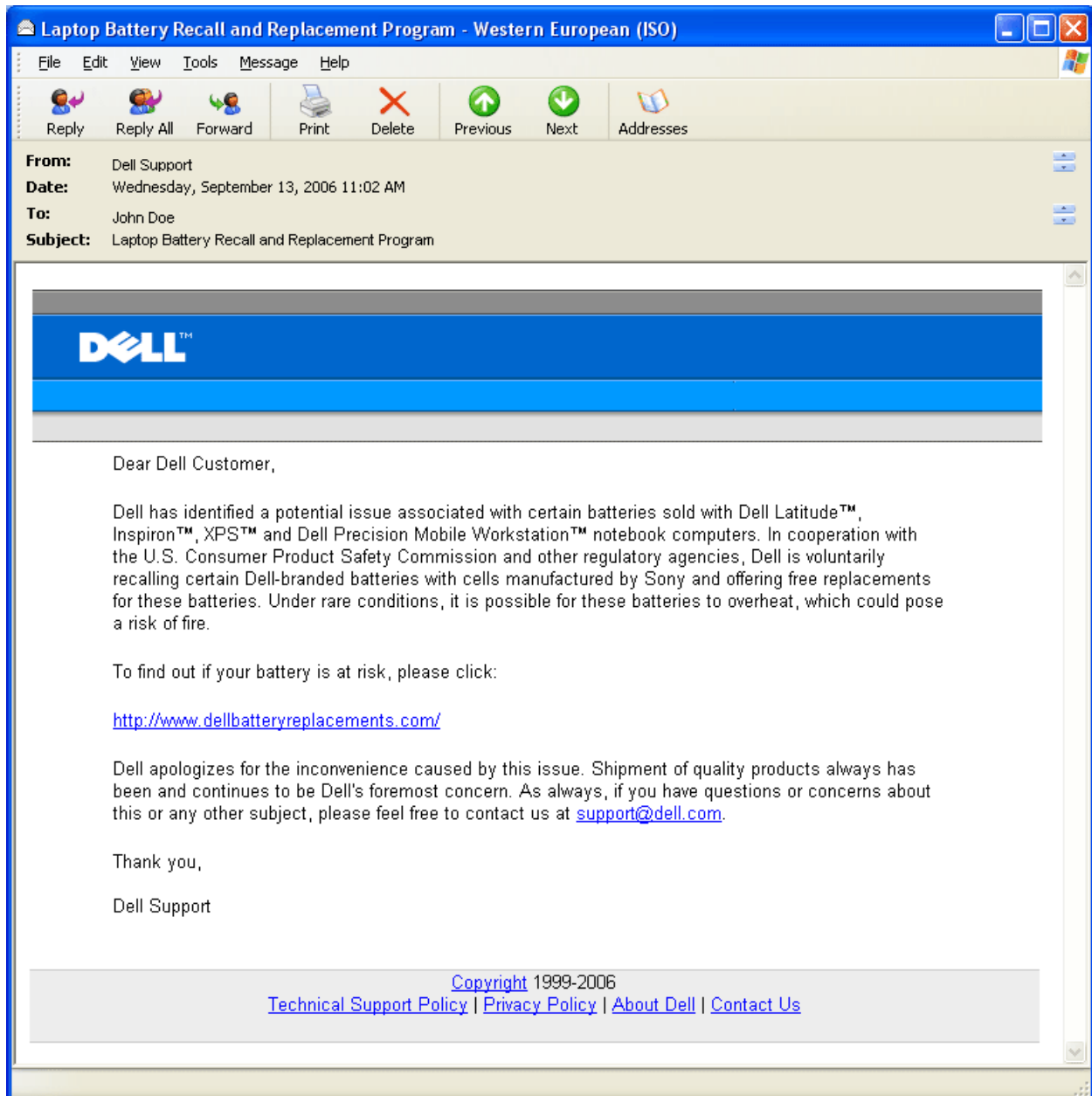


Figure 7: (Email Stimulus 2a; phishing) Dell battery phishing message. The URL <http://www.dellbatteryreplacements.com> is not in use; the real URL is <http://www.dellbatteryreplacements.com>. Rated 3.62.

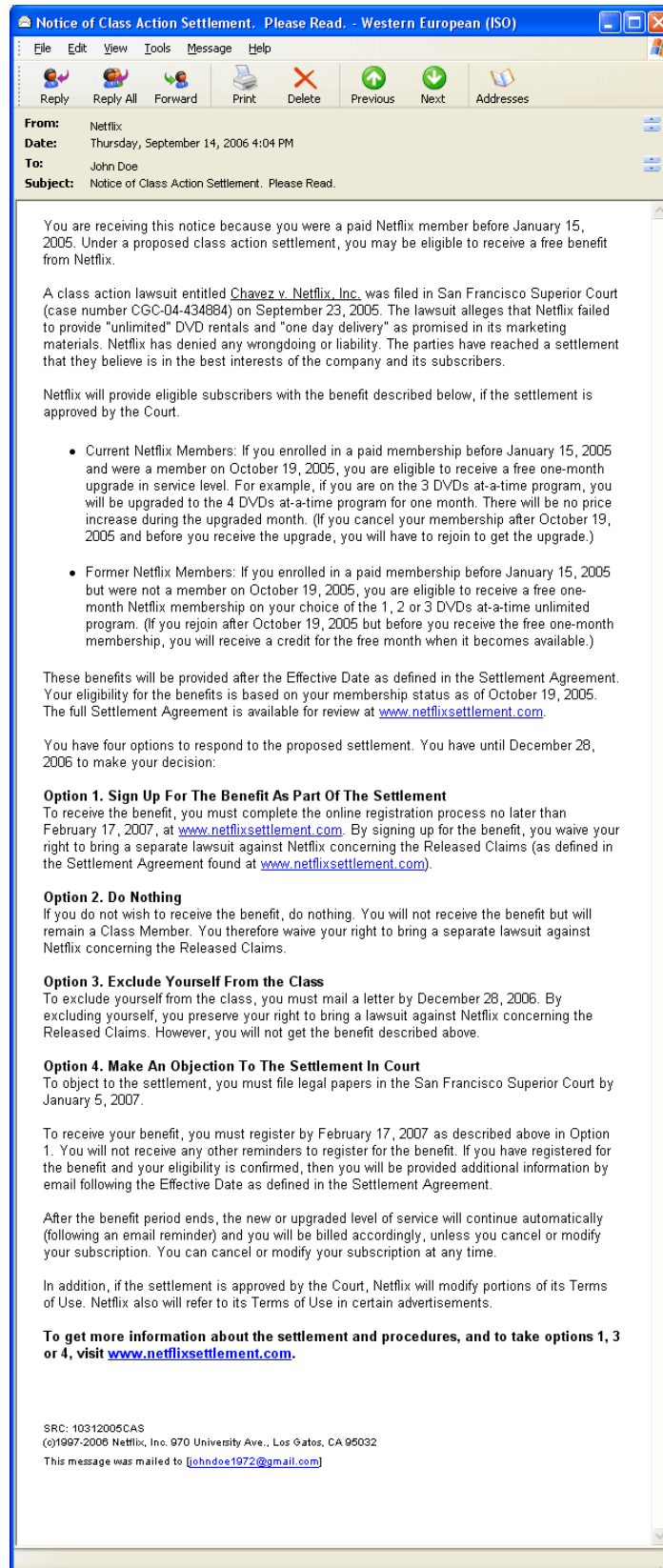


Figure 8: (Email Stimulus 2b; authentic) Authentic email (i.e., as sent) notification of Netflix class action settlement. Dates have been changed to appear relevant at time of testing. The URL, <http://www.netflixsettlement.com> – the legitimate site, is apparently a third party domain. Rated 2.75.

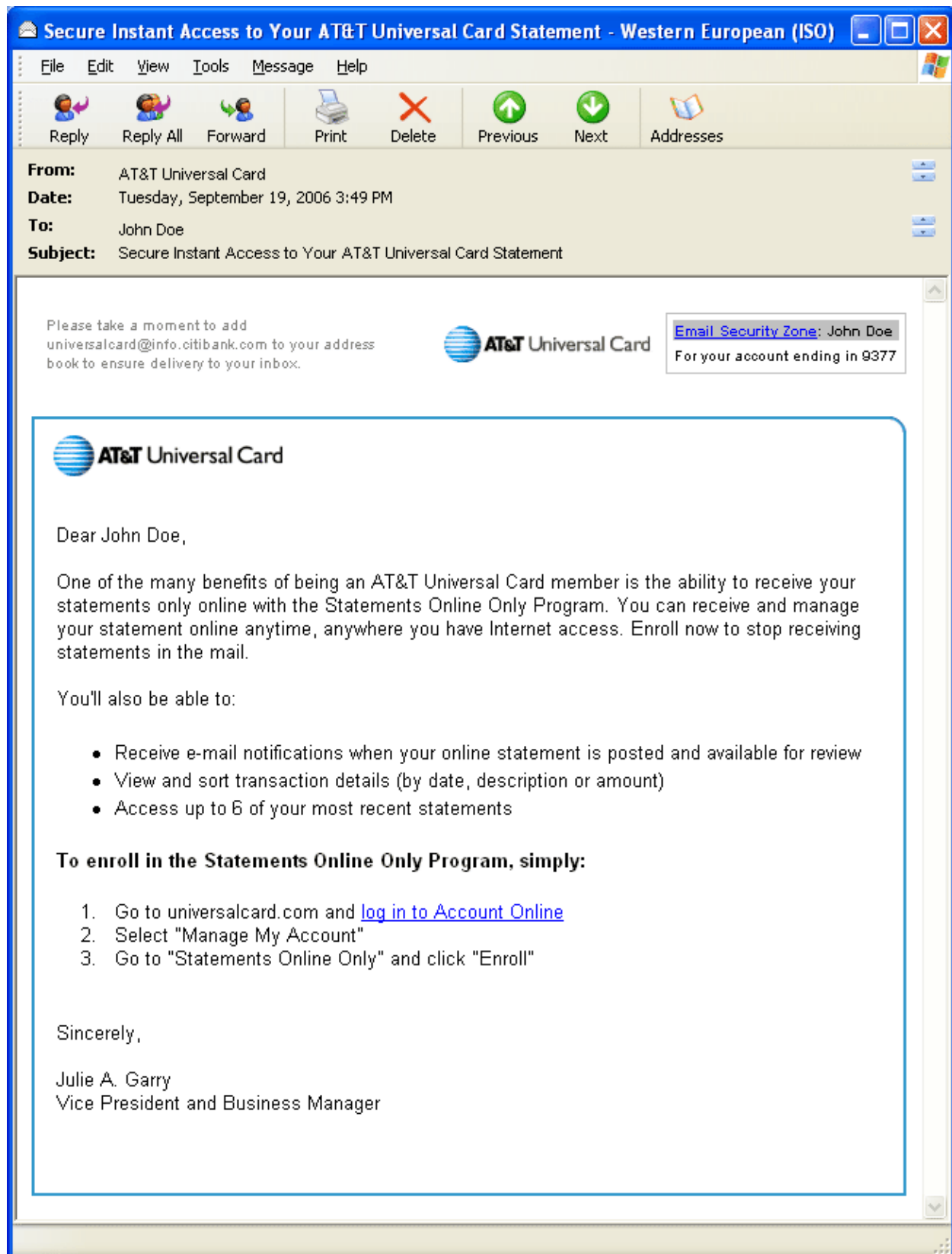


Figure 9: (Email Stimulus 3a; authentic) Legitimate AT&T Universal Card payment notification *without* small print legalese at bottom. Rated 3.05.

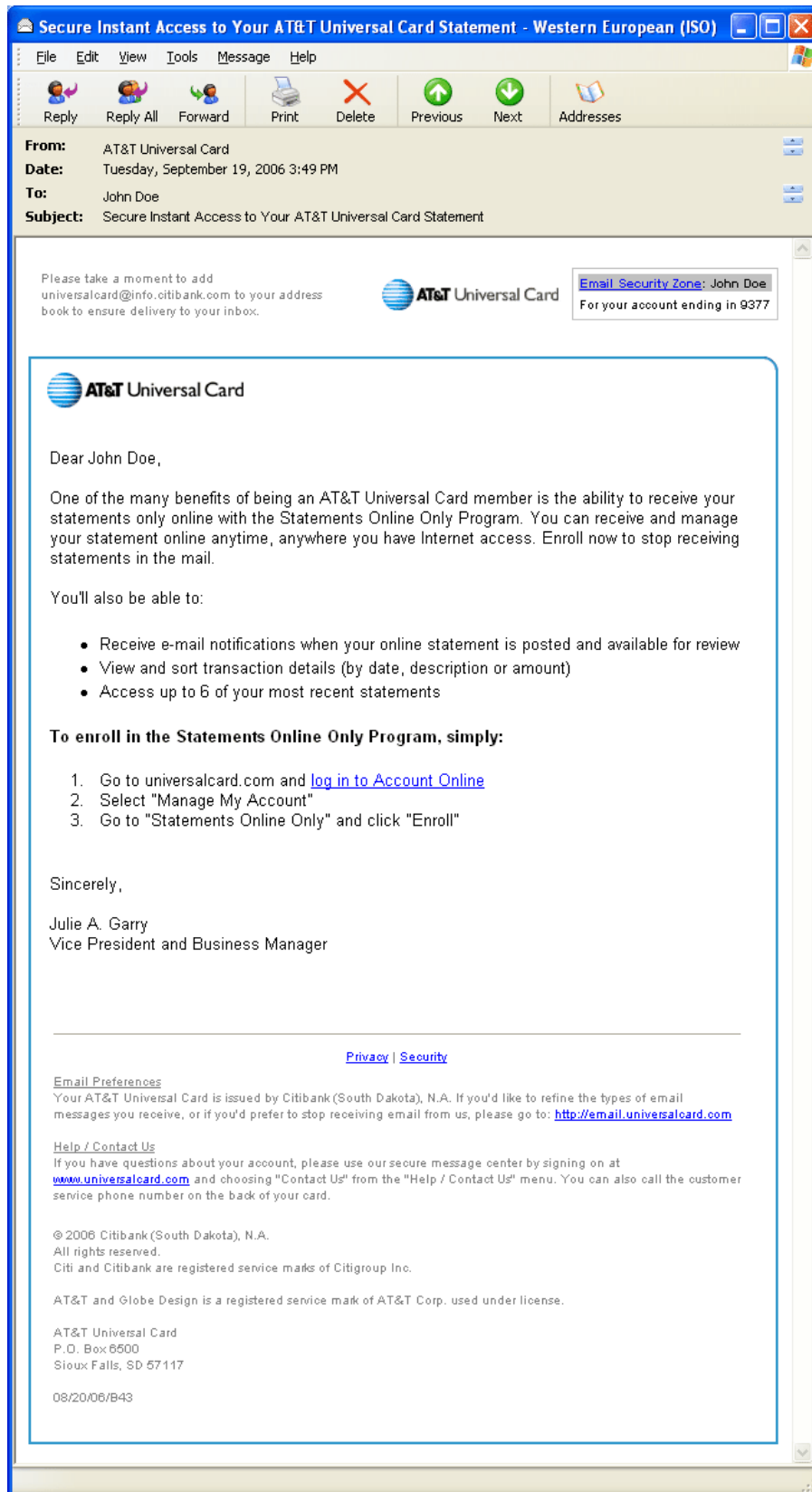


Figure 10: (Email Stimulus 3b; authentic) Legitimate AT&T Universal Card payment notification *with* small print legalese at bottom (i.e. as sent). Rated 3.63. ³⁰

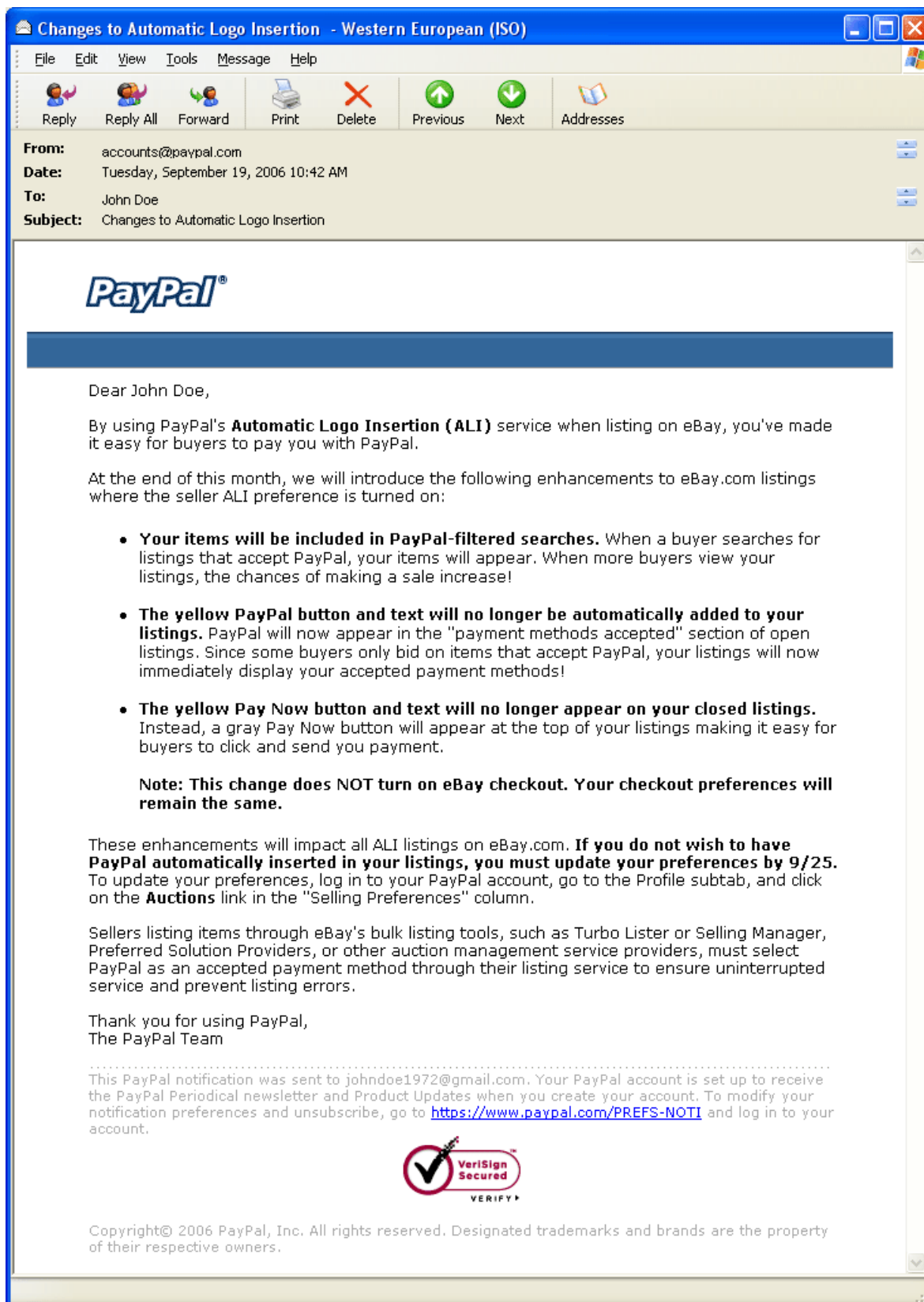


Figure 11: (Email Stimulus 4a; authentic) Legitimate PayPal policy change notification with Verisign logo added. Rated 3.19.

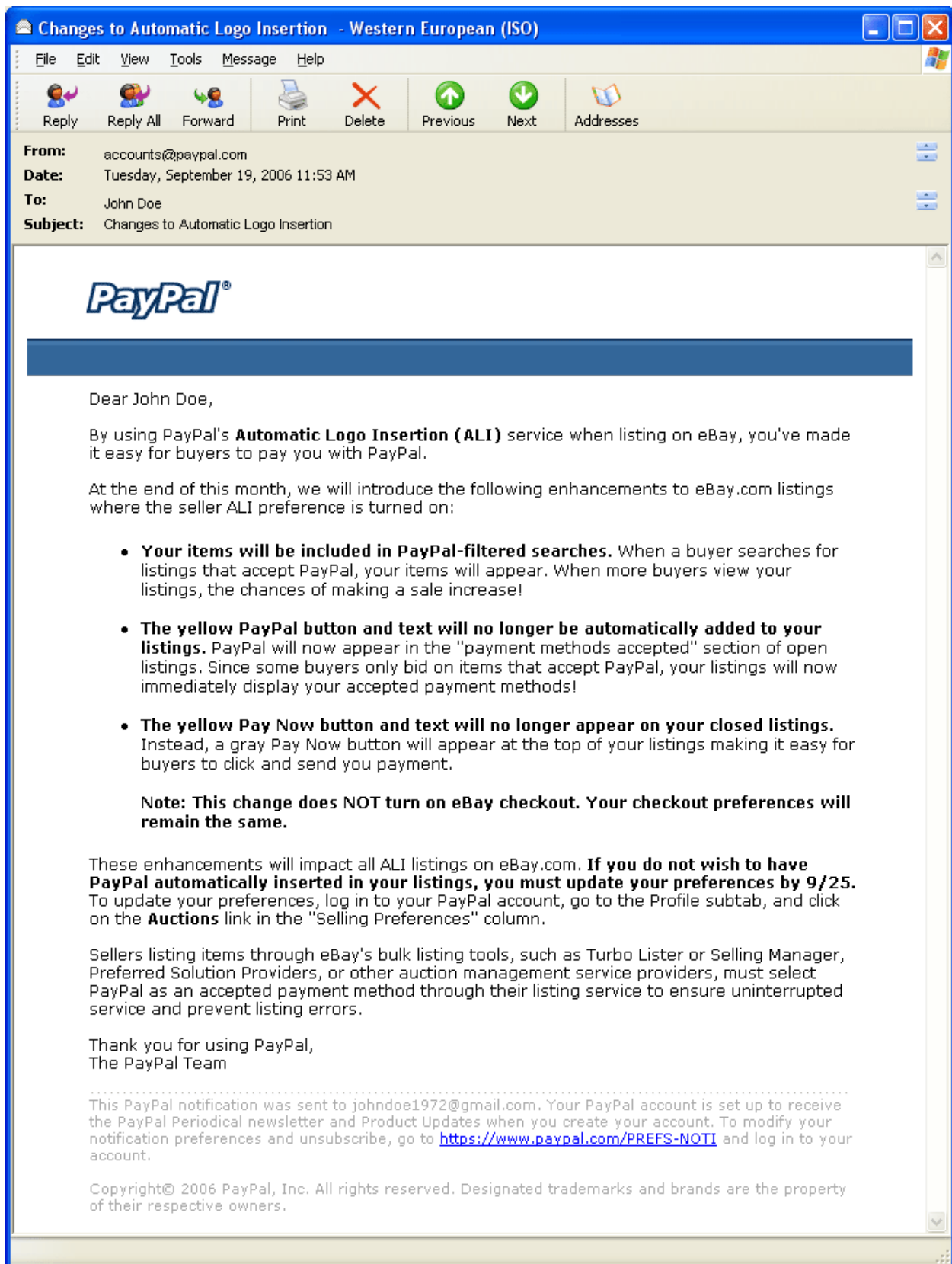


Figure 12: (Email Stimulus 4b; authentic) Legitimate PayPal policy change notification without Verisign logo (i.e. as sent). Rated 3.23.

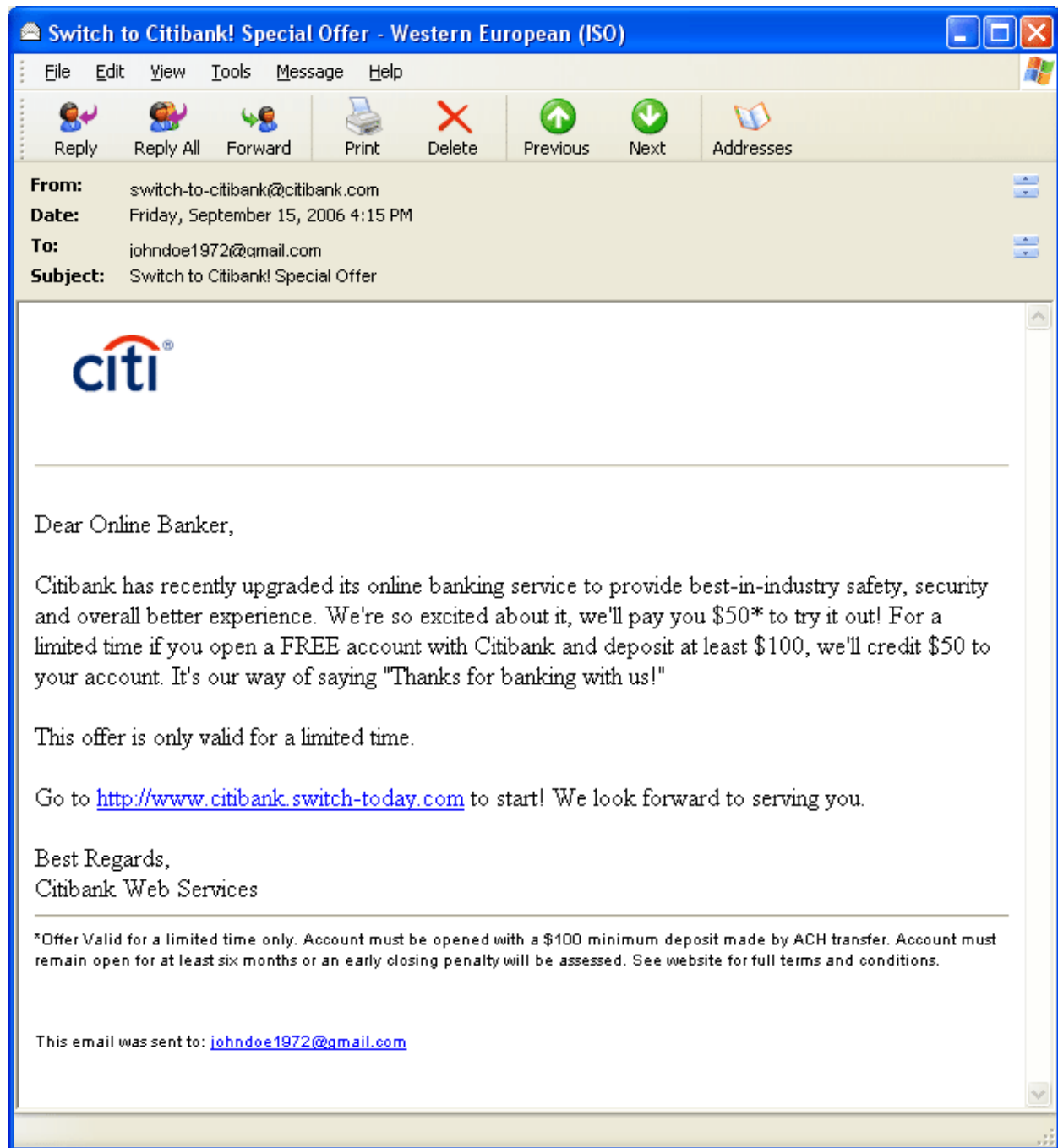


Figure 13: (Email Stimulus 5a; phishing) Citibank phishing message with no trust logos. Rated 2.40.

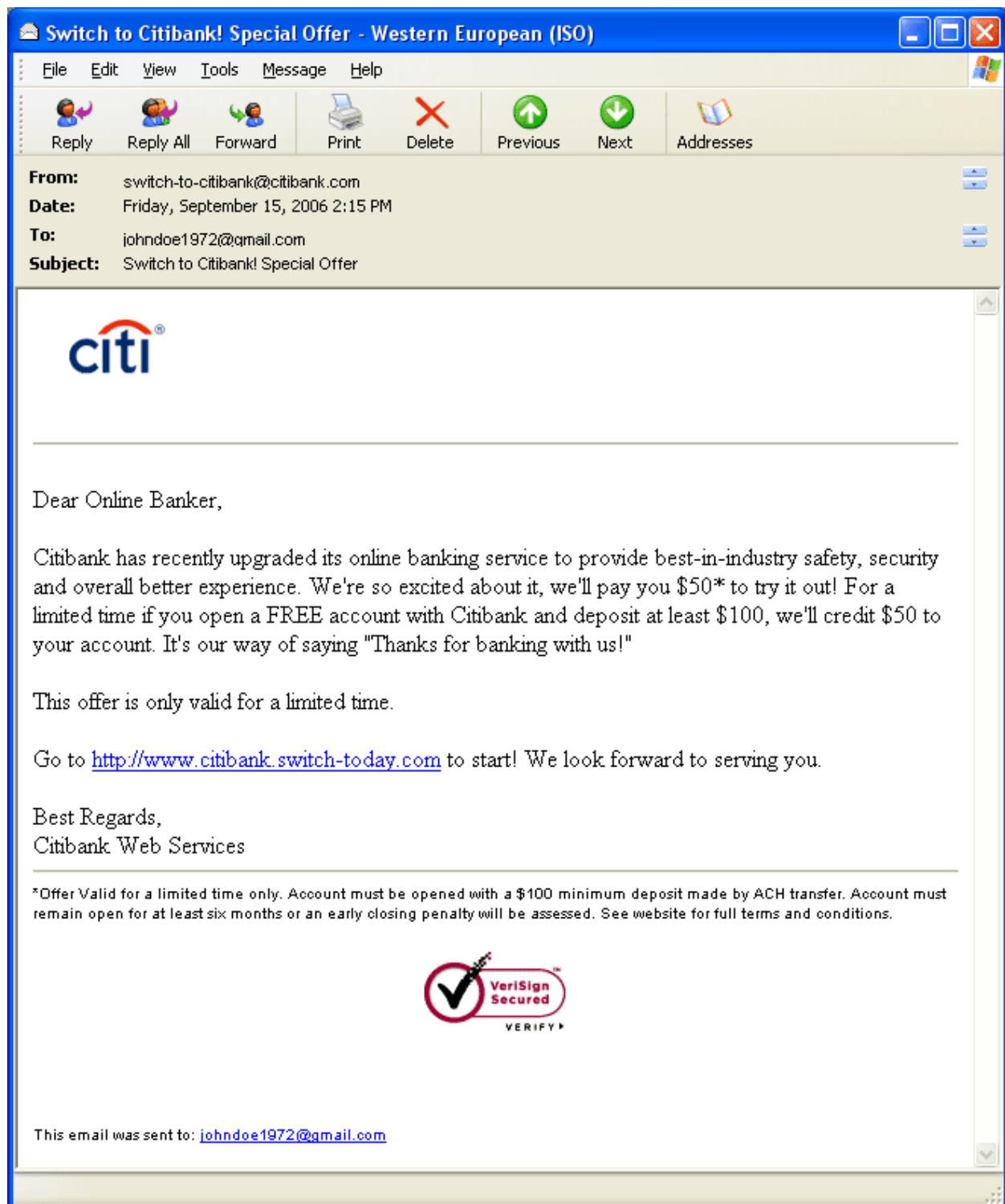


Figure 14: (Email Stimulus 5a; phishing) Citibank phishing message with Verisign logo. Rated 2.73.

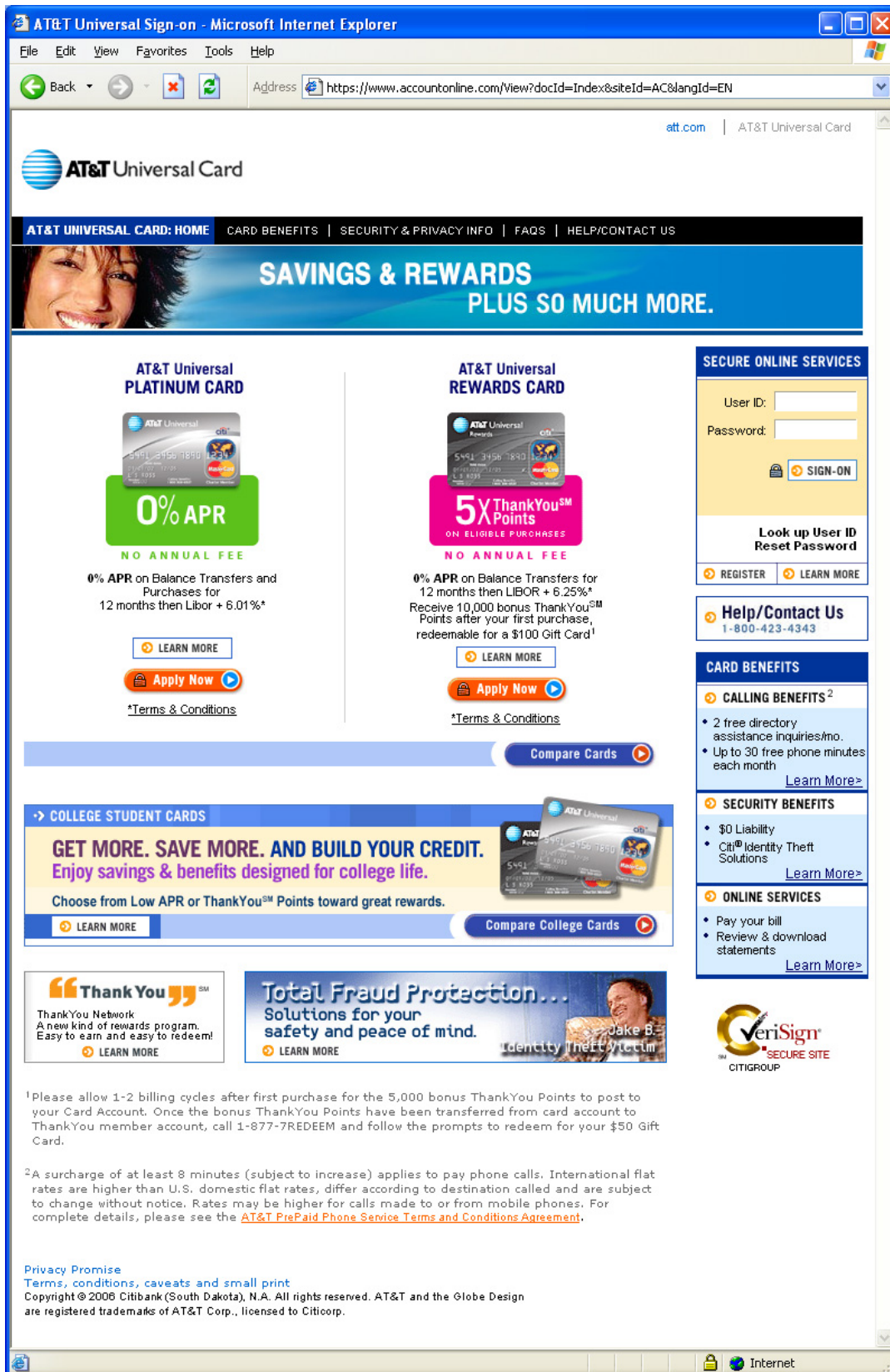


Figure 15: (Web Stimulus 1a; authentic) Authentic AT&T Universal Banking site with complex URL <https://www.accountonline.com/View?docId=Index&siteId=AC&langId=EN>. Rated 2.76.

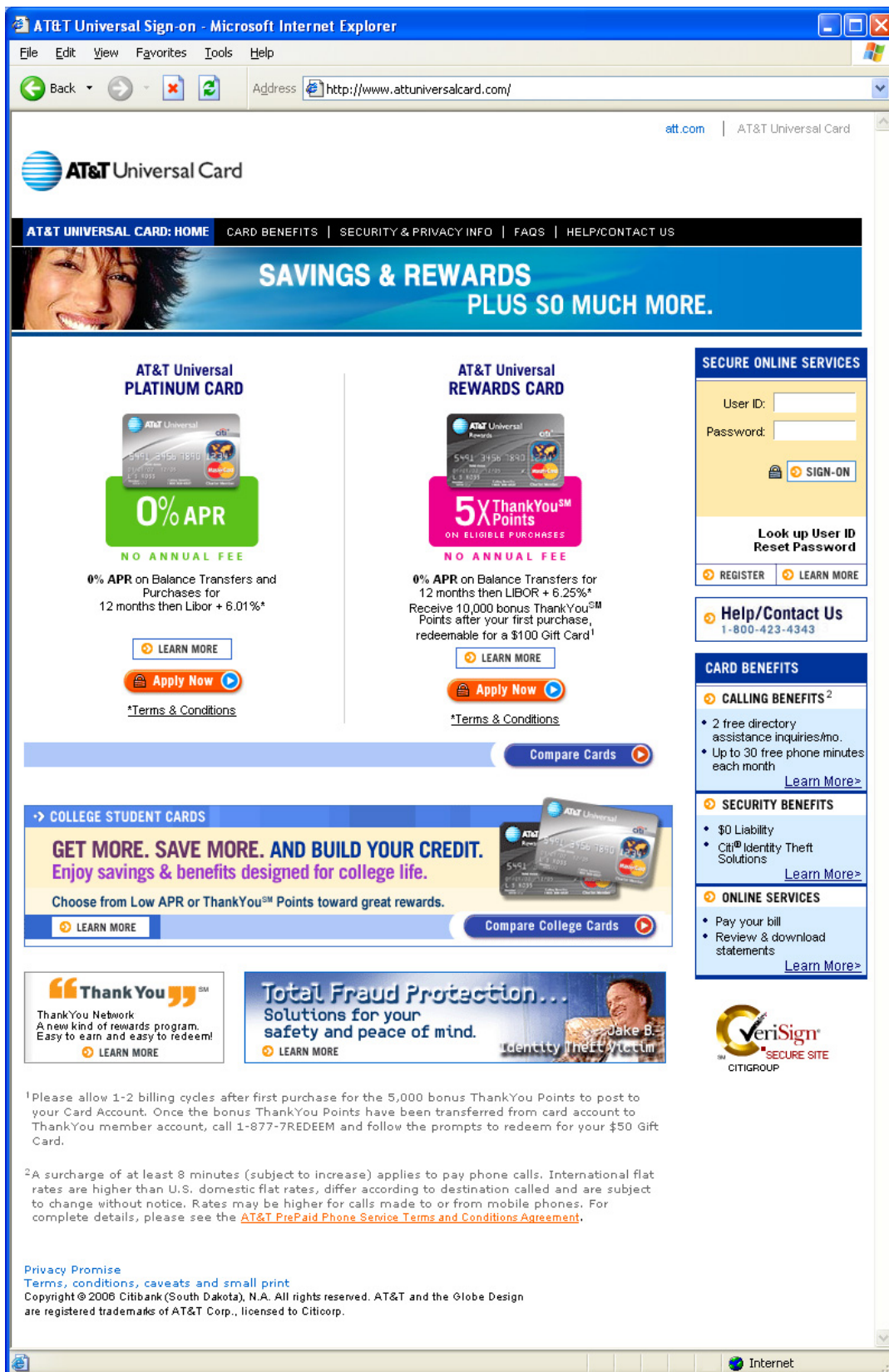


Figure 16: (Web Stimulus 1b; phishing) AT&T Universal Banking site with unused URL <http://attuniversalcard.com>. Note that this is not SSL. Rated 3.24.



Figure 17: (Web Stimulus 2a; phishing) Citibank credit card phishing site with original art and unused URL <http://www.citicardmembers.com/>. Rated 3.11.



Figure 18: (Web Stimulus 2b; phishing) Citibank credit card phishing site with modified art to match URL <http://www.citicardmembers.com/>. Rated 3.43.



Figure 19: (Web Stimulus 3a; phishing) PayPal phishing site with unused URL <http://www.ebaygroup.com/paypal>. Note eBay logo underneath PayPal logo. Rated 3.35.



Figure 20: (Web Stimulus 3b; authentic) Authentic PayPal site when linking from eBay; <http://www.paypal.com/ebay>. Note eBay logo underneath PayPal logo. Rated 3.66.

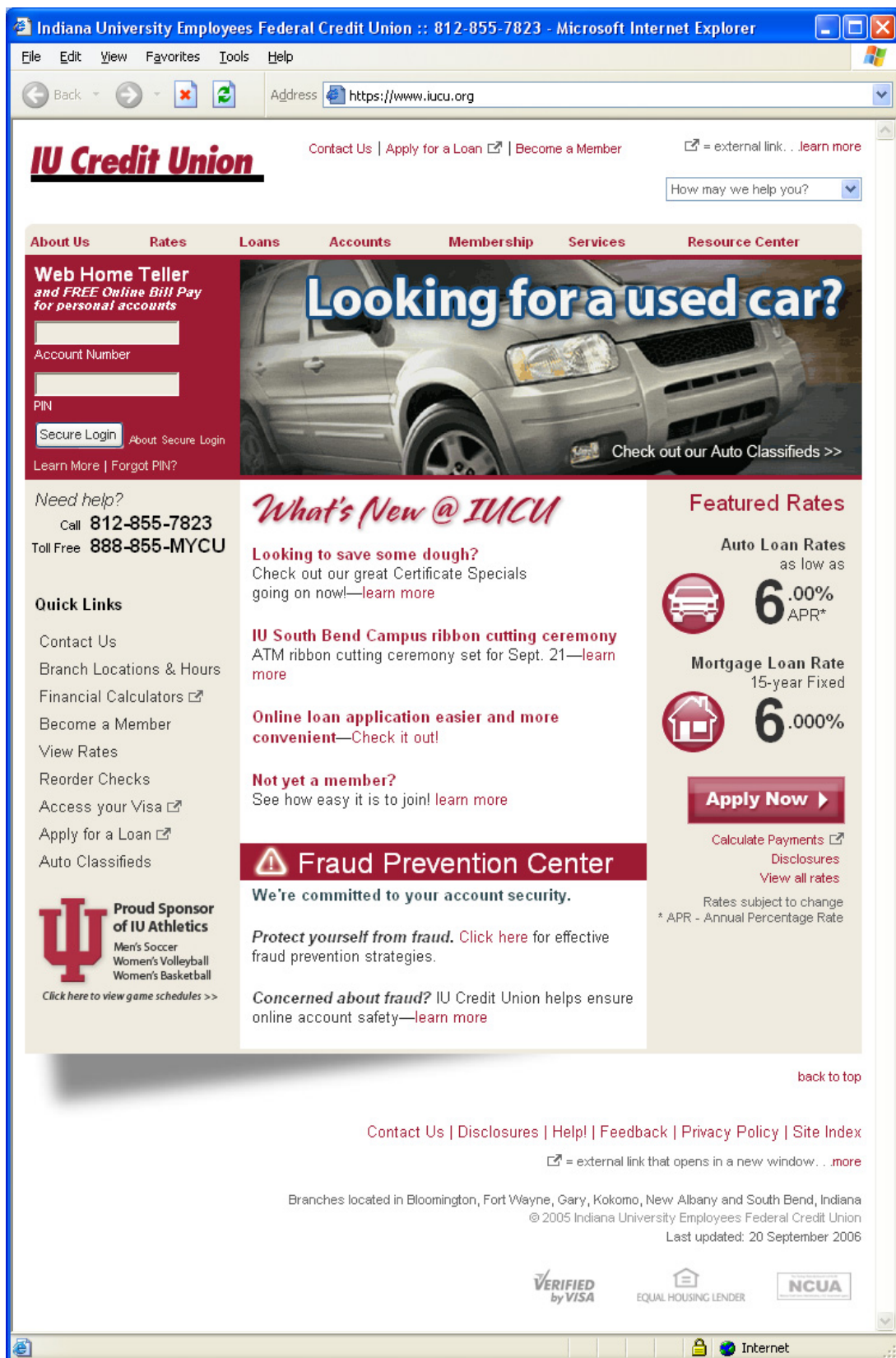


Figure 21: (Web Stimulus 4a; authentic) IUCU with authentic URL, <https://www.iucu.org>. The fraud language has been softened from the original language. Rated 3.69.

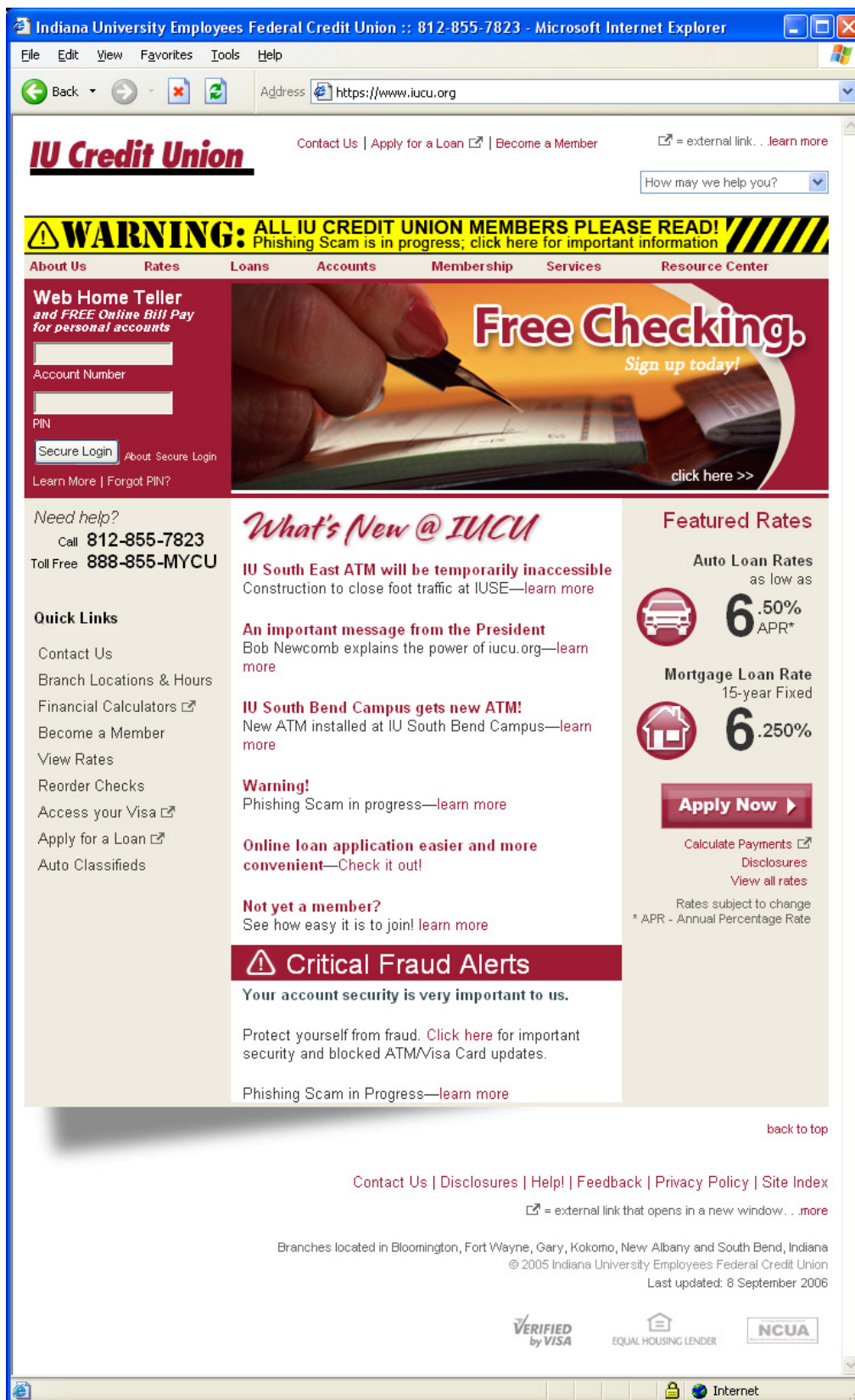


Figure 22: (Web Stimulus 4b; authentic) IUCU with authentic URL, <https://www.iucu.org>. Completely authentic website after a highly publicized phishing attack. The fraud alert language is severe. Rated 3.33.



Figure 23: (Web Stimulus 4a; phishing) Wells Fargo phishing site with unused URL <http://www-wellsfargo.com/>; no trust logos. Rated 3.17.

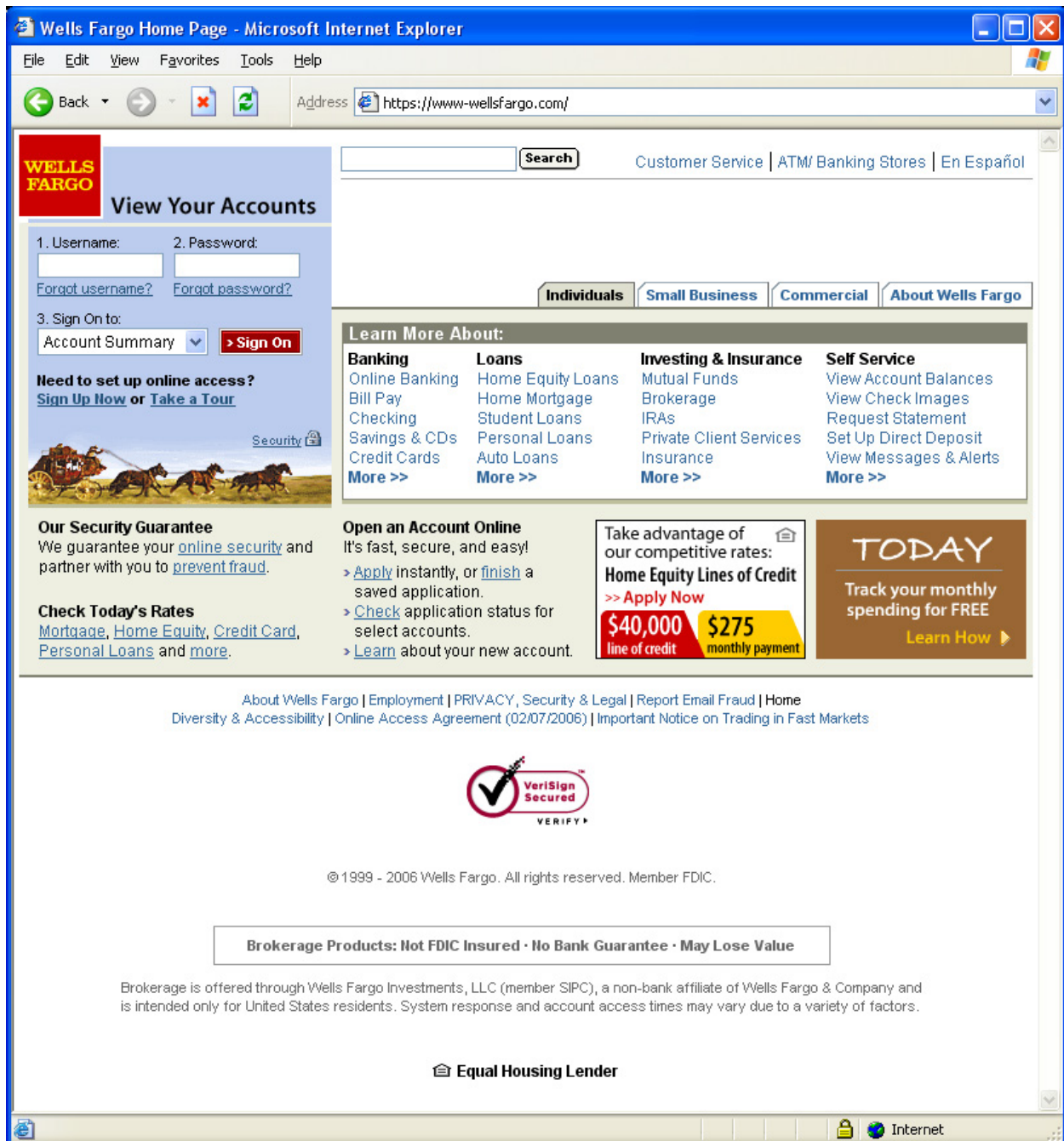


Figure 24: (Web Stimulus 4b; phishing) Wells Fargo phishing site with unused URL `http://www-wellsfargo.com`; added Verisign logo and `https`. Rated 3.50.

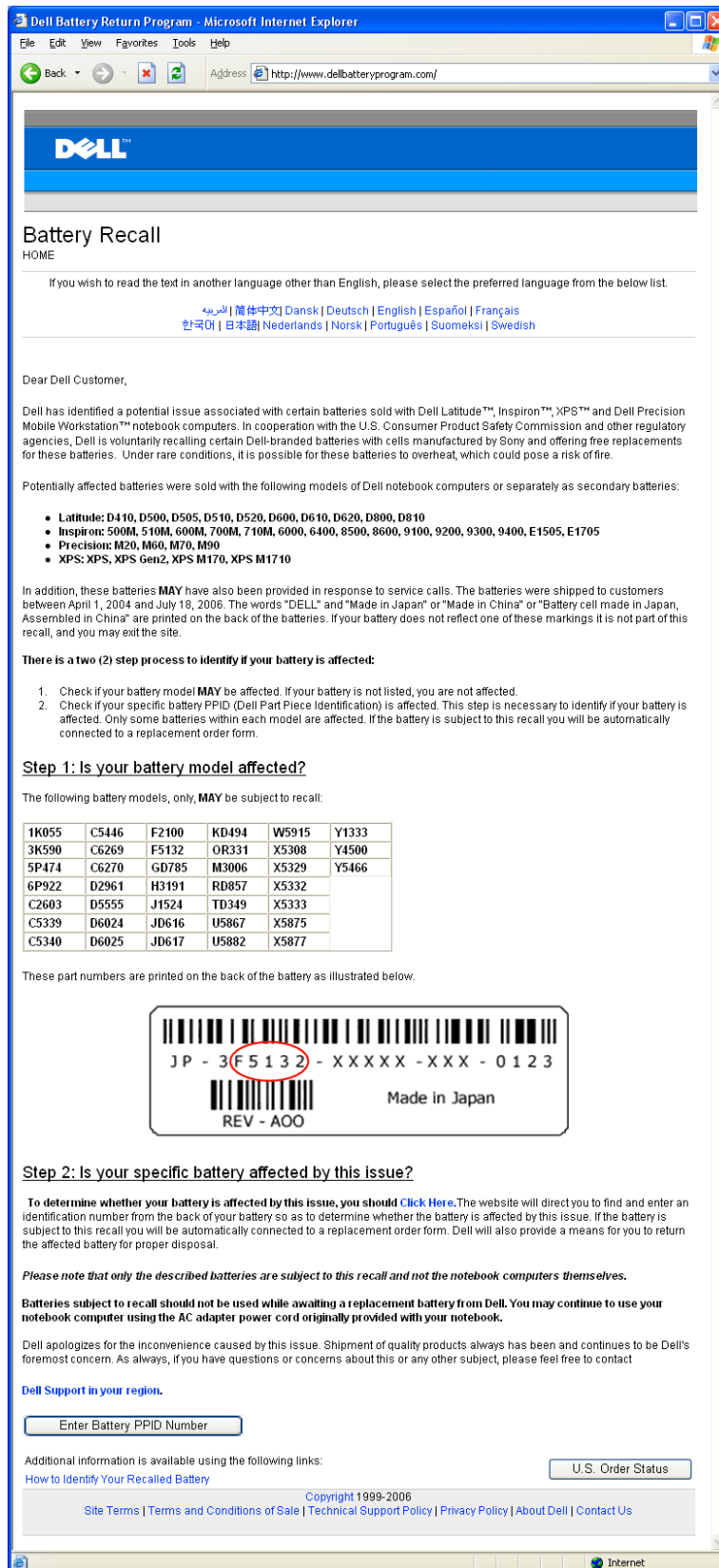


Figure 25: (Web Stimulus 6a; authentic) Authentic Dell battery website. Note that the URL <http://www.dellbatteryprogram.com> appears to be a third party site. Rated 3.54.

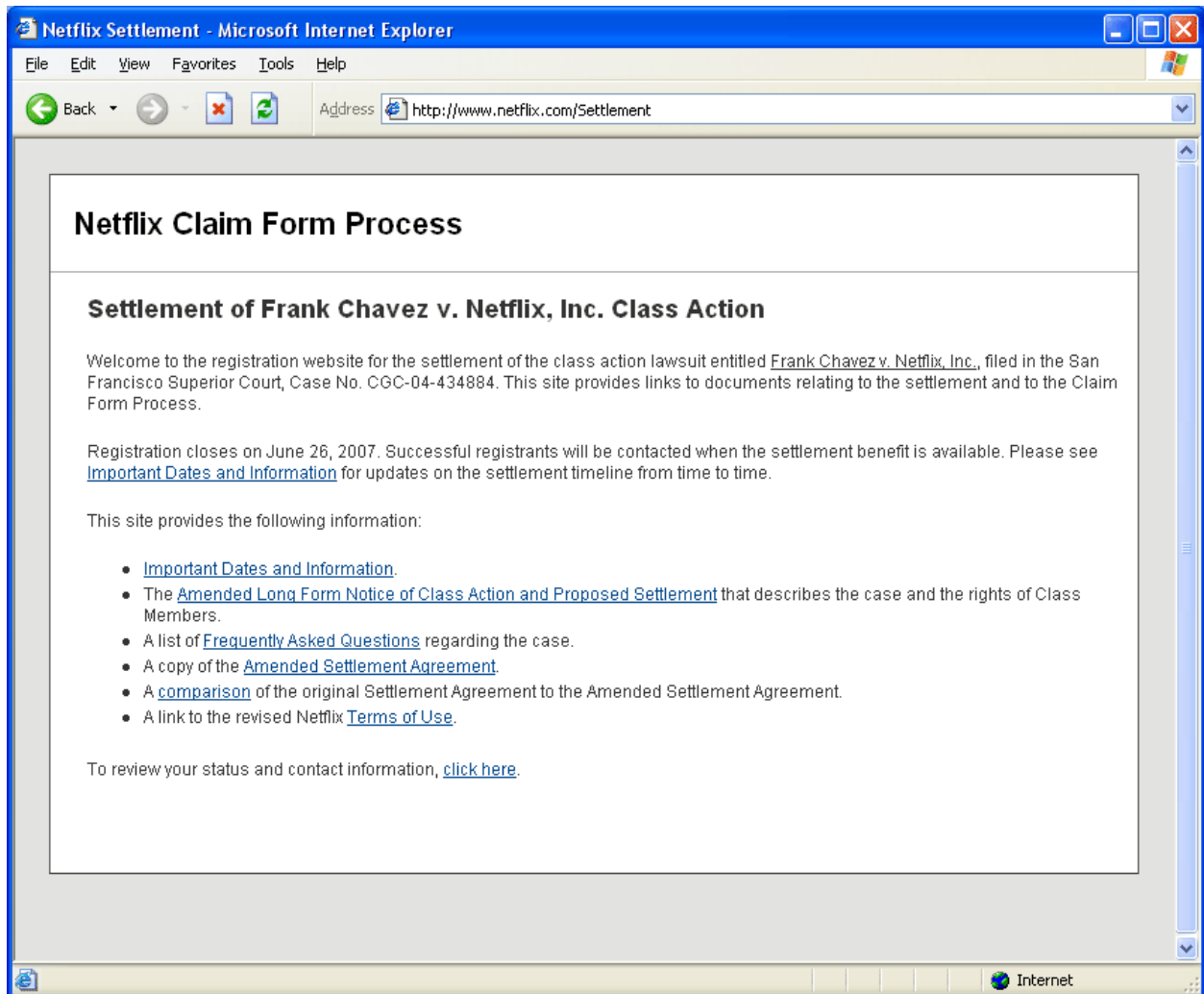


Figure 26: (Web Stimulus 6b; authentic) Authentic Netflix Settlement page with modified dates. The URL <http://www.netflix.com/Settlement> is redirected from <http://www.netflixsettlement.com/>. Rated 2.58.