Received: 23 October 2018 | Revised: 16 July 2019 | Accepted: 23 July 2019

DOI: 10.1002/acp.3594

SHORT PAPER

WILEY

Perceptual representation of spam and phishing emails

Pooja Patel | Dawn M. Sarno | Joanna E. Lewis | Mindy Shoss | Mark B. Neider | Corey J. Bohil

University of Central Florida, Orlando, Florida

Correspondence
Pooja Patel, University of Central Florida, Orlando, FL, USA.
Email: pooja.ucf@gmail.com

Present address
Joanna E. Lewis, University of North Colorado, Greely, CO, USA

Summary

Understanding how computer users allocate attention to features of potentially dangerous emails could help mitigate costly errors. Which features are salient? How stable is attention allocation across variation in email features? We attempted to measure the mental salience of several email features common in spam and/or phishing emails. We created two email sets: one in which messages contained company logos and urgent actionable links and one without these features. Participants rated pairwise similarity of emails within each set. Multidimensional scaling (MDS) analysis was conducted to quantify psychological similarity between emails. A separate group rated the same emails for presence of five other features: important downloadable content, collecting personal information, account deletion or suspension, advertisement, and large images with clickable content. Regressing feature ratings onto the MDS coordinates revealed that similarity judgments were influenced mostly by advertisement/large images and collecting personal information, regardless of presence or absence of company logos and urgent actionable links.

KEYWORDS

cybersecurity, email, phishing, spam

1 | INTRODUCTION

The way that computer users interpret and respond to potentially dangerous email messages is a critical aspect of cybersecurity. Although there are numerous ways in which valuable private information can be targeted, email is one of the most commonly used methods in use (Tapper, 2017; Wall, 2018). As Wall (2018) points out, emails allow for a large-scale attack with greater return in less time than blog commenting or fake offers targeting inexperienced online shoppers. This problem receives a great deal of attention in companies with a large employee base who risk exposing proprietary and other sensitive information. However, email security is not of concern just to office workers but to everyone who accesses the internet.

Although highly effective, automated spam filters are not sufficient to defend against this constantly changing threat (Vishwanath, Harrison, & Ng, 2016). Filtering emails at the server level has shown to be lacking in identifying dangerous emails. For example, Clayton (2004) examined a month's worth of email distribution coming from ISP "smarthost" servers to test the sensitivity of server filtering protocols for thousands of emails. The study found that the largest error was the failure to deliver genuine emails, followed by users accidentally forwarding spam to other workers and serious virus hazards getting misidentified as spam (i.e., misidentified as mere unwanted advertising).

Rather than filtering by email content, technical administrators can utilize methods to systematically prevent spam at the domain name system (DNS) level (e.g., storing lists of sender domain names to block). Yet this blacklisting method has been found to let 20% of spam sources escape detection (Jung & Sit, 2004). At the level of individual workstations, a recent evaluation showed that Bayesian classifiers were effective 95% of the time for 4,600 email classifications (Rathod & Pattewar, 2015). But even this impressive detection rate would leave 230 emails misclassified.

These statistics make clear that email users are the last line of defense against proprietary or personal data loss. Predictions for 2019 indicate that the average business user will deal with roughly 252 emails a day, with one in 20 emails being misclassified (Team, 2015). Thus, it is imperative to develop methods of reducing susceptibility to these attacks.

One way to assist users in reducing the likelihood of engaging with potential threats would be through some sort of training. For example, an increasingly popular approach is embedded training, in which simulated phishing emails are sent by company servers along with other incoming messages to track and improve user detection rates by providing error feedback. However, studies have shown that even with training, user vigilance declines over time (Bullée, Montoya, Junger, & Hartel, 2016; Kumaraguru, Sheng, Acquisti, Cranor, & Hong, 2010). Kumaraguru and colleagues (2010) have outlined shortcomings of a variety of anti-phishing training methods. They stress that educational exercises often fail to provide sufficient skills in the detection of work-specific threats. Instead, learners receive experience with generic types of threats they might come across in their workplace. The researchers show that incorporating training examples specific to an employee's job produces longer lasting improvements in the workplace.

Other research has shown that, once individuals are trained on specific threats, they often fixate on the details of training examples, limiting their ability to generalize to other hazards (Downs, Holbrook, & Cranor, 2006). Based on a series of interviews with computer users, Downs and colleagues determined that although participants were aware of threats to privacy and information due to their online behaviors, the many ways that a cyberattack can get through to them were not fully understood. Participants showed greater sensitivity to cues (e.g., broken images and spelling errors) that they had experience with, but showed no awareness for more nuanced approaches such as well-constructed emails posing as legitimate (e.g., as an email from a peer at a similar company). Additionally, anti-phishing training has been found to increase spam detection false positives (Sheng, Holbrook, Kumaraguru, Cranor, & Downs, 2010). That is, after learning which features can be used to detect fraudulent emails, trainees begin to misclassify legitimate emails when they judge these features to be present.

On the basis of this brief review, we conclude that efforts to educate email users to avoid potentially costly errors could benefit from a better understanding of how email features are mentally construed. That is, which message cues are salient in the minds of email users? How stable are these representations across variations in message features? Our goal in the current research was to gain some insight into these questions. Understanding the mental representation created by emails may help to predict human error and aid the design of more effective anti-phishing training.

1.1 | Study overview

In the current research, we attempted to measure the mental salience of several email features that are commonly found in spam and/or phishing emails. Spam emails are defined as unwanted advertising or promotional emails, whereas phishing emails are geared towards obtaining personal information. Some authors have categorized emails as either ham (legitimate work emails) or spam (illegitimate or unwanted emails) with subgroups (e.g., Trojan horse and phishing). For the current study, however, we do not distinguish between spam and phishing emails, including both in our stimulus set. Our goal was to understand which email dimensions receive the most attention and cognitive processing, and the dimensions we examined can be found in both legitimate and dangerous emails. The difference between spam and phishing emails is a finer grained distinction than we wished to examine at this stage of research.

In the exploratory study reported here, a group of participants provided feature ratings along several stimulus dimensions for a set of emails (details below). A separate group of participants completed a pairwise comparison task in which all pairs of emails were rated according to their similarity. These participants completed the similarity rating task twice: once with a set of spam emails containing prominent *Company logos* and a message urging that they immediately click a hyperlink (+Logos/Urgency set) and again with a set of spam emails that did not contain these features (−Logos/Urgency set). Our goal was to understand the mental representation of the stimuli by analyzing similarity data using multidimensional scaling (MDS), and we wished to see whether mental representation changed when prominent stimulus features varied across the two sets of emails. Responses from the feature rating task were then regressed onto the MDS coordinates to indicate which features best described the coordinate axes of each MDS space. In other words, the MDS space represents each stimulus as a point in a "psychological" space and the dimensions (i.e., axes) of the space need to be interpreted. The regression analysis indicates which rated features provide the best account of the MDS space dimensions).

Because we did not want participants focusing on whether each email was legitimate or not (i.e., ham or spam), they were informed that all emails they would see were considered spam. The dimensions that we varied across the stimulus sets—Logos and Urgency—were selected based on pilot feature rating data collected in our lab indicating these were among the most noticeable features (detailed below in Section 2.2). Furthermore, there is evidence in the literature suggesting that messages prompting an urgent response influence cognition (Vishwanath, Herath, Chen, Wang, & Rao, 2011). Vishwanath et al. (2011) found that urgency prompts in emails influence attention such that other clues to legitimacy (e.g., spelling errors) receive less attention and mental processing. They found that urgency cues were more strongly linked to phishing susceptibility than were source (e.g., URL) or grammatical error cues.

Finally, we expected emails containing *Company logos* to be more highly trusted than those without logos. Previous research has shown that familiarity with companies or brands can induce a sense of comfort, whereas emails from less popular or unknown companies raise more concern (Anggraeni, 2015). To assess this possibility, we included a trust rating measure to gauge differences in trust between the two stimulus sets used in the similarity rating task.

2 | METHOD

2.1 | Participants

University of Central Florida students participated in exchange for course credit after providing informed consent in accord with UCF IRB protocols. Participants were randomly assigned to either the feature rating task or the similarity rating task, and none completed both. Forty-one participants completed the feature rating task online, which took approximately 30–40 min. A separate sample of 22 participants completed the similarity rating task in our lab, which took approximately 40–45 min. Participants ranged from 18 to 21 years of age, and gender was split approximately evenly between males and females. No other demographic information was collected. Because the research was exploratory and there were no groups to compare (and thus no effect size to expect), power analyses were not conducted. Instead, sample sizes were based on previous research using similar methods (i.e., MDS analysis on stimulus similarity ratings) that inspired the current research (e.g., Markman & Makin, 1998, included 24 participants in a classification study using methods comparable with the current study; see also Bohil, Higgins, & Keebler, 2014). The sample size used in the similarity rating task (for analysis with MDS) was above the minimum found to be reliable for metric recovery in *Monte Carlo* simulations testing 2D, 3D, and 4D MDS models (Rodgers, 1991).

2.2 | Stimuli

The +Logos/Urgency and −Logos/Urgency stimulus sets each contained 30 email messages. These 60 emails were ultimately used in the feature rating and similarity rating tasks described below. However, these were selected from an initial set of 200 spam or phishing emails that were collected from multiple locations (e.g., spam folders and online searches). This larger collection of emails has been used previously by Sarno, Lewis, Bohil, and Neider (in press) and by Williams et al. (2019; see also Sarno, Lewis, Bohil, Shoss, & Neider, 2017).

2.2.1 | Preliminary feature selection

To provide some preliminary understanding of the dimensions characterizing email messages, two researchers in our lab (D. S. and J. L.) rated the presence or absence of 20 features for each of the 200 stimuli—including features that have appeared in the research literature and others simply hypothesized by the researchers (see Table 1 for the full list of features included in this preliminary evaluation). In a separate pilot study, the same 200 emails were rated by a group of participants (n = 21) on the degree to which they appeared to be "spam." Regressing our lab members' averaged feature ratings onto the average spam rating for each of the 200 emails revealed the most influential predictors of the spam

TABLE 1 Email features rated for study inclusion

Email feature	β (p value)	Description of email feature
Plausible premise	.37 (<.001)	Includes a sensible story or setup for action
Company logos	.19 (<.001)	Displays a graphical brand logo
Disproportionate benefit to recipient	−.17 (.001)	Promises an unusually large compensation
Urgent actionable links	.17 (.002)	Directs reader to immediately select a link taking them someplace else
Important downloadable content	.15 (.002)	Instructs to select link to download a file
Abnormal email structure	−.13 (.02)	Unusual message shape or spacing
Collecting personal information	−.11 (.02)	Requests private information (e.g., account #)
Advertisement	−.15 (.03)	Email promotes something
Account deletion or suspension	−.13 (.04)	Suggests account is frozen/deleted forcing reader to interact with email to correct it
Large images with clickable content	−.13 (.06)	Content is an image embedded with a link
Abnormal quantity of links	−.08 (.08)	Too many links are included in the email
Spelling or grammatical errors	−.06 (.21)	Text has spelling or grammatical mistakes
Money owed	−.05 (.25)	States reader has a bill or outstanding balance
Links to log in	.05 (.33)	Includes link to login page for a website
Awkward prose	−.02 (.70)	Message text includes odd phrasing
Security threat content	.02 (.72)	States that account details require an update or change of login information
Links to unsubscribe	−.01 (.75)	Includes link to a page alleging opt-out from future messaging
Uses company links	.01 (.78)	Links similar or identical to a company's links are present
Linked order numbers	−.003 (.95)	Presents an order number for alleged transaction with a link to another website
Requires a quick response	−.002 (.97)	Email urges a swift response

Note. Email features rated during preliminary stimulus evaluation (see Section 2.2.1 for details).

ratings, $F(20, 179) = 23.66$, $p < .001$, $R^2 = .73$, $R^2_{Adjusted} = .70$. Table 1 reports the β weights and p values for each of the 20 features from this preliminary assessment.

In the current research, the tasks of primary concern (detailed in Section 3) focused on a subset of seven comparatively influential features drawn from Table 1. The features included *Company logos* (the email displayed a graphical brand logo), *Urgent actionable links* (email directed the reader to act immediately by selecting a link that takes them someplace else, e.g., to a website), *Important downloadable content* (instructed the reader to select a link to download a file that in return will protect their computer), *Collecting personal information* (requested private information), *Advertisement* (email promoted something), *Account deletion or suspension* (suggested that an account is frozen/deleted due to lack of activity or unauthorized activity forcing reader to interact with the email to correct it), and *Large images with clickable content* (email content was an image embedded with a link).

This list is drawn from the most influential features—based on the regression p values—shown in Table 1. The seven dimensions listed above were retained for further study because they are relatively objective perceptual dimensions of email messages. We omitted from consideration three other relatively influential features appearing in Table 1—*plausible premise*, *disproportionate benefit to recipient*, and *abnormal email structure*—because they lacked a clear perceptual basis, and our goal was to evaluate the stability of attentional focus across changes in visible email features. More subjective email dimensions—including those omitted here—were examined in a separate study by Williams et al. (2019).

2.2.2 | Stimulus sets for the current study

On the basis of the researchers' ratings on these seven perceptual dimensions, we created two categories of emails: 30 emails rated as containing *Company logos* and *Urgent actionable links* (+Logos/Urgency), and 30 emails rated as having neither *Company logos* nor *Urgent actionable links* (−Logos/Urgency). This number was chosen to allow as many trials in the similarity comparison task (described below) as possible in a tolerable amount of time for participants. The selection of these two dimensions among the seven available dimensions was somewhat arbitrary, although consistent with previous research examining the influence of company logos (Anggraeni, 2015) and urgent actions (Vishwanath et al., 2011). Our primary goal was not to emphasize these dimensions, but rather to assess stability in the mental representation across variation in prominent stimulus features. Prevalence of the remaining five features was controlled to the extent that neither stimulus set had a statistically greater incidence of important downloadable content, collecting personal information, advertisement, account deletion or suspension, or large images with clickable content (p values > .11 for all t test results comparing the two email sets on researcher-based feature ratings).

3 | PROCEDURE

3.1 | Feature rating task

Participants who completed the feature rating task viewed a series of email images on their computer screen, along with rating questions (these participants did not complete the similarity rating task). Participants were asked to rate from one to seven (1 = clear absence of feature, 7 = clear presence of feature) on the level of five features in each email. These features included (a) *Important downloadable content*, (b) *Collecting personal information*, (c) *Account deletion or suspension*, (d) *Advertisement*, and (e) *Large images with clickable content*. No other instructions were provided in order to avoid influencing the ratings (e.g., no mention was made of the fact that all emails were spam or phishing messages).

Participants completed the task online via Qualtrics survey software. Participants were informed that the image would remain on screen until a rating response was selected for each feature and that the task was self-paced. Each email was condensed to fit a 430 × 520 pixel space (all emails remained legible at this resolution). Each trial consisted of a single email presented at center screen with a five-feature rating scale selection window directly below. Although the emails were presented in random order between participants, the listed order of features in the rating window remained fixed on every trial. The task took roughly 30 min on average to complete 60 trials (i.e., five features rated on each of the 30 stimuli from the +Logos/Urgency and −Logos/Urgency sets).

Table 2 displays average ratings for each feature in both stimulus sets. There were no differences between ratings between the stimulus sets on any dimension, except for a higher average rating on *Important downloadable content* in the −Logos/Urgency stimuli, $t(29) = 2.49$, $p = .02$. We do not examine this difference further, though, as the feature *Important downloadable content* does not appear to influence the similarity ratings summarized by the MDS results (as described in Section 4).

3.2 | Similarity rating task

A separate group of participants completed the similarity rating task (these participants did not complete the feature rating task), during

TABLE 2 Feature rating averages for each stimulus set

Feature	+Logos/ Urgency	−Logos/ Urgency
Important downloadable content	2.32 (.40)	2.67 (.22)
Collecting personal information	1.80 (.17)	2.02 (.51)
Account deletion or suspension	2.63 (.53)	2.74 (.36)
Advertisement	3.24 (2.23)	3.07 (1.80)
Large images with clickable content	2.52 (1.97)	3.04 (1.21)

Note. The mean for each feature derived from the feature rating task is shown with standard deviation in parentheses.

which they viewed two emails side-by-side on each trial and selected a number that best represented how alike (similar) they perceived the emails to be. An instructional screen was presented asking participants to indicate how similar they believed the presented emails to be from 1 to 9 (1 = very dissimilar, 9 = very similar). The task was carried out using an E-Prime program created by Hout, Goldinger, and Ferguson (2013). Participants who completed the similarity rating task completed the task twice: once for the +Logos/Urgency stimulus set and once for the −Logos/Urgency set. Participants completed all possible combinations of pairwise comparisons for 30 emails in each set for a total of 930 similarity ratings. This included each email paired with itself for a manipulation check. Figure 1 shows a screen capture of a trial from the similarity rating task.

Participants were informed that all the emails presented were flagged as spam. This was done because our goal was to explore how mental representation changes as a function of stimulus features, rather than as a function of whether the observer believed the stimulus was legitimate or not. The email resolution matched that of the feature rating task. The stimuli remained on screen until a response was selected via keyboard number keys. The intertrial interval was 1 s. Every 40 trials participants were allowed a self-paced break.

3.3 | Trust rating task

After providing similarity ratings for all stimulus pairs in each stimulus set, participants in the similarity rating task also completed a trust rating task, during which they were asked to rate the level of trust in each email on a scale from 1 to 5 (1 = no trust, 5 = completely trust). Each image from the similarity rating task was presented in random order along with the rating scale. Completing the similarity rating task for the two stimulus sets, along with the trust rating task for these stimuli, took approximately 40 min to complete.

4 | RESULTS

The steps in our data analysis consisted of performing MDS on the similarity rating data (separately for the +Logos/Urgency and −Logos/Urgency sets), followed by regression of results from the feature rating task onto the resulting MDS stimulus coordinates. We also compared trust ratings for the two stimulus sets.

4.1 | MDS analysis

An MDS analysis was performed on the similarity rating data. MDS is a dimensionality reduction method that creates a psychological proximity space given similarity rating responses. The distance between points in this space reflects the degree of psychological difference between stimuli as experienced by the observer (Borg & Groenen, 2005). The ALSCAL algorithm (Borg & Groenen, 2005) was used for the generation of an aggregate data MDS space (i.e., we averaged the similarity ratings across participants before entering them into the analysis). We limited our attention to two-dimensional MDS

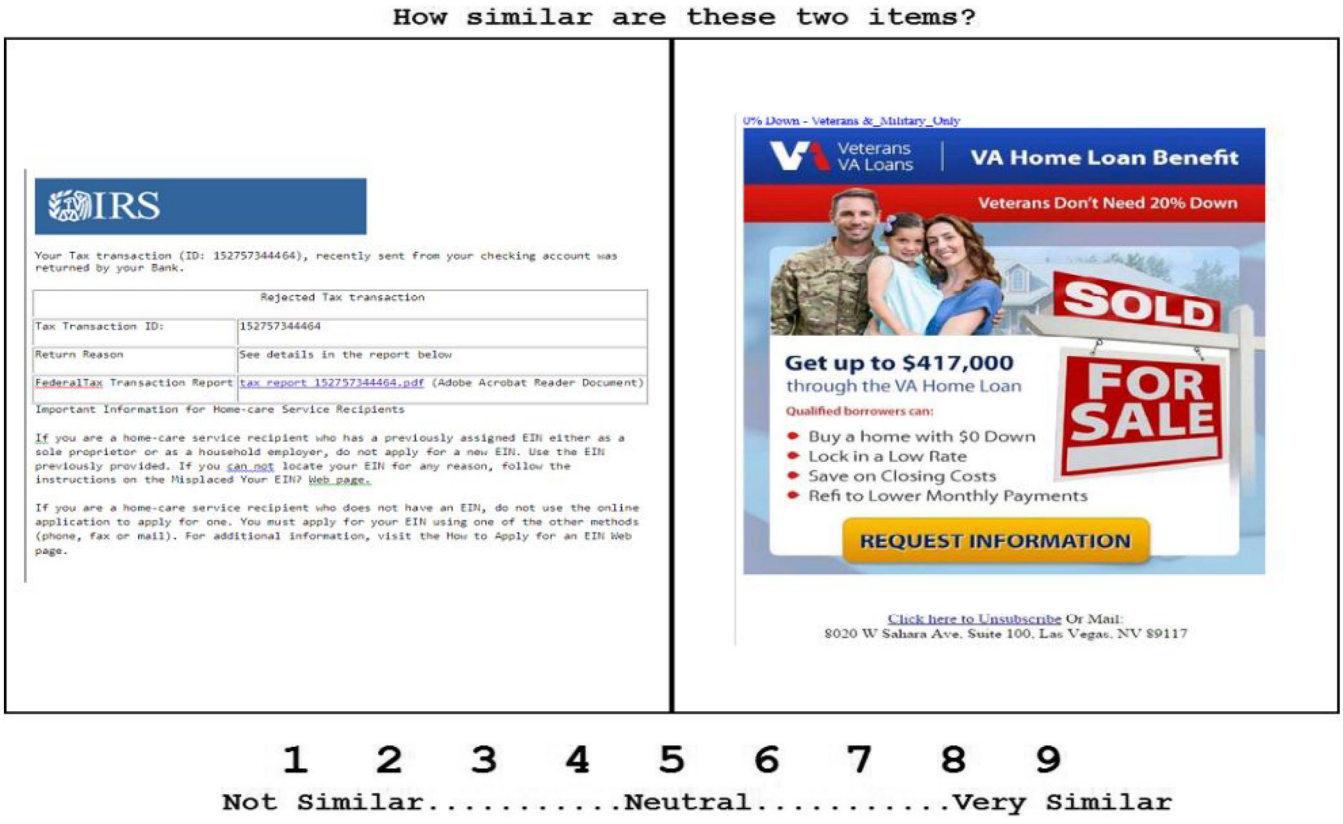

FIGURE 1 Screen capture of a trial in +Logos/Urgency condition [Colour figure can be viewed at wileyonlinelibrary.com]

spaces to simplify interpretation of the most prominent stimulus features. The +Logos/Urgency data fit (Young's S-Stress) was 0.16 with a relatively high average proportion of variance accounted for (R^2 = .87). The −Logos/Urgency data fit (Young's S-Stress) was 0.103 with a higher average proportion of variance accounted for (R^2 = .95).

Figure 2 shows the MDS space for each condition with a sample email from each quadrant of the resulting space. Figure 2a displays +Logos/Urgency analysis, and Figure 2b displays −Logos/Urgency analysis. To aid with interpretation of the most prominent dimensions as determined by the MDS analysis, we relied on the regression results described next.

4.2 | Regression of feature ratings onto MDS

The feature ratings were regressed onto the *x*, *y* coordinates of the two-dimensional MDS space for each stimulus set. A similar approach has been used in previous studies to link feature ratings to MDS space dimensions, thus providing additional clues as to interpretation of psychological dimensions underlying similarity ratings (Bohil et al., 2014; Kruskal & Wish, 1978; Markman & Makin, 1998).

The regression analysis showed a clear impact of *Advertisement* on the *x* axis (Dimension 1) for both +Logos/Urgency and −Logos/Urgency stimulus sets, whereas the *y* axis (Dimension 2) interpretation was less clear. According to the regression coefficients, several features were correlated with the *y* dimension for both sets, limiting clear interpretation of this dimension.

Further analysis revealed, however, a strong correlation between the features *Advertisement* and *Large images with clickable content*, r = .57, p < .01. As a result, we collapsed (averaged) the ratings along these two dimensions and re-ran the regression analysis using the following four feature rating dimensions: the new averaged advertising/large images dimension, along with *Important downloadable content*, *Collecting personal information*, and *Account deletion or suspension*.

This analysis revealed a much clearer dimensional interpretation. *Advertisement/Large images* was again significant for both conditions on Dimension 1. For the +Logos/Urgency set, β = −.70, $F(1, 25)$ = 41.91, p < .001, η_p^2 = .63, R^2 = .79, and $R^2_{Adjusted}$ = .75, and for the −Logos/Urgency set, β = −.78, $F(1, 25)$ = 108.80, p < .001, η_p^2 = .81, R^2 = .84, and $R^2_{Adjusted}$ = .82. As shown in Figure 2, emails displayed more or less apparent *Advertisement* features along the *x* dimension of the MDS spaces, regardless of stimulus set. Dimension 2 (the vertical dimension) was strongly driven by the dimension *Collecting personal information* for both stimulus sets. For +Logos/Urgency, β = −.72, $F(1, 25)$ = 7.60, p = .01, η_p^2 = .23, R^2 = .37, and $R^2_{Adjusted}$ = .26, and for −Logos/Urgency, β = .55, $F(1, 25)$ = 7.30, p = .01, η_p^2 = .23, R^2 = .25, and $R^2_{Adjusted}$ = .13. Thus, according to these analyses, observers consistently relied heavily on *Advertisement/Large images* and *Collecting personal information* features as a basis for their similarity judgments. The presence or absence of "Logos" and "Urgency" did not appear to produce any obvious psychological difference. It is important to point out that although the sign of regression coefficients reversed for *Collecting personal information* across sets, directionality in MDS analysis is somewhat arbitrary, with psychological similarity between emails represented by interpoint distances.

4.3 | Trust measure

After completing the similarity rating task, participants were shown every email individually and asked to rate the level of trust in each on a scale from 1 to 5. Because the similarity rating task was within-participants, we compared the trust ratings across stimulus sets (+Logos/Urgency vs. −Logos/Urgency) using a sign test. The results indicated that +Logos/Urgency (M = 2.75) trust ratings were significantly lower than for the −Logos/Urgency stimulus set (M = 3.62), Z = −4.08, p < .001.

5 | DISCUSSION

Our goal in this exploratory study was to determine the psychological representation of features commonly found in spam and phishing email messages. Participants compared several email messages in two sets—those with logos and urgency and those without—to provide data for MDS analysis for each set. A separate group of participants completed a feature rating task of the same stimuli along several stimulus dimensions. These ratings were regressed onto the 2D MDS space coordinates to provide psychological interpretation of the most influential stimulus dimensions underlying similarity ratings. Our analyses suggest that *Advertisement/Large images* and *Collecting personal information* were most salient in the minds of email observers, apparently receiving the most attention from participants during their similarity assessment of each pair of stimuli. It also appears that this allocation of attention is relatively unaffected by the presence or absence of two common features—*Company logos* and *Urgent actionable links*.

In addition, the trust measure that participants completed for each stimulus set shows that emails that included *Company logos* and *Urgent actionable links* were less trusted on average compared with the set of emails without *Company logos* and *Urgent actionable links*. This difference in trust concurs with previous findings demonstrating the influence of urgency. For example, Vishwanath et al. (2011) found that urgency cues strongly influence information processing (reducing attention to important clues to legitimacy) in the context of spam/phishing detection. Their findings are consistent with our interpretation of the current results: *Urgent actionable links* are a negative cue for trust. It would appear that the request for urgency in email messages has a stronger negative influence on trust than the positive influence of *Company logos* reported in previous studies.

Our conclusions in the current research are limited by a few design decisions. First, in the feature rating task—which produced values for regression onto the derived MDS spaces to aid in psychological interpretation of the dimensions—the features *Company logos* and *Urgent actionable links* were not included. Our goal was to evaluate psychological representation of visibly apparent dimensions in the presence

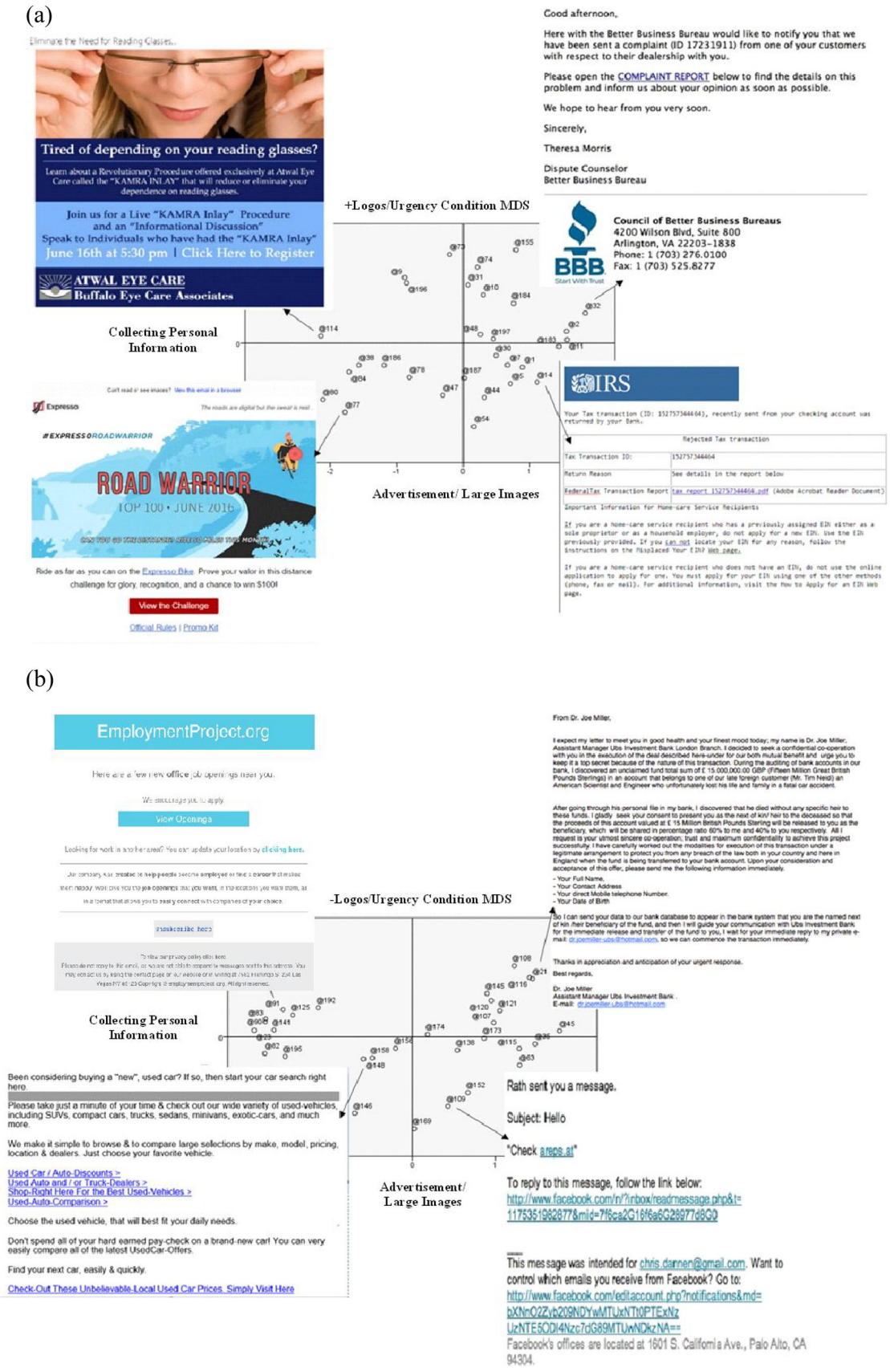

FIGURE 2 Multidimensional scaling (MDS) spaces for the (a) +Logos/Urgency set and (b) –Logos/Urgency set. Labeled points in plots represent emails. Distance between the points represent the degree of perceived similarity along the two most prominent dimensions (see text for details). An email from each quadrant has been superimposed for illustration [Colour figure can be viewed at wileyonlinelibrary.com]

of systematic variation between a subset of dimensions. Including these two features in the regression analyses should have resulted in substantially different regression coefficients along these two dimensions in the two stimulus sets. It would have been worthwhile to confirm this empirically. Also, the current design prevents us from evaluating the distinct contributions from *Company logos* and *Urgent actionable links*. Future studies may evaluate the influence of these two features by parsing them into different conditions. Our speculative conclusion is that *Urgent actionable links* were the primary influence among those two dimensions in the current study.

A question for further study is whether the same features would have been highly salient if the stimulus pool included legitimate non-spam emails in addition to spam. It is possible that perceptual weighting of email features differs when spam emails appear only occasionally. However, there is little information in the literature pertaining to understanding users' psychological representation of emails (spam or otherwise). Our goal in the current research was to identify which features receive the most attention in spam messages when there was no classification aspect (i.e., spam or not spam) to the task and to assess the stability of mental representation when some features are present or absent.

5.1 | CONCLUSIONS

There appears to be some stability in representation of emails across variation of prominent features—that is, advertisements and/or large images and collecting personal information were two prominent dimensions receiving attention from participants. This remained true regardless of whether company logos or cues to urgency were present or absent. Although prior research suggests that company logos increase trust in an email message, the presence of urgency cues may counteract this trust, as indicated by our finding of reduced trust in the presence of urgency cues. This coincides with earlier research showing an influence of urgency on cognitive processing. Perhaps urgency reduces attention to—or the credibility of—company logos. Although the current study controlled for the presence of company logos and urgency cues, in future studies, we plan to covary other features to determine the relative influence over attention of several features.

Our belief is that exploratory research of this type plays a valuable role in scientific research as it can generate new questions and hypotheses. For example, although the presence of advertising and large images seems to substantially draw the attention of participants, it is not clear that these features are critical to detection of spam/phishing emails when appearing along with legitimate email messages. Likewise, significant attention appears to be paid to whether personal information is being collected or not, which may be a compelling cue to potential danger. Future studies could focus on the predictive ability of these features on classification of spam and non-spam emails. If these features are found to be predictive, then perhaps the methods utilized in the current study would be valuable for exploring feature perception under a variety of task conditions (e.g., work settings and under time pressure) and perhaps even to individualize training.

Our primary goal in the current research was to shed light on the underlying perceptual representation of email messages. By determining which features are attended by observers over a heterogeneous set of emails, we may progressively understand and improve this aspect of cybersecurity.

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

FUNDING INFORMATION

No outside funding to report.

ORCID

Pooja Patel https://orcid.org/0000-0001-5358-9777

REFERENCES

Anggraeni, A. (2015). Effects of brand love, personality and image on word of mouth; the case of local fashion brands among young consumers. *Procedia-Social and Behavioral Sciences*, *211*, 442–447. Retrieved from. https://doi.org/10.1016/j.sbspro.2015.11.058

Bohil, C. J., Higgins, N. A., & Keebler, J. R. (2014). Predicting and interpreting identification errors in military vehicle training using multidimensional scaling. *Ergonomics*, *57*(6), 844–855. Retrieved from. https://doi.org/10.1080/00140139.2014.899631

Borg, I., & Groenen, P. (2005). *Modern multidimensional scaling: Theory and applications* (2nd ed.). New York: Springer-Verlag. Retrieved from. https://www.springer.com/us/book/9780387251509?gclid=CjwKCAjw67XpBRBqEiwA5RCocWfq8h3iwhoCbQYc4JqbDtiCiZ4k-q1IqcxRObIbAC5wBXaSmE0fmshoCIFEQAvD_BwE#otherversion=9781441920461

Bullée, J. W., Montoya, L., Junger, M., & Hartel, P. H. (2016, January). Telephone-based social engineering attacks: An experiment testing the success and time decay of an intervention. In *SG-CRC* (pp. 107–114). Amsterdam: IOS Press. https://doi.org/10.3233/978-1-61499-617-0-107

Clayton, R. (2004, July). Stopping Spam by Extrusion Detection. In CEAS. Retrieved from https://ht.transparencytoolkit.org/FileServer/FileServer/OLD%20Fileserver/conferenze%20e%20seminari/2004-07%20First%20Conference%20on%20Email%20and%20Anti-Spam%20(CEAS)/Stopping%20Spam%20by%20Extrusion%20Detection.pdf

Downs, J. S., Holbrook, M. B., & Cranor, L. F. (2006). Decision strategies and susceptibility to phishing. In Proceedings of the second symposium on Usable Privacy and Security. ACM. (pp. 79–90). https://doi.org/10.1016/10.1145/1143120.1143131

Hout, M. C., Goldinger, S. D., & Ferguson, R. W. (2013). The versatility of SpAM: A fast, efficient spatial method of data collection for multidimensional scaling. *Journal of Experimental Psychology: General*, *142*, 256–281. https://doi.org/10.1037/a0028860

Jung, J., & Sit, E. (2004). An empirical study of spam traffic and the use of DNS black lists. In Proceedings of the 4th ACM SIGCOMM conference on Internet Measurement. ACM. (pp. 370–375). https://doi.org/10.1145/1028788.1028838

Kruskal, J. B., & Wish, M. (1978) Multidimensional scaling. Sage University Paper Series on Quantitative Applications in the Social Sciences, No.

07-011, *Sage Publications*, Newbury Park. Retrieved from https://doi.org/10.4135/9781412985130

Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J. (2010). Teaching Johnny not to fall for phish. *ACM Transactions on Internet Technology (TOIT)*, *10*(2), 7. https://doi.org/10.1145/1753326.1753383

Markman, A. B., & Makin, V. S. (1998). Referential communication and category acquisition. *Journal of Experimental Psychology: General*, *127*(4), 331. Retrieved from–354. https://doi.org/10.1037/0096-3445.127.4.331

Rathod, S. B., & Pattewar, T. M. (2015, April). Content based spam detection in email using Bayesian classifier. In *2015 International Conference on Communications and Signal Processing (ICCSP)* (pp. 1257–1261). n/a: IEEE. https://doi.org/10.1109/ICCSP.2015.7322709

Rodgers, J. L. (1991). Matrix and stimulus sample sizes in the weighted MDS model: Empirical metric recovery functions. *Applied Psychological Measurement*, *15*(1), 71–77. Retrieved from. https://doi.org/10.1177/014662169101500107

Sarno, D. M., Lewis, J. E., Bohil, C. J., & Neider, M. B. (in press). Which phish is on the hook?: Phishing vulnerability for older versus younger adults. Human Factors. Retrieved from: https://doi.org/10.1177/0018720819855570

Sarno, D. M., Lewis, J. E., Bohil, C. J., Shoss, M. K., & Neider, M. B. (2017). Who are phishers luring? A demographic analysis of those susceptible to fake emails. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.*, *61*, 1735–1739. https://doi.org/10.1177/1541931213601915

Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. F., & Downs, J. (2010, April). Who falls for phish?: A demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 373–382). n/a: ACM. https://doi.org/10.1145/1753326.1753383

Tapper, J. (2017, August 1). White House officials tricked by email prankster. *CNN*. Retrieved from https://www.cnn.com/2017/07/31/politics/white-house-officials-tricked-by-email-prankster/index.html

Team, R. (2015). *Email statistics report, 2015–2019*. Palo Alto, CA, USA. Retrieved from: The Radicati Group. Inc. https://www.radicati.com/wp/wp-content/uploads/2015/02/Email-Statistics-Report-2015-2019-Executive-Summary.pdf

Vishwanath, A., Harrison, B., & Ng, Y. J. (2016). Suspicion, cognition, and automaticity model of phishing susceptibility. *Communication Research*, 0093650215627483, Retrieved from*45*, 1146–1166. https://doi.org/10.1177/0093650215627483

Vishwanath, A., Herath, T., Chen, R., Wang, J., & Rao, H. R. (2011). Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. *Decision Support Systems*, *51*(3), 576–586. Retrieved from. https://doi.org/10.1016/j.dss.2011.03.002

Wall, D. S. (2018). How big data feeds big crime. *Current History*, *117*(795), 29–34. Retrieved from. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3359972

Williams, S., Sarno, D., Lewis, J., Shoss, M., Neider, M., & Bohil, C. (2019). The psychological interaction of spam email features. *Ergonomics*. Retrieved from, *62*, 983–994. https://doi.org/10.1080/00140139.2019.1614681

How to cite this article: Patel P, Sarno DM, Lewis JE, Shoss M, Neider MB, Bohil CJ. Perceptual representation of spam and phishing emails. *Appl Cognit Psychol*. 2019;33:1296-1304. https://doi.org/10.1002/acp.3594