Personality and Individual Differences 186 (2022) 111335

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

The effects of personal values and message values on vulnerability to phishing

Avner Caspi [*], Maayan Sayag, Maya Gross, Zohar Weinstein, Shir Etgar

Department of Education and Psychology, Open University of Israel, 1st University Road, Raanana, Israel

ARTICLE INFO

Keywords:
Phishing
Human values
Personality
Social design

ABSTRACT

Phishing messages are designed to establish believability in the recipients, and to make them feel that the message is authentic. The current study examines the effect of personal values and of the values associated with the presumed sender on perceived authenticity of phishing messages. We tested two alternative hypotheses: (a) the value-congruency hypothesis contends that when the recipient's values match the underlying value of the message, the perception of authenticity increases; (b) the socially-shared values hypothesis suggests that messages that reflect the social image of the sender increase perception of authenticity. We further hypothesized that trust propensity moderates the effect of values on perception of phishing messages. Study 1 ($N = 624$) investigated and validated the values conveyed by bank messages. Study 2 ($N = 309$) tested the main hypotheses. Results supported the socially-shared values hypothesis. Trust propensity did not predict perceived authenticity of messages, and did not moderate the effect of the message value on perceived authenticity. The findings suggest that messages that fit the presumed sender's attributed values may be more risky, regardless of the receiver's values.

1. Introduction

Phishing is a criminal tactic by which people are persuaded by an online message to expose or transfer personal information to an unauthorized entity, mainly social security numbers, bank details, or passwords for online accounts (Jagatic, Johnson, Jakobsson, & Menczer, 2007; Lastdrager, 2014). Many studies have examined the effects of individual differences on susceptibility to phishing (Anawar, Kunasegaran, Mas'ud, & Zakaria, 2019; Kleitman, Law, & Kay, 2018; Lawson, Pearson, Crowson, & Mayhorn, 2020; Moody, Galletta, & Dunn, 2017). The current study aims to add to this line of research by testing the effects of personal values (Schwartz, 1992) on the detection of fraudulent messages.

The literature that examined individual differences in susceptibility to phishing shows that vulnerability to phishing associates with conscientiousness (Frauenstein & Flowerday, 2020), extraversion and openness (Pattinson, Jerram, Parsons, McCormac, & Butavicius, 2012), intelligence (Kleitman et al., 2018), self-control (Holtfreter, Reisig, Piquero, & Piquero, 2010), risk seeking (Moody et al., 2017), sensation seeking (Jones, Towse, Race, & Harrison, 2019), narcissism (Curtis, Rajivan, Jones, & Gonzalez, 2018), and the propensity to trust others (Wright, Chakraborty, Basoglu, & Marett, 2010). This long list of individual characteristics shows that vulnerability to phishing is a complex phenomenon.

People differ in their personal values, but the impact of values on vulnerability to phishing has not been explored thus far. Schwartz (1992, 2015; Schwartz et al., 2012) theorized that human values are desirable, stable, and broad goals that vary in importance and guide individuals' perception, judgement, and behavior. In its original version, the *Theory of Basic Human Values* identified ten broad values: Universalism, Benevolence, Conformity, Tradition, Security, Power, Achievement, Hedonism, Stimulation, and Self-direction. These values can be grouped into four higher order clusters of values with bipolar dimensions: self-enhancement versus self-transcendence, and openness to change versus conservation. The ten broad values are structured on a circle, such that values that are positively correlated are adjacent to one another, while values that are negatively correlated are located on an opposite place around the circle. Dozens of studies have provided empirical support to the circumflex structure of human values, its universality, and its predictive validity (Sagiv & Roccas, 2017). Fig. 1 presents the value circumflex along with a short description of each value.

[*] Corresponding author.
E-mail address: avnerca@openu.ac.il (A. Caspi).

https://doi.org/10.1016/j.paid.2021.111335
Received 31 August 2021; Received in revised form 7 October 2021; Accepted 9 October 2021

Several studies have shown that people attribute the ten values not only to themselves but also to brands and objects (Allen, 2002; Caspi, Etgar, & Kavé, 2021; Shepherd, Chartrand, & Fitzsimons, 2015; Torelli, Özsomer, Carvalho, Keh, & Maehle, 2012; Voorn, van der Veen, van Rompay, & Pruyn, 2018; Voorn, van der Veen, van Rompay, Hegner, & Pruyn, 2021; Zhang & Bloemer, 2008). Caspi et al. (2021) found that objects differ in their attributed value profiles. These profiles are socially agreed upon by members of social groups, with some cross-group differences. For example, people attribute values of conservation and self-transcendence to a washing machine and values of openness to change and self-enhancement to a television. Furthermore, evidence from consumer research shows that when one's personal values are congruent with the values that are attributed to a brand, the likelihood of purchasing that brand increases (Allen, 2002; Shepherd et al., 2015; Zhang & Bloemer, 2008). We suggest that susceptibility to phishing may also be determined by the relation between a person's values and the values conveyed by the phishing message.

Prior studies reported that the effect of personal characteristics on susceptibility to phishing is moderated by message features (Kim & Kim, 2013; Kleitman et al., 2018; Williams, Beardmore, & Joinson, 2017). Phishing messages are designed to establish believability, and to make recipients feel that the message is authentic (Wright, Jensen, Thatcher, Dinger, & Marett, 2014). To increase authenticity, phishing messages often contain cues of authority and urgency (Williams, Hinds, & Joinson, 2018), or attempt to establish a sense of relationship between the sender and the recipient. Tactics for creating a sense of relationship include reciprocity (e.g., implying that the sender has contributed to the recipient), social proof (e.g., contending that a given behavior conforms to the right social norm), consistency (e.g., implying that a given behavior is consistent with past behaviors), and scarcity (e.g., making an opportunity appear limited) (Wright et al., 2014). These features of phishing messages may correspond to personal values. For example, reciprocity behavior may be linked to values of self-transcendence (Universalism and Benevolence), and social proof may be associated with conservation (Conformity, Tradition, and Security). Furthermore, such associations between the message content and values may increase vulnerability to phishing.

It is possible that the congruency between specific features of the message and the recipient's values can affect the susceptibility to phishing. However, this matching may lead to two alternative predictions. First, when the values conveyed by the message fit the recipient's values, the recipient may believe the message, treat it as more authentic, and thus be more vulnerable to phishing. This notion rests on the assumption of the homophily tendency – the preference to gather with, attract to, and prefer people and activities that are similar to oneself (McPherson, Smith-Lovin, & Cook, 2001). A person who recognizes his or her values in a given message may have more faith in that message, whereas when the message emphasizes values that the person does not prioritize, his or her suspicion may increase.

An alternative possibility relies on the idea that messages can fit socially shared values rather than personal values. Indeed, brands have value profiles that reflect symbolic meanings, and these profiles are shared by groups (Aaker, Benet-Martinez, & Garolera, 2001; Caspi et al., 2021; McCracken, 1986; Solomon, 1986). Therefore, messages might be perceived as trustworthy if they convey values that match the socially shared "images" of the brands. For example, if a message attempts to impersonate a financial institute, it will express values that most people attribute to financial institutions, such as conformity and security. According to this hypothesis, recipients' personal values, whether congruent or incongruent with the message, will have no effect on the perception of phishing messages. Rather, messages that convey the brand's socially shared values will be perceived as authentic, whereas messages that reflect the opposing values will raise suspicion, and will fail to lure the recipient.

In addition to personal values, one's tendency to trust others may also affect susceptibility to phishing. Trust is defined as "a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another" (Rousseau, Sitkin, Burt, & Camerer, 1998, p. 395). Previous studies found that people who are high in their inclination to trust others are more vulnerable to phishing attacks. Wright et al. (2010) found that the disposition to trust is negatively correlated with phishing detection. Alseadoon, Othman, and Chan (2015) reported that trust predicted phishing susceptibility unlike other personality traits. However, some

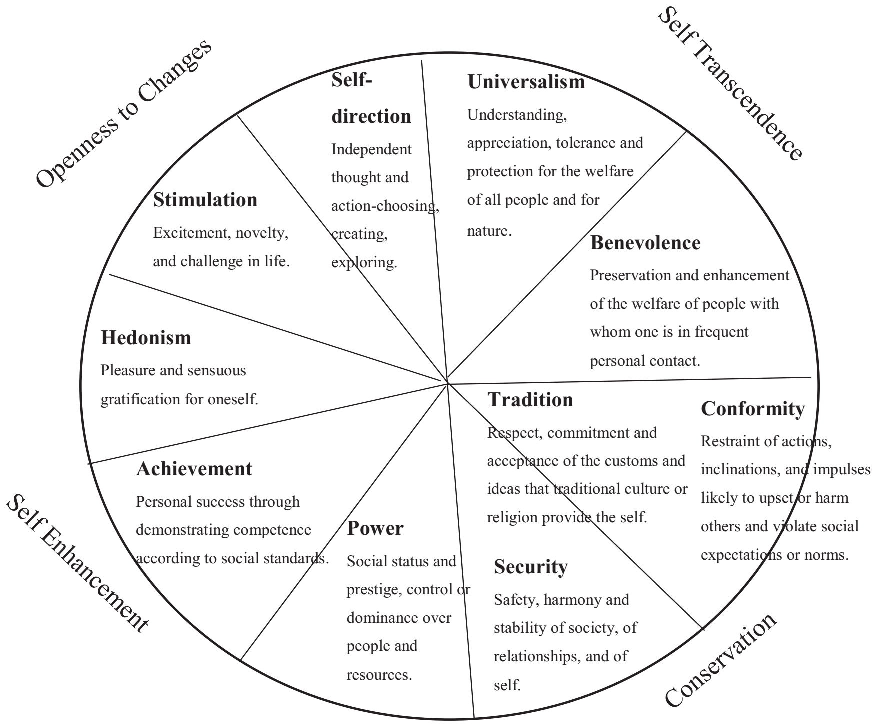
Openness to Changes
Self Transcendence
Self Enhancement
Conservation
Self-direction
Independent thought and action-choosing, creating, exploring.
Universalism
Understanding, appreciation, tolerance and protection for the welfare of all people and for nature.
Stimulation
Excitement, novelty, and challenge in life.
Benevolence
Preservation and enhancement of the welfare of people with whom one is in frequent personal contact.
Hedonism
Pleasure and sensuous gratification for oneself.
Tradition
Respect, commitment and acceptance of the customs and ideas that traditional culture or religion provide the self.
Conformity
Restraint of actions, inclinations, and impulses likely to upset or harm others and violate social expectations or norms.
Achievement
Personal success through demonstrating competence according to social standards.
Power
Social status and prestige, control or dominance over people and resources.
Security
Safety, harmony and stability of society, of relationships, and of self.

Fig. 1. Schwartz's values circumflex.

studies found that trust propensity did not predict phishing susceptibility (e.g., Moody et al., 2017; Wright & Marret, 2010), and thus the effect of trust requires additional empirical evidence, which the current study aims to provide.

Trust may moderate the effect of values on susceptibility to phishing. It has been shown that values of universalism are positively linked to interpersonal trust, whereas values of power are negatively related to it (Lönnqvist, Verkasalo, Wichardt, & Walkowitz, 2013). In addition, trust may moderate the effect of value congruency, such that among people who are high in trusting others the effect of value congruency may be more prominent than among people who are low in trusting others. The same moderating effect may also be found when phishing messages communicate the social values that recipients associate with the content or identity of the alleged sender of the message. Again, the effect of matching the social image of the sender is expected to be more prominent among people with high trust tendencies than among people with low trust tendencies.

The current study examines how personal values are related to phishing messages. We created messages that emphasized one of the ten values, and asked participants to select the most believable (i.e., authentic) messages as well as the least believable (i.e., distrustful) messages. We hypothesize that participants will choose messages that fit their value profile. For example, participants who endorse high levels of power are expected to view messages that emphasize power as believable and to view messages that emphasize universalism as unbelievable. Alternatively, we hypothesize that the selection of believable messages will reflect sensitivity to social norms rather than to personal values. For example, if participants perceive a bank as high in conformity and low in stimulation, they will view messages that reflect conformity as believable, and will mark messages that express stimulation as unbelievable. We further predict that interpersonal trust will moderate these effects.

We report two studies. In Study 1, we examined the content of ten bank messages, each conveying one of the ten values. We chose to test bank messages because most phishers pretend to represent financial institutions (Kim & Kim, 2013). Few studies measured the values that associate with banks (van Esterik-Plasmeijer & van Raaij, 2017; Voorn et al., 2021; Zhang & Bloemer, 2008) although they did not report specific value profiles. It must be noted that Herman, Heller, Cohen, Be'ery, and Lebel (2014) surveyed a representative Israeli sample and found a low level of trust in banks. The purpose of study 1 was to validate the intended value of each of the messages that we created. In Study 2, we tested our main hypotheses, by measuring personal values, believability of the ten messages, and trust propensity.

2. Study 1: values conveyed by messages

2.1. Methodology

2.1.1. Participants

Six-hundred and twenty-four participants were recruited through an Israeli Online Panel, each receiving 10 NIS (about 3$) for participation. The sample was balanced in terms of gender (50.4% male). Age ranged from 25 to 84 years old (mean = 43.9, SD = 13.2), 63% were married, 25% were single, and the rest reported another family status. All participants were Jewish, 70.2% identified as secular, 11.7% as traditional, 10.3% as orthodox, and 7.9% as ultra-orthodox.

2.1.2. Stimuli

2.1.2.1. Messages. Ten bank messages were created. They were short (between 40 and 60 words), and they began with a welcoming statement (“Dear customer”), and ended with an invitation to click on a link (“For further details click here”). The content of all messages involved financial issues, and they differed in the value that was highlighted. For example, the message that conveyed *Hedonism* read:

Dear customer,

Our bank offers dozens of bonuses that provide delightful experiences for you. Restaurants, spas, 1 + 1 cinema tickets with breakfast included, and many more. Join and enjoy your massage today!

For further details click here.

To increase content validity, five independent raters read the messages and suggested refinements before the message was finalized.

2.1.2.2. Measuring message values. For each message, participants were asked to rate the degree to which it reflected each of the ten values. We adapted the *Short and Broad Schwartz Value Survey (SVS)* questionnaire (Caspi et al., 2021; Sekerdej & Roccas, 2016; Oppenheim-Weller, Roccas, & Kurman, 2018). As in the original questionnaire, each value item consisted of the broad definition of that value type. For example, the item for security read “Safety, stability, and order. Caring for the family's security and its health”. Following Schwartz (1992), the scale ranged from −1 (the message reflects the opposite value), 0 (the message does not reflect the value), 3 (the message reflects the value), 6 (the message reflects the value to a large extent), to 7 (the message highly reflects the value).

2.1.2.3. Plausibility. We measured the plausibility that a bank will send the message to its customers, using a single item “How plausible is it that a bank will send this message to its costumers”. The scale ranged from 1 = “not at all” to 7 = “extremely plausible”.

2.1.2.4. Trustworthiness. The trustworthiness of the message was measured by a single item “To what extent is the message that you read trustworthy”. The scale ranged from 1 = “not at all” to 7 = “extremely trustworthy”.

2.1.3. Procedure

Participants were randomly assigned to one of the ten message conditions. They were presented with a list of the definitions of the ten values, they then read the message, and filled the value questionnaire. Following the rating of values, they rated plausibility and trustworthiness. Last, the participants answered demographic questions.

2.2. Results

Table 1 presents the average scores for every message on each of the ten values. For most messages, the highest score matched the intended value. For example, the highest score for the message that was created to reflect universalism was universalism. There were some messages for which the highest score was in the same higher-order dimension. For instance, the higher score for the message that expressed benevolence was universalism rather than benevolence, both comprising the self-transcendence value cluster. To confirm this observation, we looked for an interaction between message and rated value. We run repeated measures ANOVA, in which messages were a between-subject factor and values were a within-subject factor. The effect of messages was not significant, F (9, 614) = 1.610, p = .11, partial eta^2 = 0.023. That is, participants who rated different messages used similar ratings. More important, the effect of values was significant, F (9, 5526) = 14.147, $p <$ .001, partial eta^2 = 0.023. This effect suggests that participants rated the messages according to a shared social image of the sender. Fig. 2 illustrates the sender’s profile, showing high ratings of security and self-direction values, and low ratings of the tradition value. Post-hoc pairwise comparisons using Bonferroni correction revealed that security and self-direction received the highest score (p's < .05 for comparisons with most other values), and tradition received the lowest score (p's < .001 for comparisons with all other values). Last, the interaction between message and values was significant, F (9, 5526) = 14.075, p < .001, partial eta^2 = 0.171. To extract the interaction effect, we run ten separate repeated measures ANOVAs, one for each message (see Table 2 for

Table 1
Mean value scores by message (bold – highest score).

Message	N	Values									
		UN	BE	CO	TR	SE	PO	AC	HE	ST	SD
Universalism (UN)	65	**6.78**	5.32	5.23	4.28	5.48	4.42	4.77	3.75	4.34	5.77
Benevolence (BE)	61	**6.79**	6.31	5.33	4.93	5.39	3.92	4.57	4.13	4.38	5.33
Conformity (CO)	62	3.81	4.03	**4.71**	4.37	4.79	5.29	**4.71**	3.94	4.05	4.19
Tradition (TR)	61	**6.33**	6.25	6.00	5.90	6.02	4.31	4.80	4.79	4.51	4.85
Security (SE)	62	5.35	5.63	5.05	5.32	**6.53**	4.81	4.65	3.98	4.21	4.94
Power (PO)	65	3.54	4.28	4.37	3.82	4.43	**6.29**	5.35	5.34	4.77	4.82
Achievement (AC)	61	4.00	4.38	4.16	3.66	5.36	**6.13**	6.10	4.92	4.92	4.95
Hedonism (HE)	67	4.70	4.85	3.94	3.69	4.30	4.85	3.76	**6.48**	5.75	4.88
Stimulation (ST)	59	4.73	4.59	3.69	3.49	4.64	4.02	4.58	5.10	5.81	**6.86**
Self-direction (SD)	61	5.03	4.93	4.54	4.05	5.07	4.77	4.93	5.18	5.15	**5.59**
Total	624	5.10	5.05	4.70	4.34	**5.19**	4.89	4.82	4.88	4.79	**5.21**

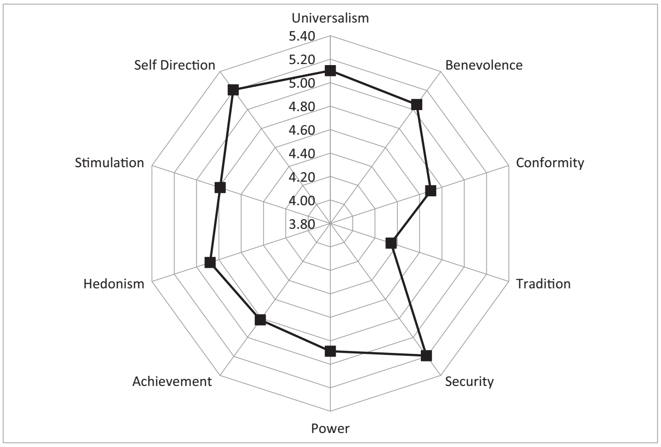

Fig. 2. Values profile: average values scores.

results). These analyses generally confirmed the observation that the value expressed in each message was recognized by the participants.

Next, we looked at plausibility and trustworthiness scores. First, we computed correlations between these two measures, and found that they were high for all messages (.631 ≤ r ≤ .801, all p's < .001). Second, we run one-way ANOVAs to test the effect of message on plausibility and on trustworthiness. Ratings of plausibility were significantly affected by message value, F (9, 614) = 2.529, p = .007, partial eta^2 = 0.036. Post-hoc pairwise comparisons using Bonferroni correction revealed that most differences were not significant. Messages that expressed security and benevolence received higher plausibility scores (see Fig. 3 for means and significant differences). Generally, participants believed the plausibility of all messages, yet their ratings were below the middle of the 7-point scale [grand mean = 3.78, SD = 1.87, t(623) = −2.96, p = .003, Cohen's d = 0.12].

Similarly, there was a significant effect of message for the trustworthiness scale, F (9, 614) = 2.312, p = .015, partial eta^2 = 0.033. However, post-hoc comparisons revealed only a marginal difference (p = .07) between ratings of trustworthiness for messages expressing conformity and of security. All messages received scores below the midpoint of the scale [grand mean = 3.40, SD = 1.82, t(623) = −8.72, p < .001, Cohen's d = 0.33]. Fig. 4 presents the mean scores.

2.3. *Discussion*

Study 1 aimed to validate the values that the messages conveyed. For most messages, the intended value was recognized. Furthermore, the profile of values that participants attributed to the messages may portray

Table 2
Repeated measures ANOVA for each message.

Message	F (df)	Partial eta^2	Pairwise comparison (mean difference)[a]									
			UN	BE	CO	TR	SE	PO	AC	HE	ST	SD
Universalism	17.600 (9, 576)**	0.216		1.46**	1.55*	2.51**	1.31*	2.37**	2.02**	3.03**	2.45**	1.02*
Benevolence	22.181 (9, 540)**	0.270	−0.48		0.98*	1.38**	0.92*	2.39**	1.74**	2.18**	1.93**	0.98**
Conformity	5.619 (9, 549)**	0.084	0.90*	0.68		0.34	−0.08	−0.58	0.00	0.77	0.66	0.52
Tradition	11.319 (9, 540)**	0.159	−0.43	−0.34	−0.10		−0.12	1.59*	1.10	1.12	1.39	1.05
Security	10.018 (9, 549)**	0.141	1.18*	0.90*	1.48**	1.21*		1.73*	1.89**	2.55**	2.32**	1.60**
Power	11.986 (9, 576)**	0.158	2.75**	2.02*	1.92*	2.48**	1.86*		0.94	0.95	1.52**	1.48*
Achievement	14.590 (9, 540)**	0.196	2.10**	1.72**	1.93**	2.44**	0.74	−0.03		1.18*	1.18*	1.15*
Hedonism	25.707 (9, 594)**	0.280	2.78**	2.63**	3.54**	3.79**	3.18**	2.63	3.72**		1.73**	2.60**
Stimulation	21.464 (9, 522)**	0.270	1.09	1.22	2.12**	2.32**	1.17	1.80**	1.24*	0.71		−1.05
Self-direction	3.002 (9, 540)*	0.048	0.56	0.66	1.05*	1.54*	0.53	0.82	0.66	0.41	0.44	

UN – Universalism; BE – Benevolence; CO – Conformity; TR – Tradition; SE – Security; PO - Power; AC – Achievement; HE – Hedonism; ST – Stimulation; SD – Self-direction.

[a] In each row we present the comparison of the intended value (i.e., the value highlighted in the message) with all other values. Positive difference reflects a higher score for the highlighted value.

* p < .05.

** p < .001.

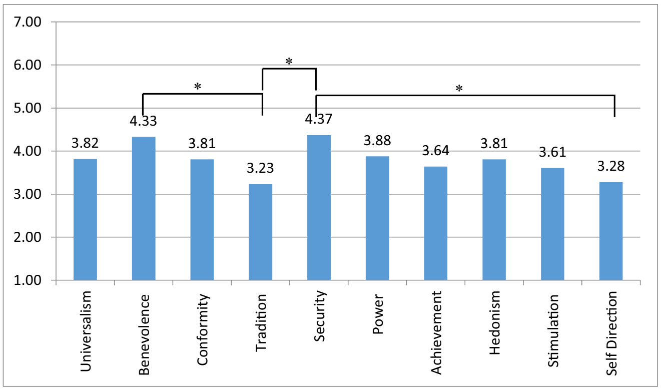

Fig. 3. Message plausibility ratings (* $p < .05$ for the difference between pairs of messages).

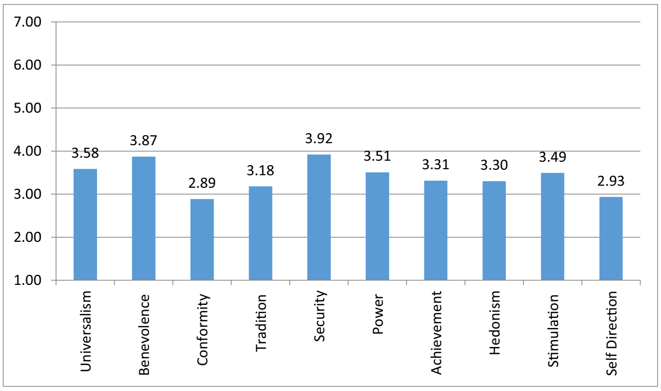

Fig. 4. Message trustworthiness ratings.

the social image of a bank. According to these findings, this image refers to values of security, self-direction, universalism, and benevolence. Given that this profile is based on a between-subjects design, in which each participants read only one message, it requires further testing. Furthermore, security and self-direction are opposing values, universalism seems less related to banks in general, and power was not scored among the highest values of the messages that we presented, contrary to what one would expect. Indeed, the current study was not designed to characterize the value profile of the message sender, and thus the results must be interpreted with caution.

As for believability, the data do not point to values that either increase or decrease message believability. Participants rated the plausibility that a bank would send these messages at an intermediate level, and rated message trustworthiness similarly. Thus, participants neither highly believed nor highly suspected any of the messages. We note that the task included no explicit instruction that might elicit suspicion. Furthermore, each participant read only one message, and could not compare messages.

3. Study 2: value congruency and phishing

3.1. Methodology

3.1.1. Participants

Three-hundred and nine members of an Israeli online panel (age range 18–73 years, mean = 40.6, SD = 13.54; 54.7% men) were paid 20 NIS (about $6) for participation in Study 2. All participants were Jewish, fluent Hebrew speakers, 86.7% were born in Israel, 55.7% were married, 31.1% were single, and the rest reported another family status. Sixty-five percent of the sample identified as secular, 22.7% as traditional, and the rest as either orthodox or ultra-orthodox.

3.1.2. Stimuli

3.1.2.1. Human values. Participants completed the 46-item version of the SVS (Schwartz, 1992), validated for cross-cultural use (Gandal, Roccas, Sagiv, & Wrzesniewski, 2005). In this questionnaire value items cover the ten different values described in Schwartz's theory. Each item was followed by a short explanatory phrase in parentheses [e.g., WEALTH (material possessions, money)]. Respondents rated the importance that they attributed to each item on a 9-point scale from “opposed to my values” (−1) through “not important” (0) to “of supreme importance” (7). We used the standard indexes recommended by Schwartz (1992) to measure the priority given to each of the ten values.

3.1.2.2. Categorizing messages as authentic or non-authentic. We used the ten messages from Study 1. Participants were requested to read all messages and to select two messages that they believed were the most reliable and authentic, and two other messages that they found to be the most unreliable and non-authentic. All messages were presented at once on a screen, but order of message presentation was random across participants. Half of the participants were asked to first choose the most authentic messages and then the non-authentic ones; the other half were asked to choose the non-authentic messages first. No message could be selected as both authentic and non-authentic. A message rated as authentic was scored 1, a message rated as non-authentic was scored −1, and a message that was rated as neither authentic nor non-authentic received a score of 0.

3.1.2.3. Trust. We used Yamagishi and Yamagishi’s (1994) questionnaire to measure participants' beliefs about honesty and trustworthiness in others. The questionnaire includes six statements such as “Most people are basically honest”. The scale ranged from 1 = definitely disagree to 5 = totally agree. Cronbach's alpha was 0.85.

3.1.2.4. Information about the existence of bank accounts. Since the message was presumably sent by a bank, we assumed that having more than one bank account may moderate the impact of its content (i.e., its value) on its perceived authenticity. To control for this possibility, participants were asked whether they held several bank accounts in different banks. If they answered yes, they were asked how many additional bank accounts they had (e.g., two, three, four, or above four).

3.1.2.5. Financial distress. Being in financial distress may increase the likelihood of being lured by dubious messages. This possible predictor of phishing victimhood has not received much research attention. To control for a possible influence of financial distress on detecting phishing, participants were asked to what extent they felt financial distress in the last weeks, using a five-point scale from 1 = rarely to 5 = almost always.

3.1.2.6. Phishing awareness. Participants were asked if they were familiar with the definition of phishing (24.3% answered that they were unfamiliar with its meaning). Those who were familiar with phishing (75.7%) were asked if they were afraid of becoming a victim of a phishing attack, and if they ever experienced a phishing attack. We assumed that awareness to phishing, and even more so experiencing phishing, would interact with participants' judgement of the messages.

3.1.3. Procedure

The study had two parts. The first part lasted about 20 min, and included the SVS (Schwartz, 1992). To avoid carry-over from reporting personal values to the message selection task, participants were invited to complete a second questionnaire after a few days. This part lasted approximately 15 min and included a short explanation about phishing, the message selection task, and questions about trust, bank accounts, financial distress, and phishing awareness.

3.2. Results

Table 3 presents descriptive statistics (means, SDs, and correlations) of all the studied variables. We found that conformity associated positively with being afraid of a phishing attack ($r = .133$). Other personal values did not correlate with this measure. Only 23 participants reported that they had been victims of a phishing attack. Universalism correlated positively ($r = .149$) and achievement correlated negatively ($r = -.131$) with being a victim of phishing.

The findings show clearly that there is no congruency between the values that one endorses as personal values and the values of the messages that were considered authentic or non-authentic, with only one exception for the value of hedonism. To test differences in message categorization, we analyzed the data in two complementary approaches. First, we tested differences in classifying messages as either authentic (or believable) or as non-authentic (or suspicious) with the use of a frequency analysis. This analysis treated each message separately (see Table 4). Messages that conveyed tradition, power, hedonism, and benevolence were selected more often as authentic, whereas messages that conveyed stimulation and self-direction were selected more often as non-authentic. Findings for the message that conveyed benevolence replicated the plausibility measure used in Study 1. Ratings of all other values differed from the results of Study 1, and ratings of the message conveying self-direction contradicted the findings of Study 1.

Second, we ran a repeated-measures ANOVA using the scores for each message (means and SDs are presented in Table 3). This analysis compared categorization of each message to all other messages. The difference between messages was significant, $F\ (9, 2772) = 11.922$, $p < .001$, partial eta^2 = 0.037. Post-hoc pairwise comparisons using Bonferroni adjustment revealed significant differences between messages, as shown in Fig. 5. The most authentic messages according to this analysis were those expressing tradition, power, hedonism, and benevolence. The least authentic messages were those reflecting stimulation and self-direction, corroborating the previous analysis.

Next, we conducted ten separate regression analyses to examine the moderating role of trust. In each regression the predictor was a personal value (e.g., power) and the dependent variable was the authenticity score of the corresponding message (e.g., the message that conveyed power). Trust was the moderator. We used PROCESS version 3.4.1 macro for SPSS to run these analyses. The results are presented in Table 5. Only the interaction between personal self-direction and trust reached significance. This interaction is presented in Fig. 6. The results contradict our predictions. Among participants who were *high* on trust, those who endorsed high levels of self-direction *suspected* the message that conveyed self-direction more than did participants who endorsed low levels of self-direction. Among participants who were low on trust there was no effect. Note that the message that conveyed self-direction was categorized as non-authentic relatively often, and as the interaction reveals, this tendency emerged particularly in participants who endorsed high levels of self-direction.

In the next analysis, we tested the moderating effect of trust on categorizing messages, with the assumption that messages would be more believable when they reflected the social image of banks. To test this hypothesis we ran a repeated-measures ANOVA, using message categorization as a within-subject factor and trust as a between-subject factor. We divided participants into two groups: below the median on the trust scale (median = 3.50, $n = 151$) and above the median ($n = 158$).

The interaction effect was not significant, $F\ (9, 2763) = 1.030$, $p = .413$, partial eta^2 = 0.003. Fig. 7 presents the results. Adding number of bank accounts, financial distress, and phishing awareness as covariates did not change this result. In sum, trust did not moderate the effect of value congruency or the effect of message value.

3.3. Discussion

Participants in Study 2 suspected the authenticity of a message when it denoted certain values that did not conform to the values of the message sender, regardless of their personal values. Messages that conveyed benevolence, tradition, power, or hedonism seemed authentic, whereas messages that conveyed stimulation and self-direction did not seem authentic. The value of hedonism showed a personal value congruency effect that did not emerge for the other values. Although personal values associated with some variables that affected the degree of believability, personal values did not predict message categorization as either authentic or non-authentic.

Our findings show further that benevolence correlated positively with trust propensity and power correlated negatively with it. These correlations are similar to Lönnqvist et al.'s (2013) results, except that Lönnqvist et al. found positive correlations between universalism and trust rather than between benevolence and trust. Nevertheless, both values encompass the same higher-order value dimension, namely self-transcendence. In addition, we found that participants who were high on conformity, were also more afraid of phishing attacks. Lastly, participants who were high on universalism, or low on achievement, reported greater frequencies of experiencing actual phishing. Again, despite these correlations, personal values did not affect one's categorization of messages as either authentic or suspicious.

Trust propensity did not predict perceived message authenticity and did not interact with personal values in predicting message categorization. The null effect of trust replicated some prior studies (Moody et al., 2017; Wright & Marret, 2010), but contradicted others (Alseadoon et al., 2015; Wright et al., 2010). A recent study shows that credulity – the *willingness* to believe in someone or something in the absence of reasonable proof, and not general trust – the *actual* confidence in others, may increase older adults' vulnerability to fraud (Shao et al., 2019). Further research will have to test this distinction.

4. General discussion

Several individual differences may influence susceptibility to phishing. We predicted that personal values might also affect this susceptibility. Our results show that conformity associates with worrying about phishing attacks, and that universalism and achievement associate with being a victim of phishing. However, personal values did not relate to the degree to which people believed in message authenticity. Instead, the results highlight the importance of message characteristics (Williams et al., 2018; Wright et al., 2014) and illustrate the importance of the value profile of the alleged brand in generating the authenticity of the message.

While most participants know what phishing *is*, very few reported that they encountered such an attack. We do not know whether this low rate of self-reported actual susceptibility represents true rates in the population, implying that very few people actually transfer money or expose personal details to online imposters, or whether it reflects the fact that only few people actually receive and notice phishing messages. Previous research has suggested that low exposure to email-delivered phishing attacks results in relatively low accuracy in reporting malicious messages (Sawyer & Hancock, 2018). Advances in automatic filtering-out of phishing messages reduce such exposure, thereby paradoxically decreasing the ability to identify fraud and increasing vulnerability. Our study raises the possibility that people who endorse high levels of conformity and high levels of achievement are more immune to this 'prevalence paradox', and are inherently more suspicious. On the other hand, people who endorse high levels of universalism may tend to trust others more, and to be exploited more often.

The findings of this study suggest that a believable message is a message that aligns with the social image of the source. According to Hovland, Janis, and Kelley (1953), people are more likely to accept a message when the source presents itself as credible. This tendency may

Table 3
Means, SDs, and correlations of all variables in Study 2. [a]

	Mean	SD	Gender	Age	Trust	Accounts	Financial distress	Fear of phishing	Victim of phishing
Gender (1 = women, 2 = men)									
Age	40.62	13.55	.022						
Trust	3.38	0.66	-.026	.139*					
Number of bank accounts	1.29	0.53	.095	.167**	.026				
Financial distress	2.32	1.10	-.048	-.115*	-.086	-.088			
Fear of phishing (0 = no, 1 = yes; N = 234)	54.7% yes		.026	.106	-.015	.093	-.009		
Victim of phishing (0 = no, 1 = yes; N = 234)	7.4% yes		.013	.018	.050	.054	.050	.109	
Personal value:									
Universalism	4.57	0.62	**-.158****	.097	-.002	**-.114***	.066	.028	**.149***
Benevolence	5.04	0.65	-.061	.017	**.181****	.029	.009	.128	-.046
Conformity	4.90	0.62	-.005	.028	-.075	.028	.052	**.133***	-.080
Tradition	3.76	0.98	.010	**-.143***	.069	-.046	.032	-.059	-.118
Security	5.10	0.70	-.027	**.241****	.035	.071	-.042	-.052	.058
Power	3.12	1.20	**.194****	**-.118***	**-.113***	.054	-.094	-.092	.058
Achievement	4.60	0.70	-.001	**-.138***	-.047	.069	**-.193****	.037	**-.131***
Hedonism	4.06	1.21	.042	.052	-.111	.061	-.026	-.083	-.024
Stimulation	3.99	1.08	**.149****	-.060	-.021	-.004	.087	-.067	.065
Self-direction	4.90	0.59	-.074	.057	.045	-.073	.083	.071	.013
Message categorization:									
mUniversalism	-0.00	0.59	-.071	.060	-.026	-.018	-.084	.056	.013
mBenevolence	0.10	0.63	.011	-.074	.053	-.029	-.084	-.014	-.156*
mConformity	-0.06	0.58	.022	.068	.032	-.064	.078	-.139*	.123
mTradition	0.17	0.67	-.082	.055	-.043	.017	.067	.010	-.046
mSecurity	-0.01	0.59	-.077	.083	.010	.006	.003	.048	.024
mPower	0.13	0.68	.087	-.041	.061	-.015	-.004	.054	-.102
mAchievement	-0.06	0.53	-.094	-.029	.003	.041	-.117*	.181**	.030
mHedonism	0.12	0.68	.166**	-.065	-.007	.076	-.011	-.038	.050
mStimulation	-0.17	0.62	.021	-.074	-.035	.014	.076	-.085	.041
mSelf-direction	-0.21	0.63	-.021	.031	-.049	-.033	.054	-.056	.053

	UN	BE	CO	TR	SE	PO	AC	HE	ST	SD
Personal value:										
Benevolence	-.065									
Conformity	-.230**	.174**								
Tradition	-.270**	.005	.190**							
Security	-.060	-.046	.204**	-.039						
Power	-.341**	-.456**	-.152**	-.012	-.149**					
Achievement	-.168**	.162**	-.232**	-.258**	-.338**	.000				
Hedonism	-.238**	-.393**	-.165**	-.293**	-.146*	.289**	-.044			
Stimulation	.013	-.318**	-.388**	-.323**	-.374**	.175**	.051	.282**		
Self-direction	.067	.105	-.194**	-.407**	-.060	-.310**	.134*	-.033	.022	
Message categorization:										
mUniversalism	.021	.083	.017	-.005	-.041	-.076	-.007	.031	-.072	.053
mBenevolence	.062	.020	.003	-.**115***	.011	-.**130***	.069	.016	-.029	.**152****
mConformity	-.001	-.017	-.037	-.067	.091	.027	.042	.058	-.026	-.062
mTradition	-.067	.050	.048	.083	-.046	-.045	.037	-.024	.069	-.**132***
mSecurity	.092	.096	.030	-.077	-.043	-.057	.058	-.079	-.002	.027
mPower	-.098	-.005	.006	.049	-.015	.073	-.015	-.033	.022	.008
mAchievement	.062	-.044	-.099	-.087	-.082	.042	.079	.054	.003	.**113***
mHedonism	-.093	-.**134***	-.081	.077	-.015	.106	-.071	.**160****	.090	-.053
mStimulation	.069	-.089	-.055	.039	.045	.046	-.069	-.042	.002	-.043
mSelf-direction	-.010	.051	.**155****	.065	.088	.004	-.101	-.**141***	-.077	-.030

	mUN	mBE	mCO	mTR	mSE	mPO	mAC	mHE	mST
mBenevolence	-.008								
mConformity	-.114*	-.080							
mTradition	-.195**	-.263**	-.080						
mSecurity	-.141*	-.051	-.087	-.030					
mPower	-.112*	-.144*	-.134*	-.048	-.136*				
mAchievement	-.022	-.128*	-.129*	-.044	-.127*	-.149**			
mHedonism	-.080	-.050	-.137*	-.114*	-.251**	-.180**	-.062		
mStimulation	-.125*	-.138*	-.138*	-.217**	-.092	-.055	-.022	-.083	
mSelf-direction	-.134*	-.137*	-.037	-.094	-.021	-.158**	-.157**	-.142*	-.124*

[a] * p < .05; ** p < .001. UN – Universalism; BE – Benevolence; CO – Conformity; TR – Tradition; SE – Security; PO - Power; AC – Achievement; HE – Hedonism; ST – Stimulation; SD – Self-direction; Message values are preceded by m

Table 4
Distribution of messages categorization, and McNemar test results.

Value	Categorized as authentic	Categorized as non-authentic	Not categorized	McNemar Chi^2 test
Universalism	53	54	202	0.000
Benevolence	**78**	47	184	7.200*
Conformity	43	63	203	3.406
Tradition	**100**	48	161	17.574**
Security	52	54	203	0.009
Power	**94**	54	161	10.277**
Achievement	34	53	222	3.724*
Hedonism	**91**	55	163	8.390*
Stimulation	38	**91**	180	20.961**
Self-direction	35	**99**	175	26.619**

Majority is presented in bold.

[*] $p \leq .05$.

[**] $p < .001$.

explain why phishers impersonate credible institutions, such as banks. To earn credibility, the message should be compatible with the shared social knowledge attributed to the source. We used bank messages in a setting that raises suspicions. Banks are credible institutions, but as our results show, to elicit believability or to invoke suspicion, the message must be congruent with the bank's social image or incongruent with this image. Put it in Hovland et al.'s (1953) exact words, "the degree of confidence in the speaker's intent to communicate the assertions he considers most valid" (p. 21) must not be violated to establish trustworthiness in the message.

Study 1 and Study 2 have portrayed a somewhat different social profile of banks. These portrays may result from two key differences between the studies: the design and the level of suspiciousness. As for design, the first study used a between-subject design whereas the second used a within-subject study. As such, only the second study involved comparisons between messages when the variance that resulted from participants was controlled for. Suspiciousness was explicitly stated only in the second study, which may also contribute to the differences in the results of the two studies. To describe the social value profile of banks,

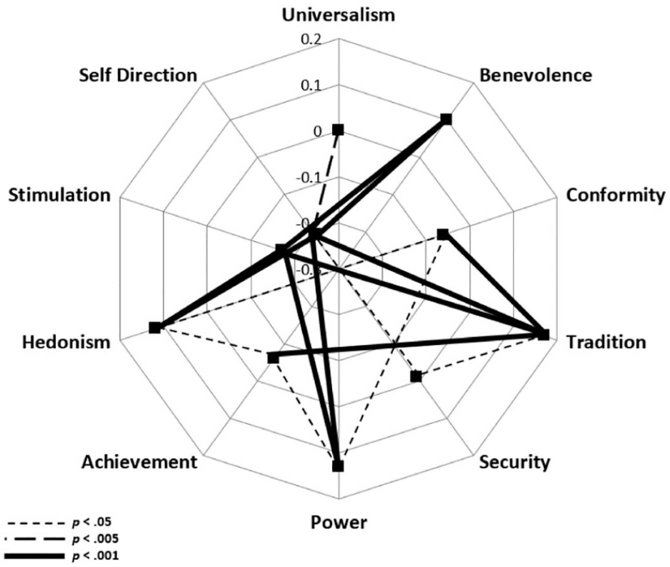

Fig. 5. Mean message categorization (score above 0 denotes that the message was categorized as authentic more often than as non-authentic; score below 0 denotes that the message was categorized as non-authentic more often than as authentic). Lines represent significant differences.

further studies are required that will be designed to test this issue directly.

Our findings provide no evidence in support of the value-congruency hypothesis. The single positive correlation between a personal value and perceived message authenticity was found for the value of hedonism. Indeed, people who endorse high levels of hedonism appreciate pleasures or sensuous gratification for themselves. Banks may *afford* hedonistic experiences, but it would be difficult to argue that banks *reflect* hedonistic values. Prior studies that measured the values that associate with banks (van Esterik-Plasmeijer & van Raaij, 2017; Voorn et al., 2021; Zhang & Bloemer, 2008) did not report specific value profiles. Furthermore, it is possible that different social groups or cultures may attribute somewhat different values to the same brand (Caspi et al., 2021). Thus, further studies of the value profiles of banks are required to address this finding.

4.1. Limitations

Measuring phishing by classifying messages as authentic versus non-authentic is very common in studies of phishing. Nevertheless, Jones et al. (2019) argue that this method has some limitations. Responding to a phishing message involves a series of actions that differs in their level of vulnerability: Opening the message may be risky if the message contains an embedded image which is automatically downloaded; clicking on a link within a message may increase the level of vulnerability; and disclosing personal details on a linked webpage is even more dangerous. A simple binary categorization of authenticity does not distinguish between each of these behaviors and may in fact inflate detection rates (Parsons, McCormac, Pattinson, Butavicius, & Jerram, 2013). However, recent work by Hakim et al. (2021) shows that phishing emails that were more tempting in the real world were rated as more suspicious by participants in the lab.

Furthermore, in the current study all messages were presented together and the participant was asked to select among them. This procedure provides a partial simulation of real-world situations, in

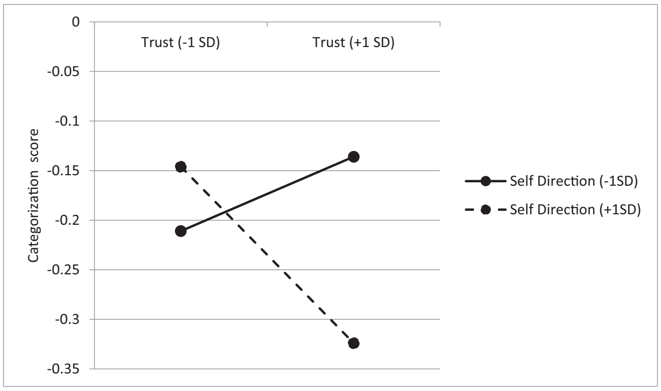

Fig. 6. Self-direction by Trust interaction.

Table 5
Regressions assessing the moderating role of trust on message categorization.

Message	R^2	F(3,305) =	Value			Trust			Interaction		
			Coefficient (SE)	*t*	95%CI	Coefficient (SE)	*t*	95%CI	Coefficient (SE)	*t*	95%CI
Universalism	0.003	0.301	0.021 (0.054)	0.349	−0.085, 0.128	−0.023 (0.051)	−0.463	−0.123, 0.076	−0.061 (0.081)	−0.748	−0.220, 0.099
Benevolence	0.004	0.434	0.011 (0.056)	0.196	−0.100, 0.122	0.049 (0.055)	−0.894	−0.059, 0.157	−0.052 (0.082)	−0.633	−0.214, 0.110
Conformity	0.008	0.785	−0.023 (0.054)	−0.415	−0.129, 0.084	0.032 (0.050)	0.639	−0.067, 0.131	−0.111 (0.087)	−1.286	−0.282, 0.059
Tradition	0.020	2.090	0.057 (0.039)	1.473	−0.019, 0.134	−0.037 (0.058)	−0.635	−0.151, 0.077	0.098 (0.053)	1.843[a]	−0.007, 0.202
Security	0.002	0.224	−0.037 (0.048)	−0.773	−0.132, 0.058	0.010 (0.051)	0.197	−0.090, 0.110	−0.020 (0.076)	−0.269	−0.170, 0.120
Power	0.010	1.063	0.046 (0.032)	1.407	−0.018, 0.110	0.073 (0.059)	1.232	−0.043, 0.189	−0.101 (0.045)	−0.212	−0.099, 0.079
Achievement	0.007	0.711	0.062 (0.043)	1.436	−0.023, 0.148	0.004 (0.045)	0.091	−0.085, 0.094	0.027 (0.062)	0.429	−0.096, 0.149
Hedonism	0.026	2.736*	0.090 (0.032)	2.808**	0.027, 0.153	0.010 (0.058)	0.165	−0.105, 0.124	0.018 (0.047)	0.372	−0.076, 0.111
Stimulation	0.004	0.418	−0.001 (0.033)	−0.016	−0.066, 0.065	−0.032 (0.054)	−0.601	−0.138, 0.073	0.047 (0.050)	0.943	−0.051, 0.146
Self-direction	0.017	1.746	−0.044 (0.061)	−0.727	−0.165, 0.076	−0.046 (0.054)	−0.866	−0.152, 0.059	−0.163 (0.079)	−2.061*	−0.318, 0.077

[a] $p < .07$.
* $p < .05$.
** $p < .005$.

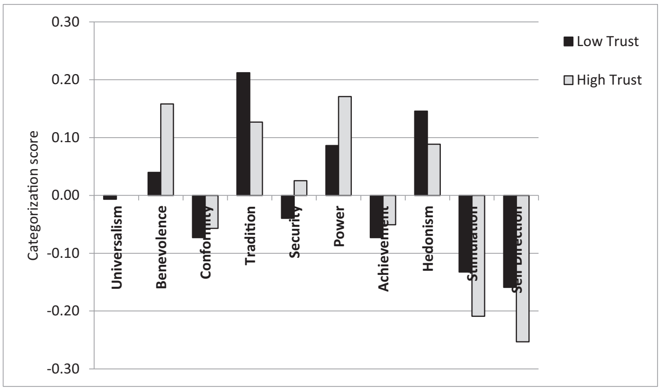

Fig. 7. Interaction between values and trust.

which phishing messages appear unexpectedly, among hundreds of credited messages that raise no suspicion, and without any context or reference that allow deliberate comparisons. Individuals who are aware that they are taking part in a phishing study may be more suspicious, and this awareness may result in a bias toward 'phishing' decisions (Anandpara, Dingman, Jakobsson, Liu, & Roinestad, 2007).

Our attempt to generate messages that reflected a single and recognizable value (Study 1) was generally successful. However, the construction of messages was not perfect, and the tradition message was not recognized as reflecting the intended value. Thus, the conclusions about the luring effect of a tradition message sent from a bank should be cautious.

Last, all messages in the current study were from banks, although other institutions may also be used in phishing messages, such as government offices, public agencies, or private institutions. We believe that users might be misled by messages that emphasize the values that associate with the sender, regardless of its identity.

4.2. Summary

We found that participant believe in messages that reflect the socially attributed values that associate with the institution that allegedly sent the message. We suggest that these messages may be more dangerous. Personal values did not interact with attributed values in this judgement, therefore no value congruency between the message content and one's values was evident. This finding may contribute to understanding the success of phishing attacks, since the phishers need not customize their message to a specific audience. Personal trust did not influence message believability, and did not moderate the effect of message values. People who endorsed high levels of conformity and achievement may take more cautionary steps against phishing, while people who endorse high levels of universalism might be more vulnerable.

CRediT authorship contribution statement

Avner Caspi: Conceptualization, Methodology, Formal analysis, Writing- Original draft preparation. **Maayan Sayag**: Methodology, Software, Formal analysis. **Maya Gross**: Methodology, Investigation. **Zohar Weinstein**: Methodology, Validation: **Shir Etgar:** Writing- Reviewing and Editing.

Acknowledgement

This work was supported by the Blavatnik Interdisciplinary Cyber Research Center [Grant #0610114651].

References

Caspi, A., Etgar, S., & Kavé, G. (2021). *Attributing Values to Devices*. Manuscript submitted for publication.

Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007). Social phishing. *Communications of the ACM, 50*(10), 94–100.

Lastdrager, E. E. (2014). Achieving a consensual definition of phishing based on a systematic review of the literature. *Crime Science, 3*(1), 1–10. https://doi.org/10.1186/s40163-014-0009-y

Anawar, S., Kunasegaran, D. L., Mas'ud, M. Z., & Zakaria, N. A. (2019). Analysis of phishing susceptibility in a workplace: A big-five personality perspectives. *Journal of Engineering Science and Technology, 14*(5), 2865–2882. Author. (2021) http://jestec.taylors.edu.my/Vol%2014%20issue%205%20October%202019/14_5_30.pdf.

Kleitman, S., Law, M. K. H., & Kay, J. (2018). It's the deceiver and the receiver: Individual differences in phishing susceptibility and false positives with item profiling. *PLoS ONE, 13*(10), Article e0205089. https://doi.org/10.1371/journal.pone.0205089

Lawson, P., Pearson, C. J., Crowson, A., & Mayhorn, C. B. (2020). Email phishing and signal detection: How persuasion principles and personality influence response patterns and accuracy. *Applied Ergonomics, 86*, Article 103084. https://doi.org/10.1016/j.apergo.2020.103084

Moody, G. D., Galletta, D. F., & Dunn, B. K. (2017). Which phish get caught? An exploratory study of individuals′ susceptibility to phishing. *European Journal of Information Systems, 26*(6), 564–584. https://doi.org/10.1057/s41303-017-0058-x

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology, 25*, 1–65.

Frauenstein, E. D., & Flowerday, S. (2020). Susceptibility to phishing on social network sites: A personality information processing model. *Computers & Security, 94*, Article 101862. https://doi.org/10.1016/j.cose.2020.101862

Pattinson, M., Jerram, C., Parsons, K., McCormac, A., & Butavicius, M. (2012). Why do some people manage phishing e-mails better than others? *Information Management & Computer Security, 20*(1), 18–28. https://doi.org/10.1108/09685221211219173

Holtfreter, K., Reisig, M. D., Piquero, N. L., & Piquero, A. R. (2010). Low self-control and fraud offending, victimization, and their overlap. *Criminal Justice and Behavior, 37*(2), 188–203. https://doi.org/10.1177/0093854809354977

Jones, H. S., Towse, J. N., Race, N., & Harrison, T. (2019). Email fraud: The search for psychological predictors of susceptibility. *PLoS ONE, 14*(1), Article e0209684. https://doi.org/10.1371/journal.pone.0209684

Curtis, S. R., Rajivan, P., Jones, D. N., & Gonzalez, C. (2018). Phishing attempts among the dark triad: Patterns of attack and vulnerability. *Computers in Human Behavior, 87*, 174–182. https://doi.org/10.1016/j.chb.2018.05.037

Wright, R., Chakraborty, S., Basoglu, A., & Marett, K. (2010). Where did they go right? Understanding the deception in phishing communications. *Group Decision and Negotiation, 19*(4), 391–416. https://doi.org/10.1007/s10726-009-9167-9

Schwartz, S. H. (2015). Basic individual values: Sources and consequences. In T. Brosch, & D. Sander (Eds.), *Handbook of value: Perspectives from economics, neuroscience, philosophy, psychology and sociology* (pp. 63–84). Oxford, UK: Oxford University Press.

Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., Ramos, A., Verkasalo, M., Lönnqvist, J.-E., Demirutku, K., Dirilen-Gumus, O., & Konty, M. (2012). Refining the theory of basic individual values". *Journal of Personality and Social Psychology, 103*, 663–688. https://psycnet.apa.org/doi/10.1037/a0029393.

Sagiv, L., & Roccas, S. (2017). What personal values are and what they are not: Taking a cross-cultural perspective. In S. Roccas, & L. Sagiv (Eds.), *Values and behavior: Taking a cross-cultural perspective* (pp. 3–13). Cham: Springer.

Allen, M. W. (2002). Human values and product symbolism: Do consumers form product preference by comparing the human values symbolized by a product to the human values that they endorse? *Journal of Applied Social Psychology, 32*(12), 2475–2501. https://doi.org/10.1111/j.1559-1816.2002.tb02752.x

Shepherd, S., Chartrand, T. L., & Fitzsimons, J. V. (2015). When brands reflect our ideal world: The values and brand preferences of consumers who support versus reject society's dominant ideology. *Journal of Consumer Research, 42*(1), 76–92. https://doi.org/10.1093/jcr/ucv005

Torelli, C. J., Özsomer, A., Carvalho, S. W., Keh, H. T., & Maehle, N. (2012). Brand concepts as representations of human values: Do cultural congruity and compatibility between values matter? *Journal of Marketing, 76*(4), 92–108. https://doi.org/10.1509/jm.10.0400

Voorn, R. J., van der Veen, G., van Rompay, T. J., & Pruyn, A. T. (2018). It takes time to tango: The relative importance of values versus traits in consumer brand relationships. *Journal of Consumer Behaviour, 17*(6), 532–541. https://doi.org/10.1002/cb.1737

Voorn, R. J., van der Veen, G., van Rompay, T. J., Hegner, S. M., & Pruyn, A. T. H. (2021). Human values as added value(s) in consumer brand congruence: A comparison with traits and functional requirements. *Journal of Brand Management, 28*, 48–59. https://doi.org/10.1057/s41262-020-00210-w

Zhang, J., & Bloemer, J. M. (2008). The impact of value congruence on consumer-service brand relationships. *Journal of Service Research, 11*(2), 161–178. https://doi.org/10.1177/1094670508322561

Kim, D., & Kim, J. H. (2013). Understanding persuasive elements in phishing e-mails. *Online Information Review, 37*(6), 835–850. https://doi.org/10.1108/OIR-03-2012-0037

Williams, E. J., Beardmore, A., & Joinson, A. N. (2017). Individual differences in susceptibility to online influence: A theoretical review. *Computers in Human Behavior, 72*, 412–421. https://doi.org/10.1016/j.chb.2017.03.002

Wright, R. T., Jensen, M. L., Thatcher, J. B., Dinger, M., & Marett, K. (2014). Influence techniques in phishing attacks: An examination of vulnerability and resistance. *Information Systems Research, 25*(2), 385–400. https://doi.org/10.1287/isre.2014.0522

Williams, E. J., Hinds, J., & Joinson, A. N. (2018). Exploring susceptibility to phishing in the workplace. *International Journal of Human-Computer Studies, 120*, 1–13. https://doi.org/10.1016/j.ijhcs.2018.06.004

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*(1), 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

Aaker, J. L., Benet-Martinez, V., & Garolera, J. (2001). Consumption symbols as carriers of culture: A study of japanese and spanish brand personality constructs. *Journal of Personality and Social Psychology, 81*(3), 492–508. https://psycnet.apa.org/doi/10.1037/0022-3514.81.3.492.

McCracken, G. (1986). Culture and consumption: A theoretical account of the structure and movement of the cultural meaning of consumer goods. *Journal of Consumer Research, 13*(1), 71–84. https://doi.org/10.1086/209048

Solomon, M. (1986). Deep-seated materialism: The case of Levi's 501 jeans. *Advances in Consumer Research, 13*, 619–622.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*, 393–404. https://doi.org/10.5465/amr.1998.926617

Alseadoon, I., Othman, M. F. I., & Chan, T. (2015). What is the influence of users' characteristics on their ability to detect phishing emails? In N. C. Pee, Y. A. Rahim, M. A. Othman, M. F. I. Othman, & H. A. Sulaiman (Eds.), *Lecture Notes in Electrical Engineering: Volume 315. Advanced computer and communication engineering technology: Proceedings of the 1st international conference on communication and computer engineering* (pp. 949–962). Springer.

Wright, R. T., & Marret, K. (2010). The influence of experiential and dispositional factors in phishing: An empirical investigation of the deceived. *Journal of Management Information Systems, 27*(1), 273–303. https://doi.org/10.2753/MIS0742-1222270111

Lönnqvist, J.-E., Verkasalo, M., Wichardt, P. C., & Walkowitz, G. (2013). Personal values and prosocial behaviour in strategic interactions: Distinguishing value-expressive from value-ambivalent behaviours. *European Journal of Social Psychology, 43*, 554–569. https://doi.org/10.1002/ejsp.1976

van Esterik-Plasmeijer, P. W., & van Raaij, W. F. (2017). Banking system trust, bank trust, and bank loyalty. *International Journal of Bank Marketing, 35*(1), 97–111. https://doi.org/10.1108/IJBM-12-2015-0195

Herman, T., Heller, E., Cohen, C., Be'ery, G., & Lebel, Y. (2014). *The Israeli democracy index 2014*. Jerusalem: The Israel Democracy Institute.

Sekerdej, M., & Roccas, S. (2016). Love versus loving criticism: Disentangling conventional and constructive patriotism. *British Journal of Social Psychology, 55*(3), 499–521. https://doi.org/10.1111/bjso.12142

Oppenheim-Weller, S., Roccas, S., & Kurman, J. (2018). Subjective value fulfillment: A new way to study personal values and their consequences. *Journal of Research in Personality, 76*, 38–49. https://doi.org/10.1016/j.jrp.2018.07.006

Gandal, N., Roccas, S., Sagiv, L., & Wrzesniewski, A. (2005). Personal value priorities of economists. *Hum. Relat., 58*(10), 1227–1252. https://doi.org/10.1177/0018726705058911

Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion, 18*(2), 129–166. https://doi.org/10.1007/BF02249397

Shao, J., Du, W., Lin, T., Li, X., Li, J., & Lei, H. (2019). Credulity rather than general trust may increase vulnerability to fraud in older adults: A moderated mediation model. *Journal of Elder Abuse & Neglect, 31*(2), 146–162. https://doi.org/10.1080/08946566.2018.1564105

Sawyer, B. D., & Hancock, P. A. (2018). Hacking the human: The prevalence paradox in cybersecurity. *Human Factors, 60*(5), 597–609. https://doi.org/10.1177/0018720818780472

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion; Psychological studies of opinion change*. New Haven, CT, USA: Yale University Press.

Parsons, K., McCormac, A., Pattinson, M., Butavicius, M., & Jerram, C. (2013). Phishing for the truth: A scenario-based experiment of users' behavioural response to emails. In *IFIP international information security conference* (pp. 366–378). Berlin, Heidelberg: Springer.

Hakim, Z. M., Ebner, N. C., Oliveira, D. S., Getz, S. J., Levin, B. E., Lin, T., Lloyd, K., Lai, V. T., Grilli, M. D., & Wilson, R. C. (2021). The Phishing Email Suspicion Test (PEST) a lab-based task for evaluating the cognitive mechanisms of phishing detection. *Behavior Research Methods, 53*, 1342–1352. https://doi.org/10.3758/s13428-020-01495-0

Anandpara, V., Dingman, A., Jakobsson, M., Liu, D., & Roinestad, H. (2007). Phishing IQ tests measure fear, not ability. In *Proceedings of the 11th international conference on financial cryptography and 1st international conference on usable security* (pp. 362–366). Scarborough, Trinidad, Tobago.