

Designing Warning Messages for Detecting Biased Online Product Recommendations: An Empirical Investigation

Bo Xiao

Information Technology Management Department, Shidler College of Business,
University of Hawaii at Manoa, Honolulu, HI, United States, boxiao@hawaii.edu

Izak Benbasat

Sauder School of Business, University of British Columbia, 2053 Main Mall,
Vancouver, BC V6T 1Z2, Canada, izak.benbasat@sauder.ubc.ca

ABSTRACT

The increasing adoption of product recommendation agents (PRAs) by e-commerce merchants makes it an important area of study for Information Systems researchers. PRAs are a type of web personalization technology that provides individual consumers with product recommendations based on their product-related needs and preferences expressed explicitly or implicitly. While extant research mainly assumes that such recommendation technologies are designed to benefit consumers and focuses on the positive impact of PRAs on consumers' decision quality and decision effort, this study represents an early effort to examine PRAs that are designed to produce their recommendations on the basis of benefiting e-commerce merchants (rather than benefitting consumers) and to investigate how the *availability* and the *design* of warning messages (a potential detection support mechanism) can enhance consumers' performance in detecting such biased PRAs. Drawing on Signal Detection Theory, the literature on warning messages, and the literature on message framing, we identified two *content design characteristics* of warning messages -- the *inclusion* of risk-handling advice and the *framing* of risk-handling advice -- and investigated how they influence consumers' detection performance. The results of an online experiment reveal that a simple warning message without accompanying advice on how to detect bias is a two-edged sword, as it increases correct detection of biased PRAs (*hits*) at the cost of increased incorrect detection (*false alarms*). In contrast, including in warning messages risk-handling advice about how to check for bias (particularly when the advice is framed to emphasize the *loss* from not following the advice) increases correct detection, but more importantly, also decreases incorrect detection. The patterns of findings are in line with the predictions of Signal Detection Theory. With an enriched understanding of how the *availability* and the *content design* of warning messages can assist consumers in the context of PRA-assisted online shopping, the results of this study serve as a basis for future theoretical development and yield valuable insights that can guide practice and the design of effective warning messages.

Keywords: Electronic Commerce, Product Recommendation Agent, Personalization, Bias, Signal Detection Theory, Manipulative Practices, Warning, Message Framing, Online Experiment

1. Introduction

With the rapid growth of electronic commerce (e-commerce), consumers' purchase decisions are increasingly made in online environments (Xiao & Benbasat, 2007, 2014). As reported by Sandler Research¹, global online retail is forecasted to reach US\$1,248.7 billion in value in 2017, representing a compound annual growth rate (CAGR) of 14.6% for the five-year period 2012-2017. Digital marketplaces offer consumers great convenience, immense assortment of products, and significant amount of product-related information. However, due to the cognitive constraints of human information processing, identifying products that meet customers' preferences is not an easy task (Xiao & Benbasat, 2007). As a result, many e-commerce websites (e.g., myproductadvisor.com, easy-wine.net, openrice.com, amazon.com, jinni.com, appmatcher.com, tripbase.com, olayforyou.com) have implemented product recommendation agents (PRAs) to assist consumers in product search, assessment, and selection.

PRAs are a type of web personalization technology that provides individual consumers with product recommendations based on their product-related needs and preferences expressed explicitly or implicitly (Xiao & Benbasat, 2007). PRAs have the potential to improve the quality of decisions made by consumers when searching for and selecting products online, while at the same time reducing their information overload and search complexity (Xiao & Benbasat, 2007). Research by Limelight Networks has identified personalized recommendations as an e-commerce feature that is highly valued by consumers, a feature that can help online companies create brand loyalty (Limelight, 2010), which explains the phenomenal increase in the adoption of PRAs by online merchants in different industries (e.g., electronics, automobiles, movies, books, cosmetics, food and wines, tourism, real estate) (Chau, Ho, Ho, & Yao, 2013). eBay's decision to acquire recommendation technology Hunch², Amazon's acquisition of Goodreads (one of the popular reading recommendation engines)³, and more recently Apple's

¹ <http://www.sandlerresearch.org/industry-views/opportunities-for-online-shopping-in-different-countries>

² <https://gigaom.com/2011/11/21/why-ebay-is-buying-recommendation-engine-hunch/>

³ <http://www.engadget.com/2013/03/28/amazon-acquires-goodreads-kindle-recommendation-engine/>

acquisition of BookLamp and its book recommendation engine⁴ further signal the importance of recommendation technologies to e-commerce leaders. In fact, among the Internet Retailer top 500 e-retailers for 2013, 76.4% feature product recommendations based on individual shopper's needs and preferences⁵.

Despite the potential for PRAs to aid consumers in decision making, the degree to which PRAs truly empower consumers depends upon their veracity and objectivity (Hill, King, & Cohen, 1996; King & Hill, 1994). The ultimate objective of an online merchant adopting PRA is to “generate more business opportunities” by providing consumers with personalized product recommendations (Chau et al., 2013, p. 180). Therefore, rather than solely maximizing the benefits of consumers in a shopping process, the merchant aims to strike a balance between consumers' benefits and business criteria (Chau et al., 2013). On the one hand, the merchant wishes to provide consumers with product recommendations personalized to their needs and preferences so as to attract and retain customers. On the other hand, the merchant often has a vested interest in recommending products from certain vendors and/or products with certain characteristics (e.g., high profit margin products, soon-to-be discontinued products), so as to attain higher-than-usual profits (in terms of mark-ups, commissions, referral fees, etc.) or reduce losses. With such mixed motivations in mind, the merchant may implement PRAs to provide recommendations that are not solely preference-matched to benefit consumers but instead *biased* toward its own interests (Chau et al., 2013; Xiao & Benbasat, 2011, 2015).

The cost of relying on a PRA designed to produce biased recommendations (referred to as biased PRA hereafter) may very well negate its fundamental value (i.e., improving decision quality while reducing decision effort), as the gain in decision making efficiency may be overshadowed by the loss in decision quality (Haubl & Murray, 2006). Consequently, biased PRAs are an important concern for e-commerce. Nonetheless, there has been scant scholarly attention directed to this phenomenon; even less studied are potential support mechanisms that can be implemented to curb the effects of biased

⁴ <http://venturebeat.com/2014/07/26/apple-has-acquired-booklamp-and-its-book-recommendation-engine/>

⁵ <http://www.internetretailer.com/2013/05/15/site-personalization-has-room-grow-among-top-500-retailers>

recommendations in PRA-assisted online shopping. The study reported in this paper aims to address this knowledge gap by focusing on one support mechanism, *warning*, and exploring how its *availability* and the *design* of its informational content can influence consumers' performance in detecting biased e-commerce PRAs.

Consumers are found to be generally vulnerable to manipulative practices in e-commerce (e.g., Grazioli, 2004; Grazioli & Jarvenpaa, 2000, 2001, 2003a, 2003b; Grazioli & Wang, 2001). Warning individuals explicitly about potential manipulative practices can arouse their suspicion and alert them to cues of manipulation, thus increasing their likelihood to detect such manipulations. However, there is the possibility that warning messages can also lead to individuals perceiving manipulation when it does not exist. Prior research has focused on the *presence or absence* of warning messages on detection performance, without investigating design variables that affect the effectiveness of the warning messages, an additional gap this study aims to fill.

Message content (rather than source credibility or attractiveness) is the most significant predictor of individuals' attitude and behavior (Fishbein & Ajzen, 1981). This study focuses on the *content* of warning messages and examines the effects of two content design characteristics, namely, the *inclusion* of risk-handling advice and the *framing* of such advice, on consumers' *performance* in detecting *biased* recommendations from online PRAs. More specifically, this study aims to answer two key questions:

- (1) Will the inclusion of risk-handling advice in a warning message enhance consumers' performance in detecting biased recommendations from a PRA?
- (2) Will the framing (positive or negative) of the risk-handling advice (included in the warning message) influence consumers' performance in detecting biased recommendations from a PRA?

An online experiment was conducted to address these two research questions. The results of the study show that *warning* consumers explicitly about potential PRA bias enhances their performance in detecting bias in the PRA's recommendations. However, warnings may also lead to false alarms. The results of this study provide strong evidence that the most effective mechanism in supporting consumers

in bias detection is to provide them with warning messages that include advice (or tips) for addressing the risk of PRA bias and frame the advice to emphasize the loss of not following such advice.

To our knowledge, this is the first empirical study that examines and explicitly compares the effectiveness of different warning mechanisms in supporting consumers in detecting biased product recommendations. Not only does this study contribute to the IS literature by examining potential factors related to *warning* that influence consumers' detection performance, but it also fills a void in existing warning research by shifting the focus from the presence/absence of warning messages to their design characteristics. Further, by enhancing the understanding of the broad phenomenon of manipulative business practices in e-commerce, this study contributes to the concerted effort by government agencies, consumer protection organizations, and industry associations to combat such practices.

The remainder of this paper is organized as follows. Section 2 reviews previous literature and discusses the theoretical foundation for the design of warning messages. Section 3 presents the research model and develops hypotheses. The research method and results of hypothesis testing are reported in section 4 and 5, and the paper concludes with a discussion of the results, contributions of the study, and limitations and suggestions for future research.

2. LITERATURE REVIEW AND THEORETICAL FOUNDATIONS

The section reviews prior research on biased PRAs and introduces the theory and literature guiding the design of warning messages (i.e., Signal Detection Theory, the literature on warning messages, and the literature on message framing).

2.1 Prior Research on Biased PRAs

Extant research on PRAs generally assumes that such recommendation technologies are designed to provide recommendations to benefit consumers and focuses on the positive impact of PRAs on consumers' decision quality, decision effort, as well as their evaluation and acceptance of the PRAs (see a comprehensive review by Xiao & Benbasat, 2007). However, not all PRAs are altruistic; rather, many are designed not only to assist consumers but also to steer them in a particular direction, which makes them

“double agents” (Haubl & Murray, 2006). Hence, e-commerce PRAs have the potential to both aid and influence consumers in their decision making (Cosley, Lam, Albert, Konstan, & Riedl, 2003; Haubl & Murray, 2003; Senecal, 2003; Senecal & Nantel, 2004).

There has only been limited research into the dynamics and effects of PRA bias. Aksoy and Bloom (2001) introduced bias into a PRA by using product attribute importance weights that deviated from the weights expressed by PRA users to generate recommendations. They found that users of the biased PRA had higher search effort, reduced decision quality, and degraded perception of the usefulness of the recommendations. Cosley et al. (2003) altered a movie PRA’s predicted rating (i.e., the PRA’s prediction of a user’s liking of a movie, based on the user’s profile) to be either higher, lower, or the same as the actual predicted rating. They found that such alteration significantly lowered users’ satisfaction with the PRAs’ recommendations. Haubl and Murray (2003) found that product attributes selectively included in a PRA’s preference elicitation interface had a significant impact on consumers’ revealed preferences (in terms of the relative weight they attached to different product attributes), and subsequently their purchase decisions. Focusing on PRAs for digital cameras, Xiao and Benbasat (2015) showed that consumers’ product choice at e-commerce website was significantly influenced by biased recommendations provided by the PRAs. However, while the biased recommendations led to impairment in consumers’ actual (or objective) decision quality, they did not reduce consumers’ perceived (or subjective) decision quality, hence highlighting the serious risk consumers would face when dealing with biased PRA. Chau et al. (2013) found that biased music recommendations led to high level of distrust in a PRA’s competence and integrity, which in turn influenced online users’ interactions with the PRA. To our knowledge, none of the previous studies have explored interventions that can support consumers in detecting bias in the PRAs’ recommendations.

2.2 Theoretical Foundations

In this study, we draw on Signal Detection Theory, the literature on warning messages, and the literature on message framing to identify the design characteristics of warning messages. We also

subscribe to Signal Detection Theory to differentiate the possible outcomes of the task of detecting biased PRAs and to explain the pathways by which warning messages, and their design characteristics, influence consumers' performance in detecting biased PRAs.

2.2.1 Signal Detection Theory

Signal Detection Theory (Davis & Parasuraman, 1981; Green & Swets, 1966) provides a general framework to describe and study decisions made in uncertain or ambiguous situations (Wickens, 2001). It differentiates between two classes of events — *noise* (i.e., the background) and *signal* (i.e., stimulus that deviates from background noise and thus may be detected) — and explains the performance of individuals who strive to determine the presence of a signal. The theory also recognizes two types of outcomes when individuals strive to differentiate a signal from background noise: *Hits* (when signals are successfully detected) and *false alarms* (when background noises are incorrectly identified as signals). Whereas *hits* are indicators of *detection success*, *false alarms* indicate *detection failure*. Superior detection performance is achieved when the rate of *hits* is high and the rate of *false alarms* is low. In the context of detecting bias in online PRAs' recommendations, two types of outcomes are of interest: A *hit* occurs when a consumer perceives bias in the recommendations of a biased PRA, and a *false alarm* occurs when a consumer perceives bias in the recommendations of an unbiased PRA.

Signal Detection Theory also accounts for two factors influencing the outcomes of error detection: Discriminant ability and decision threshold⁶ (Green & Swets, 1966). *Discriminant ability* refers to the ease with which the person making the judgment distinguishes the signal from the noise (i.e., in the context of this study, how easy it is for an individual to determine whether a recommendation is biased or not) and is affected by both personal and situational characteristics (Scott, 2006). Discriminant ability is high when the individual making the judgment is experienced or well trained in detection tasks. For instance, an experienced interrogator in a criminal investigation is more likely to detect lies in a suspect's confession. Discriminant ability is also high when the individual making the judgment is equipped with

⁶ *Discriminant ability* and *decision threshold* are also referred to as *sensitivity* and *bias* (or *response/decision bias*), respectively, in prior literature (e.g., Scott, 2006).

the right tools or information. For instance, a polygraph machine or information about the criminal history of the suspect can help the interrogator detect lies. Finally, discriminant ability is high when the signal is strong. For instance, some suspects are just bad liars, and thus an interrogator can easily tell that they are lying.

In detecting signals, many times there are ambiguous cases. *Decision threshold* refers to the level at which the individual making the judgment sets the cutoff point for ambiguous cases (Scott, 2006). It is the decision point at which the individual determines that a signal is present. Decision threshold is also affected by individual differences and situational characteristics, such as general propensity to (dis)trust and beliefs about the base rate of a signal (Scott, 2006). For instance, some individuals are more distrustful than others and are thus more likely to assume that any given statement is likely a lie. Some situations also lead us to expect lies. For instance, a statement made by a car salesman in a sales negotiation with customers would be more likely to be taken as not truthful when compared to the same statement made by the same car salesman to a friend in a casual chat (Scott, 2006).

Signal Detection Theory suggests that support mechanisms aimed to enhance consumers' performance in detecting PRA bias can be designed to (1) improve consumers' *discriminant ability* (for instance, by advising consumers on how to access the necessary knowledge to detect anomalies resulting from bias) and/or (2) lower consumers' *decision threshold* for judging ambiguous cues as anomalies (for instance, by enhancing consumers' sensitivity to potential bias via explicit warnings).

2.2.2 Warning Messages, Decision Threshold, and Discriminant Ability

One of the most studied detection support mechanisms in prior research for inducing sensitivity to potential bias is the provision of *warnings*. Existing evidence indicates that *warning* individuals explicitly about potential manipulative practices can arouse consumers' suspicion and alert them to anomalies resulting from manipulations (G. R. Miller & Stiff, 1993; Parasuraman, 1984; Stiff, Kim, & Ramesh, 1992). However, the positive effect of warning messages on detection success has received mixed empirical support. Whereas Biros et al. (Biros, 1998; Biros, George, & Zmud, 2002), Grazioli (2004), and

George et al. (2004) showed that suspicious receivers had higher detection performance when compared to unsuspecting ones, others (e.g. George & Maret, 2004; Grazioli & Wang, 2001; Maret & George, 2005; Tilley, 2005) did not show the positive influence of induced suspicion on detection performance. Still others (e.g., Burgoon, Buller, Ebesu, & Rockwell, 1994; Parasuraman, 1984) revealed that warning messages enhanced consumers' performance in detecting anomalies at the cost of increased false alarms. These equivocal findings suggest that the key question may not be whether warning messages improve detection success but rather how they should be designed (e.g., in terms of informational content, format, vividness, format, warning time and location) (Argo and Main 2004; Stewart and Martin, 1994) to enhance their effectiveness. Nevertheless, prior detection studies have mostly focused on the presence or absence of warning messages on detection performance, without exploring *design variables* that may affect the effectiveness of the warning messages, a novel question this study aims to answer.

In the public policy and marketing literature, there has been extensive research on the effectiveness of product warning messages (e.g., warning labels on tobacco packages) (see Argo & Main, 2004). Consumer product warnings serve two major purposes: to prevent consumers from engaging in behaviors that may lead to undesirable consequences, and to promote appropriate product usage behavior (Argo and Main 2004). An effective consumer product warning is one that not only communicates clearly the nature of the risk facing consumers, but also provides advice on how to avoid that risk (Kelley, Gaidis, & Reingen, 1989; Ross, 1981). The provision of such advice on what consumers could do to avoid the negative consequences described in the warning message will equip consumers with necessary knowledge and skills to manage the risky situation, thus enhancing their motivation to change their behavior in accordance with the advice received. From a meta-analysis and literature review of prior research on warning messages involving fear appeals, Witte and Allen (2000) conclude that a warning message is most effective when it also includes information about how to avert or minimize the risk highlighted in the message.

Literature on warning messages (particularly research on consumer product warning) suggests that, in the context of our study, the effectiveness of warning messages can be enhanced by improving the

design of their informational content. More specifically, in addition to warning consumers about the potential risk of PRA bias (and thus lowering their *decision threshold* in judging ambiguous cases), the message should also include advice on how to handle this risk (so as to increase consumers' *discriminant ability* in detecting bias). Nonetheless, in order to persuade consumers to adopt the risk-handling advice included in the warning message, the advice (which can be considered a persuasive message by itself) needs to be communicated in a way to maximize its impact on consumers' behavioral outcome (Rothman, Bartels, Wlaschin, & Salovey, 2006); this is where insights from prior research on message framing are particularly valuable.

2.2.3 Message Framing as Persuasive Communication Strategy

Message framing is a persuasive communication strategy aimed at motivating behavior through the presentation of equivalent appeals or opinions framed in terms of either *gains* or *losses* (Gerend & Sias, 2009). Levin, Schneider, and Gaeth (1998) distinguish among three different kinds of framing – risky choice framing, attribute framing, and goal framing. In *risky choice framing*, introduced by Tversky and Kahneman (1981) and traditionally associated with the term “framing”, the outcomes of a potential choice involving options differing in level of risk are described in different ways (e.g., “1/3 chance 600 people will be saved” vs. “2/3 chance 600 people will die”). In *attribute framing*, a single attribute within any given context is the subject of the framing manipulation (e.g., “75% lean” vs. “25% fat”). Finally, *goal framing* is designed to influence the implicit goals that an individual adopts by specifying the consequences of a particular behavior in either positive or negative terms. Of the three types of framing, goal framing is often applied in persuasive communications context and thus is relevant to our study.

A *goal framing* effect occurs when a persuasive message has different appeal depending on whether it stresses the positive consequences of performing an act or the negative consequences of not performing the act. A review of prior research (see Levin, Gaeth, & Schreiber, 2002; Levin et al., 1998) reveals *equivocal* empirical evidence for the relative impact of positively-framed messages (emphasizing gains) vs. negatively-framed ones (emphasizing losses) in influencing a given behavior. For instance,

Meyerowitz and Chaiken (1987) found that women were more apt to engage in breast self-examination (BSE) when presented with information stressing the negative consequences of not performing the BSE than when presented with information stressing the positive consequences of performing the BSE. Likewise, Banks et al. (1995) showed that women receiving loss-framed message (emphasizing risks of not being screened) were more likely to obtain a mammogram in the next 12 months than those who received gain-framed message (emphasizing benefits of being screened). On the other hand, Detweiler and colleagues (Detweiler, Bedell, Salovey, Pronin, & Rothman, 1999) found that beachgoers who had received a gain-framed brochure about skin cancer were significantly more likely to seek out free sample of sunscreen. Rothman et al. (Rothman, Salovey, Antone, Keough, & Martin, 1993) also showed that gain-framed message had greater impact on the use of sunscreen for females. Such mixed findings have led researchers to explore moderating variables that can enhance, eliminate, or even reverse the goal framing effect.

One potential moderating variable has to do with the nature of the behavior advocated in the persuasive message (Rothman et al., 2006; Rothman & Salovey, 1997). Prospect theory postulates that while people are *risk taking* when a decision problem is formulated in terms of loss, they are *risk averse* when they are faced with gain-framed decision problem (Tversky & Kahneman, 1981). Consistent with this perspective, the impact of a given frame (positive or negative) on behavior “should depend on whether the behavior under consideration is perceived to reflect a *risk averse* or *risk seeking* course of action” (Rothman et al., 2006, p. S205), which in turn depends on “the extent to which people perceive the behavior will afford an unpleasant outcome” (Rothman et al., 2006, p. S205). More specifically, when people are considering a *detection* behavior (e.g., cancer screening or annual health checkup) that is likely to involve some unpleasant outcome (e.g., detecting a health problem) and is thus considered *risky*, negatively-framed appeals or opinions should be more persuasive. In contrast, positively-framed appeals or opinions are more effective when people are considering a *prevention* behavior (e.g., applying sunscreen or using mouth rinse) that is less likely to involve unpleasant outcomes (Rothman et al., 2006; Rothman & Salovey, 1997) and is thus considered a *risk averse* course of action. In other words, the

performance of *detection (prevention)* behavior is best facilitated by a *negatively*-framed (*positively*-framed) message (Detweiler et al., 1999). Existing empirical evidence is largely supportive of the *detection-prevention* classification (See Rothman et al., 2006 for a review).

Literature on message framing suggests that, in the context of our study, the risk-handling advice included in the content of the warning message can be framed either positively (emphasizing the *gain* or *positive* implication associated with following the advice) or negatively (emphasizing the *loss* or *negative* implication associated with not following the advice). Such framing falls into the category of *goal-framing*, as it aims to influence the implicit goals that consumers adopt. The relative effectiveness of the two *framing* strategies in enhancing consumers' detection performance will be explored in this study.

3. RESEARCH MODEL AND HYPOTHESES DEVELOPMENT

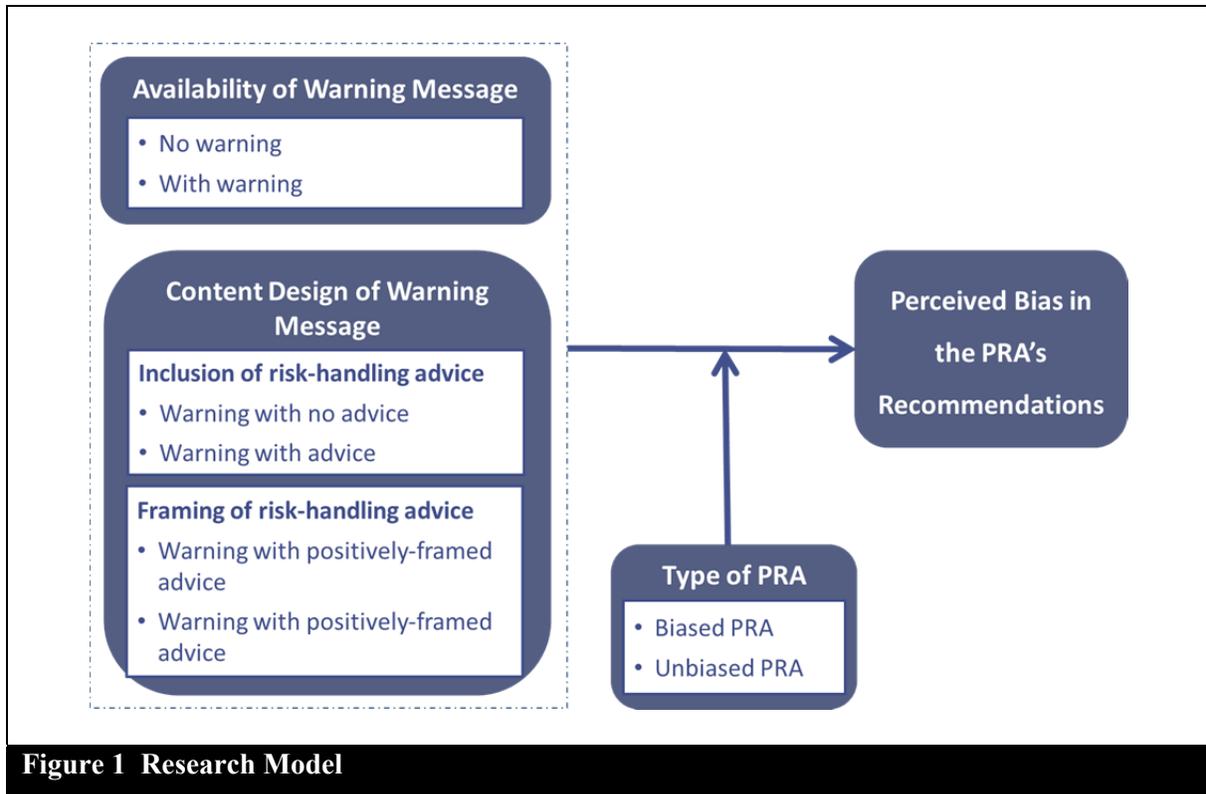
Just like other manipulative online practices, biased recommendations from PRAs may present serious risks (financial and/or social) to consumers shopping on the Internet (Grazioli, 2004; Grazioli & Jarvenpaa, 2000, 2001, 2003a, 2003b; Grazioli & Wang, 2001; Xiao & Benbasat, 2011). This study focuses on one support mechanism, *warning*, that has been shown in prior research to help consumers better detect online manipulations (e.g., Biros et al., 2002; Grazioli, 2004). While extant detection research is characterized by a predominant interest in the presence or absence of warning messages on detection performance, this study goes one important step further in identifying design characteristics of warning messages and theorizing about their effects on consumers' performance in detecting PRA bias,.

Drawing from Signal Detection Theory, the literature on warning messages, and the literature on message framing, we posit that the *availability* and *content design* of warning messages affect consumers' performance in detecting PRA bias via their influence on consumers' *decision threshold* and *discriminant ability*, respectively. More specifically, we propose that:

- The mere *availability* of warning messages may not enhance consumers' detection performance as they increase *both* hits and false alarms via lowering consumers' *decision threshold*;

- Warning messages designed with proper *content* will improve consumers' detection performance by increasing hits without simultaneously increasing false alarms via improving consumers' *discriminant ability*.

To empirically test these propositions, we developed the following Research Model (Figure 1) and a set of testable hypotheses described next.



Availability of warning message, as the name suggests, refers to whether a warning message⁷ (an intervention aimed to lower *decision threshold*) is conveyed to consumers about the possibility of PRA bias.

Content design of warning message refers to the two content design characteristics of warning messages, namely, the *inclusion* of risk-handling advice in warning messages and the *framing* of the risk-

⁷ The warning message can be of any one of following three types: Warning with no advice, warning with positively-framed advice, or warning with negatively-framed advice.

handling advice, which give rise to three different types of warning messages⁸: i) Warning with no advice, ii) warning with positively-framed advice, and iii) warning with negatively-framed advice. Consumers who receive a *warning with no advice* are warned of the risk of PRA bias but are *not* provided with advice for handling such risk (i.e., strategies for detecting bias in the PRA's recommendations). Consumers who receive the other two types of warning are provided with a warning message that includes risk-handling advice (an intervention aimed to enhance *discriminant ability*) framed either *positively* (thus accentuating the gain from following the advice) or *negatively* (thus stressing the loss from not following the advice).

Type of PRA refers to whether the PRA at the e-commerce website is an unbiased one or a biased one. While an **unbiased** PRA bases its recommendations solely on consumers' expressed preferences, a **biased** PRA recommends products on the basis of additional business criteria (e.g., profit margin and/or commission from selling particular products) (Chau et al., 2013) and thus may not recommend products that best match the consumers' preferences.

Consumers' detection performance is reflected in their performance in detecting bias in the PRA's recommendations. Prior research in deception detection suggests that identifying anomalies (i.e. deviations from what is expected, usual, and normal) resulting from strategic manipulation is the first and critical step toward successful detection (E. J. Johnson, Lohse, & Mandel, 2001; P. E. Johnson, Grazioli, & Jamal, 1993; Morrison & Robinson, 1997; Robinson, 1996). Thus, in this study, bias is assessed by the level of *perceived anomaly* in the PRA's recommendations. **Perceived bias in the PRA's recommendations** refers to the extent to which consumers believe that the PRA's recommendations deviate from what is expected, usual, and normal.

3.1 Availability of Warning Message

Prior research has identified the provision of *warnings* as an intervention that can sensitize individuals to manipulation-related anomalies (G. R. Miller & Stiff, 1993; Parasuraman, 1984; Stiff et al., 1992). Likewise, in the context of our study, explicit *warnings* about potential bias effected by PRA are

⁸ In the research model, "warning with advice" refers to warning with either positively-framed advice or negatively-framed advice.

expected to lower consumers' *decision threshold* in judging ambiguous cues as anomaly (e.g., by mitigating individuals' general propensity to trust others and/or increasing the expected base rates of anomalies) (Scott, 2006), thus enhancing the likelihood for consumers to perceive bias in the recommendations of a *biased* PRA and resulting in higher rates of *hits*. However, Signal Detection Theory suggests that lowered decision threshold in an individual's judgment may also lead to more situations to be inaccurately perceived as anomalous, resulting in higher rates of *false alarms* (which occur when consumers perceive bias in the recommendations of an *unbiased* PRA). Findings from several studies (e.g., Burgoon et al., 1994; Parasuraman, 1984) have confirmed that warning messages enhanced consumers' performance in detecting bias at the cost of increased false alarms. We thus hypothesize:

H1: Compared to those receiving *no warning*, consumers who are provided with a *warning* will be more likely to perceive bias in the recommendations of both a *biased* PRA (H1a) and an *unbiased* PRA (H1b), thus resulting in more *hits* and more *false alarms*⁹.

3.2 Content Design of Warning Message – Inclusion of Risk-Handling Advice

According to Signal Detection Theory, interventions aimed to help individuals improve detection performance can be designed to increase individuals' *discriminant ability* (e.g., by providing them with necessary knowledge to discriminate signal from noise) (Scott, 2006). In comparison to those affecting one's *decision threshold*, interventions aimed to influence *discriminant ability* tend to be more effective in achieving high rates of hits without simultaneously increasing rates of false alarms (Scott, 2006).

Warning with risk-handling advice is an intervention that can potentially enhance consumers' *discriminant ability*, and does so by providing consumers facing an uncertain judgment decision with more information about the situation at hand and encouraging them to make independent verification using appropriate tools. Consumers who follow the advice¹⁰ (either *positively-framed* or *negatively-*

⁹ Please note that our hypotheses make predictions that are ordinal in nature: we are not claiming that certain experimental (i.e., warning) conditions will result in zero false alarms; rather, we make comparative predictions.

¹⁰ An example advice (used for this study) is to verify the PRA's recommendations by comparing the recommended products with other, non-recommended products in the same brand using functionalities such as searching-by-brand (i.e., a brand-based browsing option that allows consumers to view all the products in a specific brand) and/or a

framed) included in the warning message by engaging in the behavior advocated in the advice will perform better in identifying bias in a PRA.

In contrast, a warning message with no risk-handling advice does not by itself improve consumers' discriminant ability to compensate for the impact of lowered decision threshold. Simply *warning* consumers about the risk of PRA manipulation (without simultaneously providing them with strategies to handle such risk) will place consumers in a heightened state of alertness and increase their likelihood of perceiving bias in the recommendations of an *unbiased* PRA (resulting in increased *false alarms*) as well as in the recommendations of a *biased* PRA (resulting in increased *hits*).

In sum, we expect that consumers who receive a *warning message* that includes *risk-handling advice* will have more *hits* and fewer *false alarms* than those who receive a simple *warning with no advice*. Therefore,

H2: Compared to those receiving a *warning with no advice*, consumers provided with a *warning* message that includes *advice* for handling risk will be more likely to *perceive bias* in a biased PRA (H2a) and less likely to *perceive bias* in an unbiased PRA (H2b), resulting in more *hits* and fewer *false alarms*.

3.3 Content Design of Warning Message – Framing of Risk-Handling Advice

The risk-handling advice included in a warning message is effective to the extent that consumers who receive the warning message follow the advice and adopt the course of action advocated in the advice. Therefore, the risk-handling advice (a persuasive message by its nature) needs to be communicated in a manner that maximizes its impact on consumers' behavior.

Prior research in message framing suggests that persuasive messages can be “framed” to either emphasize the benefits of taking action (i.e., a gain-framed appeal) or accentuate the costs of failing to take action (i.e., a loss-framed appeal) (Rothman et al., 2006). *Loss*-framed appeals are more persuasive when the behavior under consideration is a *detection* (as opposed to *prevention*) behavior associated with

comparison matrix (i.e., a graphic organizer that facilitates the in-depth comparison of two or more products side-by-side). Both functionalities allow consumers to better assess whether expectation violations have occurred.

high likelihood of an unpleasant outcome (Rothman et al., 2006; Rothman & Salovey, 1997), as potential losses (emphasized in loss-framed appeals) tend to be more motivating than potential gains (underscored in gain-framed appeals) when risky actions (exemplified by detection behavior that involves unpleasant outcomes) are contemplated (O'Keefe & Jensen, 2009). Accordingly, in the context of this study, as the task of bias detection necessarily involves the possibility of an unpleasant outcome (i.e., uncovering manipulation by the PRA), an advice in the warning message encouraging consumers to adopt the advocated risk-handling behavior (e.g., to use functionalities such as searching-by-brand and comparison matrix to verify the PRA's recommendations) is likely to be more effective in motivating consumers to take extra search action when the advice is framed as *loss* than when it is framed as *gain*. Thus, we hypothesize:

H3: Compared to those receiving a *positively-framed* warning message, consumers provided with a *negatively-framed* warning message will be more likely to *perceive bias* in a biased PRA (H3a) and less likely to *perceive bias* in an unbiased PRA (H3b), resulting in more *hits* and fewer *false alarms*.

4. RESEARCH METHOD

An online experiment was conducted to test the hypothesized relationships.

4.1 Experimental Design

A **4** (*Type of Warning Message*: No warning, warning with no advice, warning with positively-framed advice, or warning with negatively-framed advice) x **2** (*Type of PRA*: Unbiased or biased) between-subject factorial design was employed.

Design of Warning, Advice, and Framing. For participants assigned to groups with warnings, a pop-up window with warning message is displayed immediately before the start of the experimental task at the e-commerce website. E-commerce merchants (particularly those motivated to provide biased recommendations) are understandably reluctant to warn consumers of potential manipulative practices. As such, warning messages were implemented in this study as issued by a third party (e.g., a consumer

organization) independent of the e-commerce merchant. In order to convey to experimental participants that the warning message was **not** issued by (or part of) the e-commerce website that they were visiting, the pop-up window containing the message was designed with background color and font type/size/color distinct from those of the e-commerce website¹¹. Participants were asked to read the message before they accessed the e-commerce website.

As explained earlier, in this study, the risk-handling advice included in a warning message was to encourage consumers to conduct independent verification of the PRA's recommendations by comparing the recommended products with other, non-recommended products in the same brand using functionalities such as searching-by-brand and/or a comparison matrix that were available at the e-commerce website.

There are different ways to construct the framing of risk-handling advice. Rothman and Salovey (1997) note that, while positively-framed message can focus on attaining a desirable outcome or not attaining an undesirable outcome (both gains), negatively-framed messages can emphasize the attainment of an undesirable outcome or the failure to attain a desirable outcome (both losses). Accordingly, in this study, we manipulated positively-framed advice by accentuating both gains of performing the risk-handling behavior (i.e., *increased chance* of distinguishing an unbiased PRA from a manipulative one and *reduced risk* of being misled by biased recommendations) and constructed negatively-framed advice by stressing both losses of failing to perform the target behavior (i.e., *reduced chance* of distinguishing an unbiased PRA from a manipulative one and *increased risk* of being misled by biased recommendations). Because the warning message was rather long when the risk-handling advice was included, relevant texts in the advice were bolded and italicized to enhance the perceptual salience of the behavior being promoted and the positive/negative consequences. Table 1 illustrates the operationalization of different types of warning messages and the risk-handling advice (framed positively or negatively).

¹¹ In real world, the pop-up window with warning message could be implemented as a browser extension provided by a third-party. When downloaded and installed by consumers, the extension containing the warning message would load when consumers visit e-commerce stores.

Table 1 Design of Warning Messages and Risk-Handling Advice

Type of Warning Message	Components	Text of Warning Message
Warning with No Advice	<i>Simple Warning</i>	When shopping at online stores that provide automated Shopping Advisors ¹² , please be aware that some advisors may provide product recommendations biased toward certain brands or toward products with certain characteristics.
Warning with Positively-Framed Advice	<i>Simple Warning</i> <i>Positively-framed risk-handling advice</i>	<p>... (Same text as in Warning with No Advice condition here)</p> <p>Consumers are advised to verify the Shopping Advisor's recommendations by comparing recommended products with other, non-recommended products in each of your preferred brands, using functionalities such as comparison matrix and searching by brand that are available at the website.</p> <p>Research shows that consumers who <i>verify the Shopping Advisor's product recommendations</i> have <i>increased chance</i> of distinguishing an unbiased Shopping Advisor from a manipulative one and <i>reduced risk</i> of being misled by biased recommendations.</p>
Warning with Negatively-Framed Advice	<i>Simple Warning</i> <i>Negatively-framed risk-handling advice</i>	<p>... (Same text as in Warning with No Advice condition here)</p> <p>Consumers are advised to verify the Shopping Advisor's recommendations by comparing recommended products with other, non-recommended products in each of your preferred brands, using functionalities such as comparison matrix and searching by brand that are available at the website.</p> <p>Research shows that consumers who <i>fail to verify the Shopping Advisor's product recommendations</i> have <i>reduced chance</i> of distinguishing an unbiased Shopping Advisor from a manipulative one and <i>increased risk</i> of being misled by biased recommendations.</p>

Design of Websites, PRAs, and Bias. Prior experimental studies examining biased recommendations have either introduced the bias randomly (e.g., by increasing/decreasing the predicted product ratings or consumer indicated product attribute importance weight) (e.g., Aksoy & Bloom, 2001;

¹² In the experiment, a PRA is referred to as an automated shopping advisor.

Cosley et al., 2003) or presented only products from one particular product vendor (which is essentially an outright lie and is sure to induce suspicion from consumers) (e.g., Chau et al., 2013; Komiak & Benbasat, 2008). While PRA bias can be effected in various ways, this study focuses on PRAs that recommend products from among a *reduced set of products* the online merchant intends to promote (rather than from among all the products available at the e-commerce website), so as to enhance the likelihood for consumers to choose the promoted products. This is akin to manipulations in personal selling situations, for example, in which salespersons are motivated to recommend from a subset of products whose commission rates are high from product vendors or products they are pressured (by managers for instance) to sell (e.g., products whose stock levels are high in warehouse or having a high profit margin) (Chau et al., 2013), rather than recommending products that best match customers' preferences.

Two experimental websites (one featuring a biased PRA and the other providing an unbiased PRA) were developed for this study. Each website featured the same 96 digital cameras from 8 brands¹³, with 12 products in each brand. The product features for the 12 digital cameras in each brand were carefully designed such that six products (referred to as the *promoted* products) were dominated by the other six products (referred to as the *dominant* products). Each *promoted* product was paired with a dominant *product* in the same brand. In each pair, the *dominant* product performed better on two important attributes than the *promoted* product but offered the same price. Two content-filtering PRAs for digital cameras were adapted from Wang and Benbasat (2005). Each website had 48 *dominant* products and 48 *promoted* products. Table 2 illustrates how the two PRAs were designed.

¹³ While real world ecommerce websites (e.g., Amazon, Dell, and the online stores of BestBuy, Walmart, etc.) typically offer a much larger selection of alternatives from greater number of brands, prior experimental studies involving PRAs generally focus on a much smaller selection of alternatives to control the cognitive load placed upon participants. For instance, Haubl and Trifts (2000) constructed a total of 54 products (nine alternatives for each of the six brands) for their study. In the experiment carried out by Punj & Moore (2009), the number of alternatives was set at 30 in the “few alternatives” condition and at 99 for the “many alternatives” condition. Xu, Benbasat, & Cenfetelli (2014) kept the number of product alternatives in experimental website constant at 32 across all treatment conditions. Consistent with prior practice, the experimental website designed for our study featured 96 digital cameras from eight brands.

Table 2 The Design of PRAs

Both Biased and Unbiased PRAs

- The PRAs use needs-based questions(adapted from Wang & Benbasat, 2005) to elicit users' product related preferences
- The PRAs calculate a fit score for every available product based on users' expressed preferences
- The PRAs generate a list of 12 products, with six products displayed in each page¹⁴
- Users can use the built-in comparison matrix to compare products
- Users can utilize the built-in search-products-by-brand functionality to view additional products that are not recommended by the PRA

Unbiased PRA

- The PRA selects 12 products from among **all 96 available products** that have the highest fit-scores and present them in the recommendation list (ordered by fit scores)

Biased PRA

- The PRA selects 12 products from among **the 48 promoted products** that have the highest fit-scores and present them in the recommendation list (ordered by fit scores)
- None of the 48 dominant products will be included in the two-page recommendation list

A credible lie is one that will not trigger the target's suspicion. The bias that we implement in this study is a credible lie for two reasons. First, our manipulation of bias is detectable if consumers exert effort to view and compare extra products in their preferred brands. However, it is not as easily detected as the practice of recommending products from a single brand¹⁵ (even when consumers indicate a preference for other brands). Second, to make a credible lie, a liar needs to get inside the target's head (i.e., to take into account the perspective of the target)¹⁶; the PRA implemented in our study does exactly this by taking into account the expressed preferences of consumers when producing the biased recommendations.

¹⁴ Human beings are limited in the amount of information that can be processed simultaneously. According to Miller (1956), the magical number is seven, plus or minus two. In other words, the number of items that can be held in individuals' short-term memory is limited to seven, plus or minus two. Therefore, in our study, displaying six recommended products on each page is appropriate for controlling the participants' cognitive load.

¹⁵ In the experimental study conducted by Komiak and Benbasat (2008) as well as that by Chau et al. (2013), only products from one particular vendor were recommended.

¹⁶ <http://www.psychologytoday.com/blog/extreme-fear/201005/top-ten-secrets-effective-liars>

4.2 Sample

Participants for this study were 496 e-commerce shoppers recruited from a North American panel maintained by a marketing research firm, which specializes in market research sampling and custom panel recruitment. The sample size ensured a 0.8 power for the required hypothesis tests to detect a medium effect of 0.15. An invitation to participate in the study was broadcast by the marketing firm via e-mail to members of the panel. Individuals were provided with point-based incentive (redeemable for various prizes) for their assistance in the study available through the marketing firm.

4.3 Experimental Task and Procedures

All participants were randomly assigned to one of the eight experimental groups. They were told that an online camera store, ForeverCam.com, was testing an automated shopping advisor implemented to assist consumers in choosing digital cameras while shopping in the store. Their task was to visit ForeverCam.com and evaluate the shopping advisor (as well as its product recommendations) at the website. They were also provided with the camera-related preferences of a friend, Pat, and asked to use these preferences to evaluate the shopping advisor's product recommendations^{17,18}.

To motivate participants to take the experimental task seriously, participants were informed before the experiment that, in addition to the point-based incentive provided by the online marketing firm, those who provided detailed, well-supported justifications of their evaluation of the PRA would also get a \$25 cash reward.

Participants were first asked to complete a short questionnaire that collected demographic data (e.g., age, gender) and background information (e.g., level of experience with computers and online shopping, pre-existing risk perception regarding online shopping, product expertise). They were then asked to read a tutorial on how to navigate their assigned e-commerce website and use the PRA embedded in the website.

¹⁷ Consumer preferences are often constructed (rather than absolute) and thus susceptible to the influence of the shopping environment (Haubl & Murray, 2003). By providing participants with a friend's profile, we were giving them a relatively stable set of preferences that they could use to evaluate the PRA's recommendations.

¹⁸ Pat's profile was designed such that the products (available at the e-commerce website) that best matched Pat's preferences were the *dominant* products, rather than the *promoted* products.

Next, they were asked to read a tutorial on digital camera attributes and complete a quiz aimed at testing their understanding of important digital camera attributes. After that, participants were asked to read task instructions and then click on a “Start Shopping” button. Participants in the control condition (i.e., the *no warning* condition) were immediately directed to the e-commerce website to complete the experimental task. For participants in one of the three *warning* conditions, a pop-up window appeared with a warning message before they were directed to the e-commerce website. Upon the completion of the experimental task, participants were asked to fill out a questionnaire that included the measures of the dependent variable.

4.4 Measurement of Dependent Variable

The focal dependent variable, *perceived bias in the PRA’s recommendation*, was assessed by participants’ *perception of anomaly* in the PRA’s recommendations (see Table 3). An anomaly is a deviation from what is expected, usual, or normal; for instance, the presence of inconsistencies between consumers’ requirements and a PRA’s product recommendations, or the observation that products not recommended by the PRA actually fit consumers’ preferences better than those recommended (P. E. Johnson et al., 1993; P. E. Johnson, Grazioli, Jamal, & Berryman, 2001). Based on this definition, three measurement items for *perceived bias* (on 7-point scales) were newly developed for this study. Whereas a high level of *perceived bias* associated with an *unbiased* PRA indicates the presence of false alarms in detecting bias, a high level of *perceived bias* associated with a *biased* PRA indicates hits.

4.4 Pre-test

A series of pre-tests were conducted prior to the main experiment (i) to validate the measurement for *perceived bias*, and (ii) to ensure that the manipulations of warning, risk-handling advice, and framing in our study were successful and that the bias manipulation was reasonable (i.e., neither too easy nor too difficult to detect).

5. DATA ANALYSIS AND RESULTS

This section begins by reporting demographic data about participants in the experiment. Manipulation check results are reported in section 5.2. Results of hypothesis testing are presented in section 5.4.

5.1 Demographic Data

Table E1 (in E-companion) outlines the characteristics of the participants who volunteered in the experiment. More females participated in the study than males. The majority of the participants were between 30-49 years old. Over 50% of the participants used Internet for at least 20 hours each week. Also, more than half of the participants made at least five purchases online during the past 12 months. The demographic profile of the participants is similar to that of online shoppers reported elsewhere (e.g., Pew-Internet, 2008, 2009). On average, participants rated their own product expertise (with digital cameras) at 3.69 (on a seven point scale), suggesting that the participants did lack high expertise in the intended product category. There were no significant differences among the eight experimental groups in terms of product expertise or other demographic/background characteristics.

5.2 Manipulation Checks

Manipulation checks were conducted (see Table 3) for the experimental treatments.

PRA Bias. Results of ANOVA show that the treatment for *PRA bias* was successful. *Perceived bias* was significantly higher for participants in the *biased* PRA condition than for those in the *unbiased* PRA condition ($M = 3.61$ vs. 2.53 , $F(1, 494) = 73.182$, $p < 0.001$).

Availability of Warning Message. ANCOVA results show that, controlling for participants' pre-existing risk perceptions regarding shopping online, those provided with a *warning* message perceived shopping with the assistance of PRA to be significantly more risky than those not provided with such message ($M = 4.14$ vs. 3.43 , $F(1, 493) = 34.938$, $p < 0.001$). Further analysis reveals no significant difference among the three warning conditions in terms of risk perceptions associated with the use of the PRA ($M = 4.31$ vs. 4.00 vs. 4.10 , $F(2, 368) = 1.669$, $p = 0.19$).

Table 3 Manipulation Checks		
Experimental Treatment	Manipulation Check	Statistical Test
PRA Bias	<p>Measured with “perceived bias” (on 7-point Likert scale from “Strongly Disagree” to “Strongly Agree):</p> <ul style="list-style-type: none"> ▪ I noticed inconsistencies between Pat’s requirements and the shopping advisor’s product recommendations ▪ I found that, in each of Pat's preferred brands, some non-recommended products actually fit Pat's preferences better than the products recommended by the shopping advisor. That is to say, products that best meet Pat's preferences were not included in the shopping advisor's recommendations ▪ I have noticed other things (not mentioned above) that are unexpected, unusual, or abnormal in the shopping advisor's product recommendations. 	ANOVA
Availability of Warning Message	<p>Measured before the showing of the warning message (on 7-point Likert scale from “Strongly Disagree” to “Strongly Agree)</p> <ul style="list-style-type: none"> ▪ There is considerable risk involved in shopping online ▪ Shopping online could lead to undesirable consequences ▪ There is a high potential for loss involved in shopping online 	ANCOVA
	<p>Measured immediately after the showing of the warning message but before the start of the experimental task (on 7-point Likert scale from “Strongly Disagree” to “Strongly Agree)</p> <ul style="list-style-type: none"> ▪ Using the shopping advisor to select digital cameras at ForeverCam.com could lead to undesirable consequences ▪ There might be considerable risk involved in using the shopping advisor to select digital cameras at ForeverCam.com ▪ There should be no risk involved in using the shopping advisor to select digital cameras at ForeverCam.com ▪ There might be a high potential for loss involved in using the shopping advisor to select digital cameras at ForeverCam.com 	
Content Design of Warning Message – Inclusion of Risk-Handling Advice	<p>Measured subjectively with the following item (on 7-point Likert scale from “Strongly Disagree” to “Strongly Agree):</p> <ul style="list-style-type: none"> ▪ Just before you started the evaluation task at ForeverCam.com, you read a message that contains not only a warning to consumers but also practical advice for handling risks associated with the automated shopping advisor 	ANOVA
	<p>Captured objectively (with computer logs):</p> <ul style="list-style-type: none"> ▪ Whether the participants used searching-by-brand functionality¹⁹ (as advised in the warning message) 	Cross-Tab Analysis

¹⁹ As all the participants in this study made use of the comparison matrix functionality during the experimental task, we focus our data analysis on the searching-by-brand functionality.

Table 3 Manipulation Checks		
Experimental Treatment	Manipulation Check	Statistical Test
Content Design of Warning Message – Framing of Risk-Handling Advice	<p>Measured subjectively with the following two radio-button type items (after the display of the warning message to participants in the <i>warning with advice</i> experimental conditions):</p> <ul style="list-style-type: none"> ▪ I feel that ... <ul style="list-style-type: none"> • the message stresses the negative implications of failing to verify the shopping advisor's recommendations • the message stresses the positive implications of verifying the shopping advisor's recommendations ▪ I believe that ... <ul style="list-style-type: none"> • I stand to lose important benefits by failing to verify the shopping advisor's recommendations • I stand to gain important benefits by verifying the shopping advisor's recommendations 	Cross-tab Analysis

Content Design of Warning Message – Inclusion of Risk-Handling Advice. Results from both ANOVA and cross-tab analysis show that the treatment for *inclusion of risk-handling advice* (i.e., whether risk-handling advice is included in the warning message) was successful. ANOVA shows that participants' perception of the extent to which the warning message they read contained practical risk-handling advice was significantly higher in the *warning with advice* condition than in the *warning without advice* condition ($M = 5.66$ vs. 5.25 , $F(1, 370) = 4.127$, $p < 0.05$). Cross-tab analysis shows that it was significantly more likely for the participants to use the searching-by-brand functionality (the risk-handling act advocated in the advice included in the warning message) in the *warning with advice* condition than in the *warning with no advice* condition (43.1% vs. 12.9%, $\chi^2(1, N = 496) = 56.224$, $p < 0.001$).

Content Design of Warning Message – Framing of Risk-Handling Advice. Results of cross-tab analysis show that the treatment for the *framing of risk-handling advice* in the warning message was successful. It was significantly more likely for the participants to perceive the warning message they read as emphasizing the *loss* associated with and *negative implications* of failing to adopt the risk-handling behavior advocated in the message in the *warning with negatively-framed advice* condition than in the *warning with positively-framed advice* condition (65.3% vs. 43.5%, $\chi^2(1, N = 248) = 11.85$, $p < 0.01$).

5.3 Measurement Model

There is only one dependent variable for this study, namely, *perceived bias in the PRA's recommendations*. The Cronbach's alpha for *perceived bias* is 0.74, indicating good internal consistency.

5.4 Results of Hypothesis Tests

The mean values of *perceived bias* for different experimental conditions are illustrated in Figure 2.

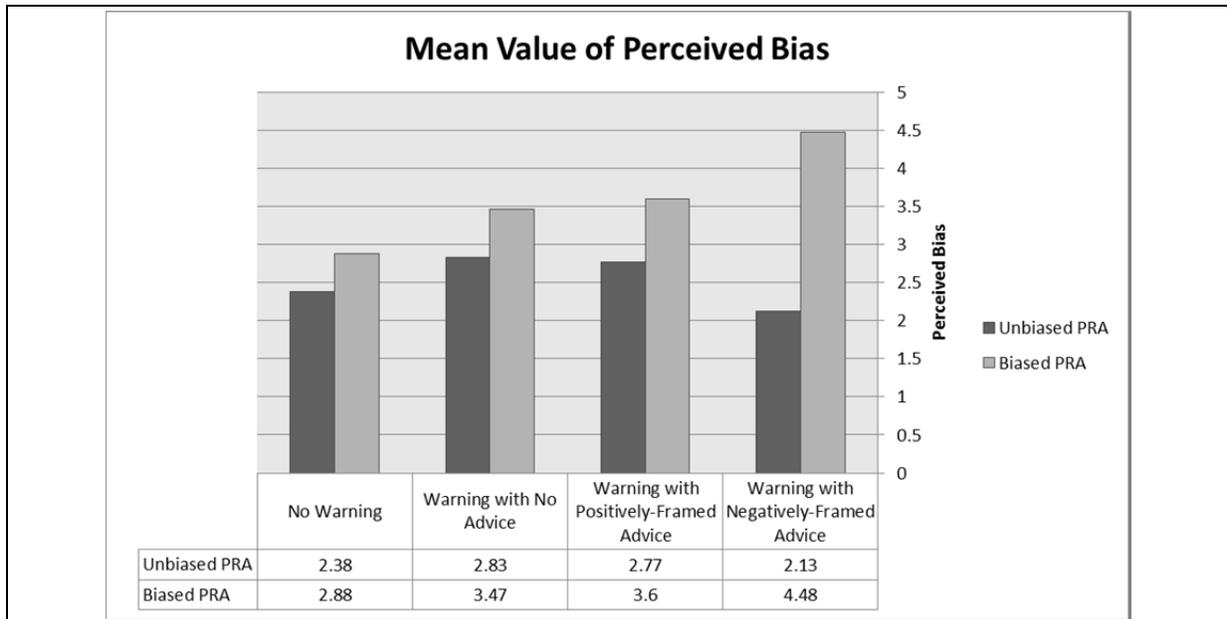


Figure 2 Mean Value of Perceived Bias in Different Experimental Conditions

An ANOVA with planned contrasts was conducted to test for the hypothesized differences in *perceived bias* among different warning conditions. Table 4 presents the results of contrast analysis.

Warning vs. No Warning: As shown in Figure 2 and Table 4, *perceived bias* was significantly higher for participants who received a *warning* than those receiving *no warning* in the *biased* PRA condition ($t(244) = 4.547, p < 0.001$) but not in the *unbiased* PRA condition ($t(244) = 1.196, p = 0.233$), suggesting that participants provided with a *warning* about potential PRA bias had significantly more *hits*, but not more *false alarms*, than those in the *no warning* condition. These results support **H1a** but not **H1b**.

Table 4 ANOVA Planned Contrasts for Noticing of Anomaly

Group A	Group B	Diff. in Mean ²⁰	df	t	Sig.	Related Hypothesis
Unbiased PRA Condition						
Warning ²¹	No Warning	0.21	244	1.196	.233	H1b
Warning with Advice ²²	Warning with No Advice	-0.38	244	2.054	.041	H2b
Warning with Negatively-Framed Advice	Warning with Positively-Framed Advice	-0.64	244	3.003	.003	H3b
Biased PRA Condition						
Warning ²³	No Warning	0.99	244	4.547	.000	H1a
Warning with Advice ²⁴	Warning with No Advice	0.57	244	2.457	.015	H2a
Warning with Negatively-Framed Advice	Warning with Positively-Framed Advice	0.88	244	3.324	.001	H3a

Further analysis comparing the *no warning* condition to each of the three *warning* conditions (i.e., warning with no advice, warning with positively-framed advice, and warning with negatively-framed advice) reveals an interaction between the availability and the content design of warning messages. More specifically, in the *biased PRA* condition, participants who received any one of the three types of warning message had significantly higher level of *perceived bias* than those receiving *no warning* ($t(244) = 2.250$, $p < 0.05$; $t(244) = 2.716$, $p < 0.01$; $t(244) = 6.039$, $p < 0.001$). In contrast, in the *unbiased PRA* condition, while participants who received either a *warning with no advice* or a *warning with positively-framed advice* had significantly higher level of *perceived bias* than those receiving *no warning* ($t(244) = 2.120$, $p < 0.05$; $t(244) = 1.842$, $p = 0.067$), the difference in *perceived bias* was not significant between participants who received a *warning with negatively-framed advice* and those receiving *no warning* ($t(244) = 1.161$, $p = 0.247$). Thus, when there was no bias in the PRA's recommendations, providing participants with a *warning message that included negatively-framed advice* did not increase their

²⁰ Difference in Mean = Group A Mean – Group B Mean

²¹ Include warning with no advice, warning with positively-framed advice, and warning with negatively-framed advice.

²² Include warning with positively-framed advice and warning with negatively-framed advice.

²³ Include warning with no advice, warning with positively-framed advice, and warning with negatively-framed advice.

²⁴ Include warning with positively-framed advice and warning with negatively-framed advice.

perception of PRA bias as did the provision of the other two types of warning messages. In sum, the provision of any type of warning message resulted in an increased *hit* rate. In addition, compared to the *no warning* condition, while *warning with no advice* and *warning with positively-framed advice* led to a higher *false alarm* rate, *warning with negatively-framed advice* did not result in an increase in *false alarms*.

Warning with Advice vs. Warning with No Advice: There was significant difference in *perceived bias* between participants who received a *warning with advice* (*positively-framed* or *negatively-framed*) and those receiving a *warning with no advice*, both in the *unbiased* PRA condition ($t(244) = 2.054, p < 0.05$) and in the *biased* PRA condition ($t(244) = 2.457, p < 0.05$), indicating that participants assigned to the *warning with advice* conditions had significantly more *hits* and significantly fewer *false alarms* than those assigned to the *warning with no advice* condition. These results support **H2a** and **H2b**.

Further analysis reveals an interaction between the two content design characteristics of warning messages (i.e., the inclusion and the framing of risk-handling advice). More specifically, there was no significant difference in *perceived bias* between the *warning with positively-framed advice* group and the *warning with no advice* group, in either the *unbiased* PRA condition ($t(244) = 0.278, p = 0.782$) or the *biased* condition ($t(244) = -0.466, p = 0.642$). On the other hand, for participants who received a *warning with negatively-framed advice*, *perceived bias* was significantly *higher* (*lower*) in the *biased* (*unbiased*) PRA condition than those receiving a *warning with no advice* ($t(244) = 3.790, p < 0.001$; $t(244) = 3.280, p < 0.01$). Thus, while the inclusion of *negatively-framed* risk-handling advice in a warning message significantly improved participants' detection performance, a warning message with *positively-framed* advice did not outperform a *warning message with no advice* in supporting bias detection.

Warning with Negatively-Framed Advice vs. Warning with Positively-Framed Advice: In the *biased* (*unbiased*) PRA condition, the *perception of bias* was significantly *higher* (*lower*) for those who received a *warning with negatively-framed advice* than for those receiving a *warning with positively-framed advice* ($t(244) = 3.324, p < 0.01$; $t(244) = 3.003, p < 0.01$), suggesting that participants assigned to the *warning with negatively-framed advice* condition had more *hits* and fewer *false alarms* than those

assigned to the *warning with positively-framed advice* condition. These results support **H3a** and **H3b**.

Supplementary analyses. Supplementary analyses were conducted to explore the pathways by which the availability and design characteristics of warning messages influence participants' performance in detecting bias. First, ANCOVA results show that, controlling for participants' pre-existing risk perceptions regarding shopping online, those provided with a *warning* message perceived shopping with the assistance of PRA to be significantly more risky than those not provided with such message ($M = 4.14$ vs. 3.43 , $F(1, 493) = 34.938$, $p < 0.001$), suggesting that the *availability of warning messages* indeed lowered participants' *decision threshold* by increasing the expected base rate of manipulative practices. Results of a regression analysis also reveal significant positive relationship between *perceived risk* of shopping with the assistance of PRA and *perceived bias* in the PRA's recommendations ($F(1, 494) = 13.122$, $p < 0.001$).

In addition, results of cross-tab analysis reveal that it was significantly more likely for the participants to use the searching-by-brand functionality (the risk-handling act advocated in the warning message that included advice) (1) in the *warning with advice* condition than in the *warning with no advice* condition (43.1% vs. 12.9%, $\chi^2(1, N = 496) = 56.224$, $p < 0.001$), and (2) in the *warning with negatively-framed advice* condition than in the *warning with positively-framed advice* condition (50% vs. 36.3%, $\chi^2(1, N = 248) = 4.751$, $p < 0.05$), suggesting that the two content design characteristics of warning messages did influence participants' *discriminant ability* (by providing them with right tool/information). An ANOVA was also run with the predictor being the *use of searching-by-brand functionality* and dependent variable being *perceived bias*. Results of the analysis show that, in the *biased PRA* condition, the level of *perceived bias* was significantly *higher* for those who used searching-by-brand functionality than those who did not ($M = 4.53$ vs. 3.23 , $F(1, 246) = 40.658$, $p < 0.001$). However, in the *unbiased PRA* condition, *perceived bias* did not differ significantly between those who used searching-by-brand functionality and those who did not ($M = 2.37$ vs. 2.58 , $F(1, 246) = 1.547$, $p > 0.1$). The results suggest that participants who used searching-by-brand functionality had higher detection performance (indicated by more *hits* and similar number of *false alarms*) than those who did not use this functionality.

Insomuch as *perceived risk*²⁵ and *use of searching-by-brand functionality*²⁶ could be considered surrogates for *decision threshold* and *discriminant ability* respectively, the results of the supplementary analyses suggest that the *availability* and *content design* (i.e., the inclusion and framing of risk-handling advice) of warning messages affect consumers' performance in detecting PRA bias via their influence on consumers' *decision threshold* and *discriminant ability*, respectively.

6. DISCUSSION AND CONCLUSIONS

This section summarizes the major findings of this study, presents its contributions to research and practice, and discusses its limitations.

6.1 Discussion of Findings

Most of the hypotheses advanced are supported by data (except for H1b, which is partially supported). The three warning mechanisms examined in this study, namely, *warning with no advice*, *warning with positively-framed advice*, and *warning with negatively-framed advice* are effective, albeit to different extents, in supporting consumers in the task of detecting PRA manipulation-based anomalies. While all three warning mechanisms were effective in enhancing the *hit* rates in detecting anomalies in a *biased* PRA (with *warning messages with negatively-framed advice* producing the greatest number of hits), providing consumers with a *warning message with no advice* or a *warning message with positively-framed advice* also increased the likelihood that consumers would perceive bias in the recommendations of an *unbiased* PRA, resulting in higher *false alarm* rates. Overall, the results of the study show that, in the PRA-assisted shopping context, the provision of *warning messages with negatively-framed advice* is the most effective mechanism in supporting consumers in bias detection.

In sum, the results of this study provide answers to the research questions that initially motivated the research. The inclusion of risk-handling advice in a warning message does not uniformly enhance

²⁵ *Perceived risk* could be considered a surrogate for *decision threshold* because heightened perception of risk will result in an increase in the expected base rate of bias, an antecedent to lowered decision threshold.

²⁶ *Use of searching-by-brand functionality* could be considered a surrogate for *discriminant ability* because, when consumers engage in independent verification of a PRA's recommendations using the searching-by-brand functionality, they will be better able to distinguish a biased PRA from an unbiased one.

consumers' performance in detecting biased recommendations from a PRA; the persuasiveness of the risk-handling advice is contingent upon how it is framed. More specifically, compared to *positively-framed advice* accentuating the gain from following the advice, *negatively-framed advice* emphasizing the loss from not following the advice is more effective in motivating consumers to engage in the risk-handling behavior advocated in the advice and thus achieve superior detection performance (indicated by more *hits* and fewer *false alarms*).

6.2 Contributions to Research

This study makes significant contributions to research. First, it contributes to IS literature by examining factors relevant to a potential detection support mechanism – warning – that influence consumers' performance in detecting biased product recommendations from e-commerce PRAs. Drawing on Signal Detection Theory, and its associated mechanisms of decision threshold and discriminant ability, as well as the literature on warning and message framing, we rationalize that a *warning* message by its nature places consumers in a heightened state of alertness and lowers their *decision threshold* in judging ambiguous cues as anomalies resulting from bias, which may increase the likelihood that consumers will perceive bias in the recommendations of both a biased PRA and an unbiased one, resulting in higher rates of false alarms as well as higher rates of hits. In addition, we contend that the *inclusion of risk-handling advice* in a warning message (a content design characteristic identified in our study) provides consumers with strategies to address the risk of PRA bias and thus enhances consumers' *discriminant ability* to compensate for the impact of lowered decision threshold, resulting in high rates of hits without simultaneously increasing false alarm rates. As the risk-handling advice included in the warning message is effective only to the extent that consumers adopt the risk-handling behavior (i.e., taking extra search action) advocated in the advice, we further postulate that, in the context of detecting PRA bias (a detection behavior involving the possibility of unpleasant outcome), *negative-framed advice* included in a warning message will be more effective than *positively-framed advice* in motivating consumers to engage in the risk-handling behavior the advice encourages.

The findings of our study provide strong evidence that although consumers are generally vulnerable to online manipulations, properly designed detection support mechanisms (such as warning) can be implemented to help consumers perform better in detecting such practices. While all three warning mechanisms examined in this study (i.e., *warning with no advice*, *warning with positively-framed advice*, and *warning with negatively-framed advice*) are effective in enhancing the *hit* rates in detecting bias, *warning with negatively-framed advice* is the only intervention that achieves this without increasing *false alarms*. Inasmuch as high hit rates and low false alarm rates are indicators of superior detection performance (Xiao, 2010), this study highlights *warning with negatively-framed advice* as the most effective mechanism in supporting consumers in detecting bias in the PRA's recommendations.

Second, this study fills a void in existing research in warning by examining the design characteristics of warning messages, rather than focusing on the presence or absence of warning messages. Integrating prior research in product warning labels with literature in message framing literature, this study identifies two *content design characteristics* of warning message (i.e., inclusion of risk-handling advice and framing of risk-handling advice) and investigates their impact on consumers' detection performance. The findings of this study enhance our understanding of the effectiveness of different warning design in improving detection performance and lay a foundation for future empirical as well as theoretical work.

Third, this study adds to an increasing body of IS research in persuasive communication (e.g., Angst & Agarwal, 2009; Cheng & Wu, 2010; Johnston & Warkentin, 2010). Extending insights from studies of goal framing effectiveness in health decision making to the context of designing warning messages to support consumers in detecting PRA bias, this study postulates and confirms previous findings with strong empirical evidence that, when the behavior being promoted is a *detection* (as opposed to *prevention*) behavior, persuasive messages framed in terms of potential losses are more effective in motivating targeted behavior than factually equivalent messages framed in terms of potential gains. The powerful effect of goal framing revealed in this study also suggests that the framing manipulation employed in this study is successful and can be adopted (and further tested) by other researchers in future research.

6.3 Contributions to Practice

This study reveals that consumers are extremely vulnerable to biased online PRAs, a finding consistent with prior research in manipulative online business practices (Grazioli, 2004; Grazioli & Jarvenpaa, 2000, 2001, 2003a, 2003b; Grazioli & Wang, 2001). The mean value of *perceived bias* was 3.61 for the biased PRA condition, which was at about the mid-point of the 7-point scale. Of all the 248 participants assigned to the biased PRA condition, only 114 of them (or 46%) had an average *perceived bias* score over 4.0, indicating that they have perceived bias in the PRA's recommendations. Even for those who had received warnings about potential PRA bias, only 99 of them (out of 186, approximately 53%) had an average *perceived* score over 4.0, suggesting the difficulty for consumers to detect bias in the PRA's recommendations. Therefore, it is imperative for government agencies (e.g., Federal Trade Commission in the United States, Office of Consumer Affairs in Canada), industry leaders (e.g., Amazon, Expedia, Dell), and consumer organizations (e.g., National Consumers League in the United States, Consumers' Association of Canada) to not only promote consumer awareness of novel manipulative practices made possible by innovative technologies supporting e-commerce but also educate consumers on strategies to address such risks. For instance, consumers may be advised to verify PRAs' recommendations by using the comparison and searching tools both available at the e-commerce website and outside of the website (for instance, at an independent, third party comparison shopping platform).

In addition to promoting consumer awareness and education, it is also important for government, industry, and consumer organizations to establish guidelines for good business practice for online companies. For example, a guideline may strongly recommend, if not require, e-commerce websites featuring a PRA to implement functionalities that allow consumers to sort/search/compare all products (not only those recommended by the PRA) by brand, price, and important product features; such functionalities will enable consumers to detect bias in the PRA's recommendations more easily. The guideline may also recommend that, in addition to providing an explanation regarding the PRA's

reasoning logic, the website should provide a fit score for each product (in the recommendation list) and detailed explanation as to how the score is calculated. Strong, enforceable self-regulation and self-policing by industry members are critical to addressing the growing problem of PRA bias.

This study also offers insights into the design and effectiveness of different warning mechanisms, not only in the context of detecting biased PRAs but also in anomaly detection context in general. Prior empirical evidence indicates that explicit *warning* about the possibility of manipulation may increase the *hit* rate at the cost of increased *false alarms* (Burgoon et al., 1994; Parasuraman, 1984). False alarms may be tolerated in an environment where the danger of strategic manipulation is real and deplorable, and thus boosting the hit rates is more important than reducing the false alarm rates (Xiao & Benbasat, 2015). However, in many cases, false alarms are costly, because they will have to be evaluated individually to confirm/disconfirm their accuracy, which demands resources, time, and effort (George et al., 2004).

From a business perspective, the most desirable detection support mechanism is one that would enhance detection performance without incurring the additional costs associated with increased false alarms (George et al., 2004). Ultimately, the purpose of any detection support mechanism is to protect consumers in e-commerce transactions rather than deterring them from engaging in e-commerce. Examining empirically the design characteristics of warning messages (a commonly used detection support mechanism), this study shows that properly designed warning messages can enhance consumers' performance (indicated by high hit rate and low false alarm rate) in detecting anomalies resulting from strategic manipulations. The results of the study further reveal that a promising intervention is to provide consumers with *warning messages that include negatively-framed risk-handling advice*, which encourages consumers to exploit the search and compare functionalities typically available at e-commerce websites to protect themselves from being misled when shopping online. It is understandable that online merchants may be reluctant to warn consumers of potential manipulative practices. Thus, the warning intervention can be implemented as browser extension (similar to existing anti-phishing, anti-fraud, and

privacy protection extensions to browsers²⁷) available from third-party consumer protection organizations. Once the extension is downloaded and installed to a browser by consumers, a warning message can be triggered when consumers visit e-commerce websites that implement PRAs. Nevertheless, if a merchant does voluntarily integrate the warning mechanism to its website, the merchant may be perceived by consumers as benevolent and honest, and the consequent increased trust may lead to more customers following the recommendations of the PRA (featured at the merchant's website) and buying from the merchant.

6.4 Limitations and Suggestions for Future Research

A number of limitations are involved with this study. First, the study examines one way by which online PRAs provide biased product recommendations to consumers, namely, to recommend products from a set of promoted products rather than from all available products. The application of this study's findings to other types of manipulation techniques requires caution. Other manipulation techniques (e.g., those designed to influence consumers' decision criteria by selectively including product attributes in the PRA's preference-elicitation interface or by providing false decisional guidance) (see Xiao & Benbasat, 2011) may have different effect on consumers' performance in detecting anomalies in PRAs, hence additional research is needed.

Second, in this study, the different types of warning messages were implemented as being issued by a third-party organization independent from the e-commerce website featuring the PRA. We also suggest that, in real world, the warning intervention (as well other similar detection support mechanisms) should be implemented as third-party, freely downloadable browser extension. While our study demonstrates the effectiveness of warning mechanisms (particularly the one that includes risk-handling advice) in supporting consumers in detecting biased PRAs, future research is needed to investigate whether and how such mechanisms can be employed effectively in practice. For instance, what factors will motivate consumers to exert the effort to download and install the browser extension? Under what circumstances

²⁷ Examples include TRUSTe Tracker Protection extension for Firefox (which discovers trackers including cookies, etags, beacons, etc.) and Netcraft Anti-Phishing extension for Chrome.

and how often should warning messages be displayed to cause minimum interruption and annoyance to consumers' shopping process? Could our suggested implementation of warning mechanisms induce a backlash (and further strategic actions) from online merchants? If so, how can their negative impact be mitigated? Additionally, while *warning with negatively-framed advice* outperforms the other two types of warning messages (i.e., *warning with no advice* and *warning with positively-framed advice*), it still results in a false alarm rate of 22.58% (with false alarm rates being 40.98% and 37.70% for consumers receiving the other two types of warning messages) (see Table A1 in Appendix), suggesting the need for further improvement in the design of warning messages.

Third, the advice (included in the warning messages) examined in this study, which encourages consumers to exert extra search effort using the search and compare functionalities available at e-commerce websites, is only one example of how risk-handling advice can be phrased to provide consumers with knowledge and skills to address the risk of PRA bias. The focus of our study is **not** on *what advice* to provide, but rather on *how to better design* warning messages and the advice included, so as to motivate consumers to change their behavior in accordance with the advice (whatever that is) included in the warning messages. Given the multitude of means by which bias can be introduced by PRAs, future research is needed to explore what the most appropriate informational content of risk-handling advice is for biases introduced from various other sources.

Fourth, in our study, the “comparison matrix” and “search-by-brand” functionalities were built into the e-commerce website. We acknowledge that, in reality, unscrupulous online merchants may be reluctant to provide such functionalities on their websites. Nevertheless, as functionalities that facilitate search and encourage discovery have become *must-have features* of e-commerce sites (Dunn, 2013; Mulpuru, Evans, & Roberge, 2013), their absence may be a cause for consumer alarm (particularly when consumers receive warning messages from an independent party). Further, consumers can make use of other websites to search and compare products of interest. Online merchants delivering biased recommendations are thus caught in a bind on whether or not to provide comparison and search functionalities on their websites – in either case, the biases may be revealed. Future research may

explicitly compare consumers' detection performance when these functionalities are present or absent on e-commerce websites.

Lastly, this study empirically examines the effect of different design characteristics of warning messages on consumers' bias detection performance. Future studies may explore other detection support mechanisms (both non-IT-based and IT-based; see Xiao & Benbasat, 2011 for examples), which can be implemented in the e-commerce context to support consumers in detecting the strategic manipulations perpetrated by online companies.

6.5 Conclusion

The increasing adoption of PRAs by e-commerce merchants makes it a desirable area of study for IS researchers. This study represents an early effort to look into PRAs that base recommendations on business criteria (rather than on consumer benefits) and examine how warning messages can be designed to support consumers in detecting biased recommendations by PRAs. We have identified two *content design characteristics* of warning message -- the inclusion of risk-handling advice and framing of risk-handling advice -- and investigated how they influence consumers' detection performance. Our results reveal that the provision of a simple warning message without advice is a two-edged sword, as it increases correct detection of biased PRAs (hits) at the cost of increased incorrect detection (false alarms). In contrast, including in warning messages risk-handling advice about how to check for bias (particularly when the advice is framed to emphasize the loss from not following the advice) increases correct detection, but more importantly, also decreases incorrect detection. Our study also provides evidence suggesting that these effects are attributable to the two theoretical mechanisms of Signal Detection Theory, namely, decision threshold and discriminant ability. With an enriched understanding of how the *availability* and the *content design* of warning messages can support consumers in the context of PRA-assisted online shopping, these results serve as a basis for future theoretical development and yield valuable insights that can guide practice.

Acknowledgments

We would like to thank the Natural Science and Engineering Research Council of Canada (NSERC) and the Social Science and Humanities Research Council of Canada (SSHRC) for their support of this research. We are grateful to the senior editor, the associate editor, and the anonymous reviewers for their constructive feedback during the review process.

References

- Aksoy, L., & Bloom, P. N. (2001, February 15-17). *Impact of Ordered Alternative Lists on Decision quality: The Role of Perceived Similarity*. Paper presented at the The Society for Consumer Research Winter Conference, Scottsdale, AZ.
- Angst, C. M., & Agarwal, R. (2009). Adoption of electronic health records in the presence of privacy concerns: the elaboration likelihood model and individual persuasion. *MIS quarterly*, 33(2), 339-370.
- Argo, J. J., & Main, K. J. (2004). Meta-analyses of the effectiveness of warning labels. *Journal of Public Policy & Marketing*, 23(2), 193-208.
- Banks, S. M., Salovey, P., Greener, S., Rothman, A. J., Moyer, A., Beauvais, J., et al. (1995). The effects of message framing on mammography utilization. *Health Psychology*, 14, 178-184.
- Biros, D. P. (1998). *The effects of truth bias on artifact-user relationships: An investigation of factors for improving deception detection in artifact produced information*. Unpublished 9839756, The Florida State University, United States -- Florida.
- Biros, D. P., George, J. F., & Zmud, R. W. (2002). Inducing sensitivity to deception in order to improve decision making performance: A field study. *MIS Quarterly*, 26(2), 119-144.
- Burgoon, J. K., Buller, D. B., Ebesu, A. S., & Rockwell, P. (1994). Interpersonal deception: V. accuracy in deception detection. *Communication Monographs*, 61(4), 303-325.
- Chau, P. Y., Ho, S. Y., Ho, K. K., & Yao, Y. (2013). Examining the effects of malfunctioning personalized services on online users' distrust and behaviors. *Decision Support Systems*, 56, 180-191.
- Cheng, F. F., & Wu, C. S. (2010). Debiasing the framing effect: The effect of warning and involvement. *Decision Support Systems*, 49(3), 328-334.
- Cosley, D., Lam, S. K., Albert, I., Konstan, J., & Riedl, J. (2003, April 5-10). *Is Seeing Believing? How Recommender Systems Influence Users' Opinions*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems, Fort Lauderdale, FL.
- Davis, D. R., & Parasuraman, R. (1981). *The Psychology of Vigilance*. London: Academic Press.
- Detweiler, J. B., Bedell, B. T., Salovey, P., Pronin, E., & Rothman, A. J. (1999). Message framing and sun screen use: Gain-framed messages motivate beach-goers. *Health Psychology*, 18, 189-196.
- Dunn, T. (2013). *Filter & sort: Improving ecommerce product findability*. Retrieved June 30, 2013, from <http://www.webcredible.co.uk/user-friendly-resources/web-usability/ecommerce-findability.shtml>
- Fishbein, M., & Ajzen, I. (1981). Acceptance, Yielding, and Impact: Cognitive Processes in Persuasion. In R. E. Petty, T. M. Ostrom & T. C. Brock (Eds.), *Cognitive Responses in Persuasion* (pp. 339-359). Hillsdale, NJ: Lawrence Erlbaum Associates.
- George, J. F., & Marett, K. (2004). *Inhibiting detection and its detection*. Paper presented at the 37th Annual Hawaii International Conference on System Sciences, Hawaii.
- George, J. F., Marett, K., & Tilley, P. (2004). *Deception detection under varying electronic media and warning conditions*. Paper presented at the 37th Hawaii International Conference on System Sciences.
- Gerend, M. A., & Sias, T. (2009). Message framing and color priming: How subtle threat cues affect persuasion. *Journal of Experimental Social Psychology*, In Press.

- Grazioli, S. (2004). Where Did They Go Wrong? An Analysis of the Failure of Knowledgeable Internet Consumers to Detect Deception Over the Internet. *Group Decision and Negotiation*, 13(2), 149.
- Grazioli, S., & Jarvenpaa, S. L. (2000). Perils of Internet fraud: An empirical investigation of deception and trust with experienced Internet consumers. *IEEE Transaction on Systems, Man, and Cybernetics*, 30(4), 395-410.
- Grazioli, S., & Jarvenpaa, S. L. (2001). *Tactics used against consumers as victims of Internet deception*. Paper presented at the Seventh Americas Conference on Information Systems.
- Grazioli, S., & Jarvenpaa, S. L. (2003a). Consumer and business deception on the Internet: content analysis of documentary evidence. *International Journal of Electronic Commerce*, 7(4), 93-118.
- Grazioli, S., & Jarvenpaa, S. L. (2003b). Deceived: under target online. *Communications of the ACM*, 46(12), 196-205.
- Grazioli, S., & Wang, A. (2001). *Looking without seeing: understanding unsophisticated consumers' success and failure to detect Internet deception*. Paper presented at the Twenty-Second International Conference on Information Systems.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley.
- Haubl, G., & Murray, K. (2003). Preference Construction and Persistence in Digital Marketplaces: The Role of Electronic Recommendation Agents. *Journal of Consumer Psychology*, 13(1&2), 75-91.
- Haubl, G., & Murray, K. B. (2006). Double Agents: Assessing the Role of Electronic Product-Recommendation Systems. *Sloan Management Review*, 47(3), 8-12.
- Haubl, G., & Trifts, V. (2000). Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids. *Marketing Science*, 19(1), 4-21.
- Hill, D. J., King, M. F., & Cohen, E. (1996). The perceived utility of information presented via electronic decision aids: A consumer perspective. *Journal of Consumer Policy*, 19, 137-166.
- Johnson, E. J., Lohse, G. L., & Mandel, N. (2001). *Designing marketplaces of the artificial: Four approaches to understanding consumer behavior in electronic environments*. Unpublished Working paper, Columbia.
- Johnson, P. E., Grazioli, S., & Jamal, K. (1993). Fraud detection: Intentionality and deception in cognition. *Accounting, Organizations and Society*, 18(5), 467-488.
- Johnson, P. E., Grazioli, S., Jamal, K., & Berryman, R. G. (2001). Detecting deception: adversarial problem solving in a low base-rate world. *Cognitive Science*, 25, 355-392.
- Johnston, A. C., & Warkentin, M. (2010). Fear appeals and information security behaviors: an empirical study. *MIS quarterly*, 34(3), 549-566.
- Kelley, C. A., Gaidis, W. C., & Reingen, P. H. (1989). The use of vivid stimuli to enhance comprehension of the content of product warning messages. *Journal of Consumer Affairs*, 23(2), 243-266.
- King, M. F., & Hill, D. J. (1994). Electronic decision aids: Integration of a consumer perspective. *Journal of Consumer Policy*, 17(2), 181-206.
- Komiak, S. Y., & Benbasat, I. (2008). A two-process view of trust and distrust building in recommendation agents: A process-tracing study. *Journal of the Association for Information Systems*, 9(12), 727-747.
- Levin, I. P., Gaeth, G. J., & Schreiber, J. (2002). A new look at framing effects: Distribution of effect sizes, individual differences, and independence of types of effects. *Organizational Behavior and Human Decision Processes*, 88(1), 411-429.
- Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, 76(2), 149-188.
- Limelight. (2010). Limelight Networks® Research Identifies Online Commerce Website Features That Create Brand Loyalty.
- Marett, K., & George, J. F. (2005). *Group deception in computer-supported environments*. Paper presented at the 38th Annual Hawaii International Conference on System Sciences, Hawaii.

- Meyerowitz, B. E., & Chaiken, S. (1987). The effect of message framing on breast self-examination attitudes, intentions, and behavior. *Journal of Personality and Social Psychology*, 52(500-510).
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81-97.
- Miller, G. R., & Stiff, J. B. (1993). *Deceptive Communication*. Newbury Park, CA: Sage Publications.
- Morrison, E. W., & Robinson, S. L. (1997). When employees feel betrayed: A model of how psychological contract violation develops. *Academy of Management Review*, 22, 226-256.
- Mulpuru, S., Evans, P. F., & Roberge, D. (2013). *Must-Have eCommerce Features*: Forrester.
- O'Keefe, D. J., & Jensen, J. D. (2009). The Relative Persuasiveness of Gain-Framed and Loss-Framed Messages for Encouraging Disease Detection Behaviors: A Meta-Analytic Review. *Journal of Communication*, 59, 296-316.
- Parasuraman, R. (1984). Sustained attention in detection and discrimination. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of Attention* (pp. 243-266). London: Academic Press.
- Pew-Internet. (2008). *Demographic profile of online shoppers*. Retrieved October 29, 2009, from <http://blogs.zdnet.com/ITFacts/?p=14022>
- Pew-Internet. (2009). *Usage over time*. Retrieved October 30, 2009, from <http://www.pewinternet.org/Static-Pages/Trend-Data.aspx>
- Punj, G., & Moore, R. (2009). Information search and consideration set formation in a web-based store environment. *Journal of Business Research*, 62, 644-650.
- Robinson, S. L. (1996). Trust and breach of the psychological contract. *Administrative Science Quarterly*, 41(4), 574.
- Ross, K. (1981). Legal and practical considerations for the creation of warning labels and instruction books. *Journal of Products Liability*, 4, 29-45.
- Rothman, A. J., Bartels, R. D., Wlaschin, J., & Salovey, P. (2006). The strategic use of gain- and loss-framed messages to promote healthy behavior: How theory can inform practices. *Journal of Communication*, 56, S202-S220.
- Rothman, A. J., & Salovey, P. (1997). Shaping perceptions to motivate healthy behavior: The role of message framing. *Psychological Bulletin*, 121, 3-19.
- Rothman, A. J., Salovey, P., Antone, C., Keough, K., & Martin, C. D. (1993). The influence of message framing on intentions to perform health behaviors. *Journal of Experimental Social Psychology*, 29, 408-433.
- Scott, E. D. (2006). Just(?) a true-false test: Applying signal detection theory to judgments of organizational dishonesty. *Business & Society*, 45(2), 130-148.
- Senecal, S. (2003). *Essays on the influence of online relevant others on consumers' online product choices*. Unpublished doctoral dissertation, University of Montreal, Montreal, Quebec.
- Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, 80(2), 159-169.
- Stiff, J. B., Kim, H. J., & Ramesh, C. N. (1992). Truth biases and aroused suspicion in relational deception. *Communications Research*, 19(3), 326-345.
- Tilley, P. A. (2005). *Training, Warning, and Media Richness Effects on Computer-Mediated Deception and its Detection*. Unpublished Doctoral Dissertation, The Florida State University.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211, 453-458.
- Wang, W., & Benbasat, I. (2005). Trust In and Adoption of Online Recommendation Agents. *J AIS*, 6(3).
- Wickens, T. D. (2001). *Elementary signal detection theory*: Oxford University Press.
- Witte, K., & Allen, M. (2000). A meta-analysis of fear appeals: Implications for effective public health campaigns. *Health Education & Behavior*, 27(5), 591-615.
- Xiao, B. (2010). *Product-Related Deceptive Information Practices in B2C E-Commerce: Formation, Outcomes, and Detection*. Unpublished Doctoral Dissertation, University fo British Columbia, Vancouver.

- Xiao, B., & Benbasat, I. (2007). E-Commerce Product Recommendation Agents: Use, Characteristics, and Impact. *MIS Quarterly*, 31(1), 137-209.
- Xiao, B., & Benbasat, I. (2011). Product-Related Deception in E-Commerce: A Theoretical Perspective. *MIS Quarterly*, 35(1), 169-195.
- Xiao, B., & Benbasat, I. (2014). Research on Use, Characteristics, and Impact of E-Commerce Product Recommendation Agents: A Review and Update for 2007-2012. In F. J. Martínez-López (Ed.), *Research Handbook on EBusiness Strategic Management* (Vol. 3, pp. 403-431): Springer.
- Xiao, B., & Benbasat, I. (2015). Consumer vulnerability to biased product recommendations in online buying: The influence of perceived personalization: University of Hawaii at Manoa.
- Xu, J., Benbasat, I., & Cenfetelli, R. T. (2014). The Influences of Online Service Technologies and Task Complexity on Efficiency and Personalization. *Information Systems Research*, 25(3), 420-436.

Appendix: Hit Rates and False Alarm Rates

Upon completion of the experimental task at the e-commerce website, participants were asked three questions (with 7-point scale) about whether they had perceived bias in the PRA’s recommendations. Following Grazioli (2004), we coded the responses of participants who answered “Mildly Agree”, “Agree”, or “Strongly Agree” to any of the three questions as “1”, indicating that these participants did perceive bias in the PRA’s recommendations. In contrast, responses of participants who answered otherwise (i.e., “Strongly Disagree”, “Disagree”, or “Mildly Disagree”) were coded as “0”, meaning that the participants did not perceive bias in the PRA’s recommendations. Finally, responses of participants (14 in total) who answered “Neutral” (i.e., the mid-point of the 7-point scale) were labeled “undecided” and excluded from the subsequent analysis. Table A1 summarizes the results of the coding practice, on the basis of which we derived the *hit rate* and *false alarm rate* for each experimental condition.

Table A1 Results of Coding Practice							
		Unbiased PRA		Biased PRA		Hit Rate	False Alarm Rate
		Bias Perceived		Bias Perceived			
		0 (No)	1 (Yes)	0 (No)	1 (Yes)		
No Warning	Count	50	10	38	21	35.59%	16.67%
Warning with No Advice	Count	36	25	23	34	59.65%	40.98%
Warning with Positively- Framed Advice	Count	38	23	21	40	65.57%	37.70%
Warning with Negatively- Framed Advice	Count	48	14	10	51	83.61%	22.58%