



Winning the Battle but Losing the War: Ironic Effects of Training Consumers to Detect Deceptive Advertising Tactics

Andrew E. Wilson¹ · Peter R. Darke² · Jaideep Sengupta³

Received: 18 May 2020 / Accepted: 29 August 2021 / Published online: 8 September 2021
© The Author(s) 2021

Abstract

Misleading information pervades marketing communications, and is a long-standing issue in business ethics. Regulators place a heavy burden on consumers to detect misleading information, and a number of studies have shown training can improve their ability to do so. However, the possible side effects have largely gone unexamined. We provide evidence for one such side-effect, whereby training consumers to detect a specific tactic (illegitimate endorsers), leaves them more vulnerable to a second tactic included in the same ad (a restrictive qualifying footnote), relative to untrained controls. We update standard notions of persuasion knowledge using a goal systems approach that allows for multiple vigilance goals to explain such side-effects in terms of *goal shielding*, which is a generally adaptive process by which activation and/or fulfillment of a low-level goal inhibits alternative detection goals. Furthermore, the same goal systems logic is used to develop a more general form of training that activates a higher-level goal (general skepticism). This more general training improved detection of a broader set of tactics without the negative goal shielding side effect.

Keywords Persuasion knowledge model · Goal systems theory · Deceptive advertising · Deception · Consumer training

Introduction

Marketing communication has been described as pervaded by deception (Boush et al., 2009), which has long been recognized as an important issue in business ethics (e.g. Attas, 1999; Stearns & Borna, 1995; Xie et al., 2015). According to the *Federal Trade Commission* (FTC; Anderson, 2019) over 40 million consumers (more than 15% of the US adult population) were victims of fraud or deception during 2017 (the most recent data available), about three-quarters of which involved advertising or sales. Regulation can be costly and

difficult (Cain, 2011; Peltzman, 1981; Russo et al., 1981), and as a result regulatory systems place a heavy burden on consumers to detect misleading tactics for themselves (Cohen, 2019; Preston, 1994; Stewart & Martin, 2004). Such observations underlie continued calls for attention to the question of how best to improve consumer competency in identifying misleading offers (e.g. Boush et al, 2009; Cialdini, 2021; Pratkanis & Aronson, 2001).

Consumers learn to detect misleading tactics through the acquisition of persuasion knowledge (PK; Friestad & Wright, 1994), and by learning to effectively neutralize, resist, and penalize misleading tactics (Boush et al., 2009; Eisend & Tarrahi, 2021; Ham & Nelson, 2019). This naturally occurs through daily trial-and-error in interacting with persuasion agents, but one branch of the PK literature has focused on fostering this learning process by training consumers to better recognize and exercise vigilance towards specific persuasion tactics (e.g. illegitimate endorsers). This specific sort of training to be vigilant in detecting a particular tactic can indeed improve its detection (e.g., training consumers to detect illegitimate sources improves detection of illegitimate sources; Sagarin et al., 2002). However, the research reported here theorizes and further shows that specific training interventions of this kind can also have an

✉ Andrew E. Wilson
awilson5@suffolk.edu

Peter R. Darke
pdarke@schulich.yorku.ca

Jaideep Sengupta
mkjaisen@ust.hk

¹ Sawyer Business School, Suffolk University, 8 Ashburton Place, Boston, MA 02108, USA

² Schulich School of Business, York University, 4700 Keele St, Toronto, ON M3J 1P3, Canada

³ Department of Marketing, HKUST, Clear Water Bay, Jiulong, Hong Kong

unfortunate side-effect, in terms of making consumers more vulnerable to other tactics they would otherwise recognize (e.g., inconspicuous limiting footnotes). We identify this side-effect by elaborating on the existing *Persuasion Knowledge Model* (Campbell & Kirmani, 2008; Friestad & Wright, 1994) to include the idea that consumers typically have numerous vigilance goals that operate at different levels in a goal hierarchy. In addition to identifying the side-effects of specific tactic training, we use this same theorizing to show that more general forms of training that highlight broader, higher level vigilance goals are more effective in helping consumers detect a range of unethical persuasion tactics.

The purpose of this paper is to build and test a model that integrates key ideas from the Persuasion Knowledge Model (PKM; Friestad & Wright, 1994) and Goal Systems Theory (Fernandez & Kruglanski, 2019) in order to better understand the range of effects that training consumers to detect misleading persuasion tactics can have. Theoretically speaking, we update the PK model by integrating ideas from Goal Systems Theory (Fernandez & Kruglanski, 2019) to propose a multi-goal, multi-level PK framework involving different types of vigilance goals. This framework leads us to make new distinctions between specific and general training procedures that encourage vigilance at different levels in the goal hierarchy, as well as novel empirical findings in terms of the negative side-effects caused by specific training and the broader detection benefits of general training. Our theorizing and findings fill an important gap in the PK literature by helping understand how multiple vigilance goals operate together in the detection of unethical tactics. These findings also have important practical implications given that many consumer welfare organizations currently use specific forms of training to help consumers cope with misleading claims, whereas our research suggests general training can offer broader benefits in detecting such claims. Our theorizing and specific predictions are developed below, followed by two experiments that test these predictions.

Research Background

Persuasion Knowledge Model

The Persuasion Knowledge Model (Campbell & Kirmani, 2008; Friestad & Wright, 1994) posits that consumers typically learn about persuasion tactics through the course of their daily experiences (Boush, et al., 1994), and are motivated to use this knowledge to form valid judgments concerning persuasive messages and agents. This is partly accomplished by the development of simple tactic recognition heuristics, which may invoke a “change-of-meaning” for the corresponding persuasive message. That is, once a consumer recognizes a persuasion tactic is being used

(e.g., flattery from a sales clerk), the meaning of the message becomes informed by the relevant persuasion knowledge (e.g., sales agents may flatter because they want to make a sale). While PK can be positive in some cases (e.g., expert sources are reliable; Isaac & Grayson, 2017), the empirical literature has been more focused on negative PK concerning consumers’ skeptical reactions towards misleading or manipulative tactics (e.g., price gouging; Ferguson et al., 2011).

Existing research on the basic PK model has focused largely on identifying cognitive and motivational variables that play a role in detection and response to the use of persuasion tactics. One major area of investigation in the PK literature concerns the salience of cues that suggest a persuasion agent has an ulterior motive. In a now classic study, Campbell and Kirmani (2000) showed that the increased salience of flattery as a persuasion tactic, along with adequate cognitive capacity to engage in careful information processing, evoked PK relating to ulterior sales motives, and led consumers to become suspicious of the sales agent (see also Chan & Sengupta, 2010; Grillo and Pizzutti, 2021; Main et al., 2007, 2019).

Another body of related research has identified a number of broad motivational factors that can improve the detection of persuasion tactics. Advertising skepticism is perhaps the main variable examined in this respect, which is defined as a general tendency to doubt advertising claims, and is measured as an individual difference variable (Obermiller & Spangenberg, 1998). Ad skepticism predicts a range of negative reactions towards advertising, including more negative attitudes towards ads, lower belief in ad claims and persuasiveness, and greater counterarguing of claims (Obermiller et al., 2005; see also Charlton & Cornwell, 2019). Especially relevant to the current concerns, skeptical consumers are also more likely to identify advertising claims as false or misleading (Obermiller et al., 2005). Numerous other studies have also shown that ad skepticism can moderate the otherwise persuasive impact of a broad range of tactics, including: covert product placements (Chan, 2020), sustainability claims (Cho & Taylor, 2020), native advertising (Chung and Kim, 2020), social cause advertising (Yang & Mundel, 2021), manipulative attention getting tactics (Germelmann et al., 2020), and brand extension appeals (Hernandez et al., 2019). Finally, a number of related motivational variables have also been shown to increase the use of PK and buffer against persuasion attempts, including: accuracy motivation (Johar & Simmons, 2000), loss prevention goals (Kirmani & Zhu, 2007), and feelings of vulnerability (Chang, 2017). For instance, Johar and Simmons (2000) found that increased accuracy motivation led to higher levels of detection for a footnote that limited an advertised product warrantee.

Training Vigilance Towards Specific Persuasion Tactics

A good deal of research concerning the PK model has examined how to improve consumers' detection of specific manipulative or misleading persuasion tactics (Table 1). All of these studies essentially alert consumers to the use of a specific misleading tactic, or make that tactic more salient, and then measure the effects this training has on the detection of that same tactic in a subsequent persuasion attempt. We refer to this as specific training, and describe the relevant studies in Table 1.

Numerous researchers have used the idea that increasing the salience of a particular tactic and/or awareness of its persuasive/manipulative intent can alert consumers to the subsequent use of the particular tactic. For instance, Kardes et al. (2006) designed an intervention to increase sensitivity to omission neglect (i.e., lack of due attention to missing information). Having consumers simply rank attributes by importance effectively increased sensitivity to missing information. Zarouali et al. (2020) used social norm information to increase the salience of sponsored content on social media, and found this increased perceptions of persuasive intent and led consumers to adjust their online account settings to restrict future advertising. A recent meta-analysis of 57 datasets concerning such disclosures of sponsored content found that making this tactic more salient increased ad recognition, perceptions of persuasive/manipulative intent, and resistance to persuasion; as well as reducing brand and source attitudes, and also source credibility (Eisend et al., 2014).

Other studies that examine specific training effects have developed somewhat more involved training procedures that educate consumers about the use of a particular persuasion tactic, and/or offer additional instructions as to the illegitimacy of the tactic. These studies are also reported in Table 1. For instance, Harris (1977) trained consumers to differentiate between stated and implied claims, which led to greater disbelief of the implied advertising claims they later viewed. Germelmann et al. (2020) examined the use of an incongruence tactic that placed product ads in media vehicles that did not match the product category (e.g., placing a computer printer ad in an automotive magazine, rather than a computing magazine, in order to attract more attention), and found that although consumers spontaneously expressed skepticism towards this tactic, these reactions intensified when additional training concerning the manipulative intent of the tactic was provided. Tessitore and Geuens (2019) compared factual training (i.e., based on scientific evidence), evaluative training (i.e., indicating inappropriateness), and no training conditions concerning the use of product placement tactics, and found that any training increased perceptions of manipulative intent and lowered purchase intentions relative to no

training. Most central to the specific training examined in the current studies, Sagarin et al. (2002) trained consumers to better distinguish between advertising endorsers who were actors rather than bona fide experts using a brief procedure in which participants examined ads with either legitimate or illegitimate spokespeople and determined whether each: (1) was an actual authority rather than an actor or impostor, and (2) possessed expertise that included the endorsed product category. This form of specific training proved effective in improving the detection of illegitimate endorsers in subsequent advertisements.

Finally, we note that the specific training approach is also very common in the educational campaigns of regulators such as the FTC (e.g., <https://www.consumer.ftc.gov/features/pass-it-on/youve-won-scams> and <https://www.consumer.ftc.gov/features/pass-it-on/home-improvement-scams/>); and consumer advocacy groups such as *Consumer Reports* (e.g., <https://www.consumerreports.org/car-repair/get-an-extended-warranty-for-your-car/> and <https://www.consumerreports.org/dietary-supplements/beware-dietary-supplements-marketed-online/>) and *Truth in Advertising* (e.g., <https://www.truthinadvertising.org/is-that-ad-deceptive/>). In each example, the strategy is to inform consumers about a particular type of deceptive tactic (e.g., you've won offers, pay-me-now scams, overpriced extended warranties, bogus online dietary supplements, or unsupported health claims) and urge vigilance towards that particular tactic in the future.

Overall, this literature suggests training consumers to detect specific tactics can be very helpful in improving vigilance towards the particular tactic involved. However, we are not aware of any prior research that considers what effects specific training can have on the detection of other, nonfocal tactics (i.e., tactics that are not the direct concern of the training involved). The current research fills this gap by theorizing and showing that specific training can, in fact, have negative side effects on the detection of other PK tactics that are not the focus of the training, and further suggests more general training that encourages vigilance at a broader level can lead to more effective coping. We develop these predictions in the next section of this paper by proposing a multigoal, multilevel extension to the traditional PK model that can account for the effects of multiple vigilance goals.

Persuasion Knowledge, Goal Systems, and Training

As the studies in Table 1 suggest, there are a large number of different misleading tactics that consumers can face in the realm of advertising, implying that multiple vigilance goals concerning these tactics must operate together in order to effectively cope with persuasion attempts. Moreover, other aspects of the PK literature suggest that broader vigilance motives such as ad skepticism and accuracy motivation can also aid in the detection of persuasion tactics. However, the

Table 1 Literature review of specific training studies and their effects by focal tactic and contribution of current research

| Representative paper | Focal tactic | Relevant IV(s), moderator(s), and/or mediators | Relevant DV(s) | Key finding(s) |
|--------------------------------------|-----------------------|--|--|---|
| Harris (1977) | Implied claims | Instructions for implied vs. asserted claims Time delay between training and response | Discrimination of implied vs. asserted claims | Instructions led to improved recognition of implied vs. asserted claims, but not after a delay |
| Sagarin et al. (2002) | Illegitimate endorser | Training treatment Endorser legitimacy Vulnerability to persuasive intent | Perceived persuasiveness Manipulative intent | Training leads to some resistance, strengthened by vulnerability |
| Kardes et al. (2006) | Omission neglect | Sensitivity to omission | Perceived information sufficiency A_{product} | Training consumer to consider criteria prior to ad exposure leads to increased sensitivity and reduced product evaluation |
| Wei et al. (2008) | Covert sponsorship | PK activation (sponsorship disclosure) Brand familiarity Perceived appropriateness | A_{brand} | PK activation led to lower brand evaluations, especially when perceived appropriateness and brand familiarity were low |
| Wentzel and Tomczak (2010) | Narrative processing | Narrative vs. expository ad execution Salience of manipulative intent | A_{ad} | Salience of persuasive intent nullifies the evaluative advantage that narrative ads otherwise hold over expository ads |
| De Kerpel et al. (2019) | Apology-appeal | Apology appeal Disclosure | A_{firm} Purchase intention (PI) | Apology appeals are persuasive in terms of enhancing A_{firm} and PI, relative to a control appeal. Disclosure of apology as a persuasive appeal extinguishes this advantage |
| Tessitore and Geuens (2019) | Product placement | Factual vs. evaluative vs. no intervention | Persuasive intent PI | Evaluative > factual > no intervention in terms of recognition of persuasive intent and resistance in terms of PI |
| Boerman et al. (2020) | Paid placement | Disclosure of paid placement Delay | Persuasive intent A_{brand} PI | Disclosure of paid brand placement increased recognition of ad, immediately and after delay, but no effect on A_{brand} or PI |
| Eisend et al. (2020) (meta-analysis) | Covert sponsorship | Disclosure (of sponsored content) | Source credibility Ad recognition Persuasive/manipulative intent Resistance A_{brand} A_{source} | Across 57 datasets, disclosure of sponsored content reduced A_{brand} A_{source} , source credibility; and increased ad recognition, persuasive/manipulative intent, and resistance |

Table 1 (continued)

| Representative paper | Focal tactic | Relevant IV(s), moderator(s), and/or mediators | Relevant DV(s) | Key finding(s) |
|--------------------------|--|---|--|---|
| Germelmann et al. (2020) | Product-medium incongruence | Congruence vs. incongruence of product-medium PK tactic training | Spontaneous thoughts about tactic Acceptability of tactic $A_{product}$, A_{brand} , A_{ad} PI | Consumers spontaneously detected and expressed skepticism towards the incongruence tactic. These negative reactions were intensified when additional instruction concerning the manipulative intent of the tactic was provided |
| Xie et al. (2020) | Salesperson dishonesty | Perspective taking training Moral self-awareness | Ethical tolerance | When consumers take the salesperson's perspective, and their own moral self-awareness is salient, they become less tolerant of salesperson lying |
| Zarouali et al. (2020) | Sponsored social media content | Social norm information vs standard sponsored label | Persuasive intent Restrictive ad setting preferences | Social norm information was more effective than a standard "sponsored content" label in indicating persuasive intent and led to stricter ad settings |
| Hwang and Jeong (2021) | Native advertising | Format similarity Format novelty | Perceived deceptiveness Ad recognition Perceived irritation | Format similarity (ad cf. content) led to perceived deceptiveness. Format novelty led to reduced advertising recognition and perceived irritation |
| Current Research | Illegitimate endorsers Restrictive footnote | Specific vs. general vs. no vigilance training | Footnote detection (nonfocal tactic) Endorser detection (focal tactic) Manipulative intent PI | Specific training to detect a particular tactic, while effective in terms of its intended effect, also has the negative side effect of making consumers more vulnerable to other nonfocal tactics. More general training that activates higher level vigilance goals leads to the detection of a broader set of tactics |

existing PK model does not speak to the question of how different vigilance goals might be structured and operate together. In order to address this issue, and provide the logic for our own predictions concerning the effects of specific and general training, it was necessary to further develop the traditional PK model by incorporating ideas from Goal Systems Theory (Fernandez & Kruglanski, 2019; Kopetz et al., 2012).

Goal Systems Theory (Fernandez & Kruglanski, 2019) posits that goals are knowledge structures governed by the same principles that characterize other cognitive structures. Specifically, goals have the property of interconnectedness, which occurs in a theoretical two-dimensional space. Related goals are vertically arrayed along a dimension of abstractness/generalizability, such that the broadest most abstract goal is situated at the top of the goal system, whereas the bottom of the system includes the most concrete/specific subgoals (Fig. 1). Connections between the goals along the vertical dimension tend to be excitatory, in the sense that activation of a higher-level goal spreads down to further activate all the various subgoals that serve it. Figure 1 depicts the proposed goal system for the vigilance goals involved in the detection of various persuasion tactics. The PK model suggests the consumer's overarching goal is the formation of valid or accurate judgments (Friestad & Wright, 1994), which is shown at the top of the figure. This accuracy goal is likely linked to multiple subordinate goals, such as the desire to: seek information (Darke et al., 1995); understand, evaluate and

integrate information (Cacioppo & Petty, 1982); and be skeptical or vigilant of marketing claims (Kirmani & Zhu, 2007; Obermiller & Spangenberg, 1998), among others. These sub-goals, in turn, should be linked to the even more specific subordinate goals that serve them. In particular, general vigilance should be linked to subordinate goals concerning the identification of specific tactics, such as: the manipulative use of flattery (Campbell & Kirmani, 2000), native ads (Hwang and Jeong, 2019), illegitimate expert endorsements (Sagarin et al., 2002), inconspicuous claim qualifications (Herbst et al., 2012), etc. Figure 1 depicts these types of subgoals at the bottom of the goal system hierarchy.

Goal System Theory further suggests that goals are arranged on a second (lateral) dimension that allows for several alternative subgoals that are associated with the same superordinate (more general) goal to be arranged at the same level of specificity/generalizability. Lateral connections have the property that activation of one goal suppresses (or inhibits) the activation of connected alternative goals at the same level in the hierarchy (Kopetz et al., 2012; Shah et al., 2002). For instance, Fig. 1 shows that the subgoals related to vigilance towards the specific tactics of manipulative flattery, native advertising, illegitimate endorsers, and inconspicuous qualifications, are all at the same level in the hierarchy, and are negatively connected to each other. This means that activation of one tactical vigilance goal (e.g., watch for illegitimate endorsers) should inhibit the activation of the other tactical vigilance goals at the same level in the hierarchy (i.e., vigilance for qualifying claims, native advertising, and flattery). This deactivation of other subgoals at the same level is known as *goal shielding*, which is a process that is usually functional in that it emphasizes the focal subgoal over distracting alternatives, thereby enabling more dedicated goal regulation (Kopetz et al., 2012; Shah et al., 2002). Finally, if *goal fulfillment* subsequently occurs, for instance when the vigilance task is successfully completed by identifying the use of the particular tactic, this tends to have the effect of inhibiting both the focal subgoal and related goals (Förster et al., 2005). If not fulfilled, the focal goal can remain activated for relatively long periods of time (Förster et al., 2007).

The proposed goal system model for PK has a number of implications for specific training interventions that involve detection of a specific tactic. First, according to goal shielding, when consumers are trained to identify a particular tactic (e.g., an illegitimate expert) and subsequently detect it, other specific detection goals that are laterally linked should be deactivated (e.g., a limiting footnote), thereby lowering detection of the latter tactic. On this basis, we propose:

H₁: A specific training intervention (vs. no training) that teaches consumers to successfully detect a par-

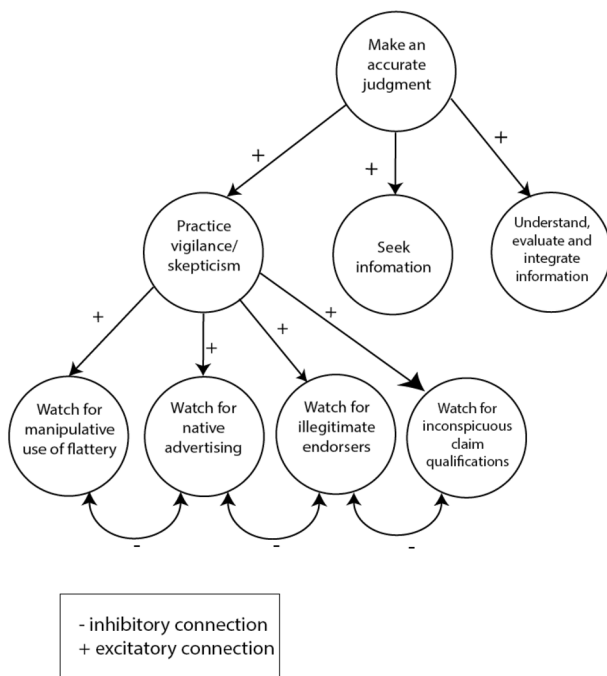


Fig. 1 The proposed goal system fragment

ticular advertising tactic will reduce detection of a second tactic in the same ad. (goal shielding side-effect prediction)

However, and importantly, our goal systems model also suggests a potential strategy for avoiding the problematic goal shielding side-effects. That is, a more general form of training that activates the higher-level vigilance goal (e.g., general advertising skepticism) should excite the vertically connected set of subordinate tactical goals, and thereby increase identification of a broader range of specific tactics (e.g., both illegitimate sources and inconspicuous qualifications should be better detected). If so, we propose:

H₂: A general training intervention that teaches consumers that vigilance and a critical mindset are effective tools for detecting misleading advertising tactics will lead to better detection of multiple tactics used in the same ad, relative to untrained controls.

Finally, our theorizing suggests that goal shielding should occur for specific training but not for general training. As mentioned above, specific training should increase detection of the illegitimate endorser tactic, and in turn satisfaction of this focal subgoal should inhibit the detection of any other subgoals that serve the same higher level vigilance goal (i.e., should lower footnote detection). This implies that goal shielding would be demonstrated by a specific-training + source-detection - footnote-detection path, where the link between specific-training and source detection is positive (i.e., specific training should increase detection of the illegitimate source) and the subsequent link between source-detection and footnote detection is negative (i.e., detection of the illegitimate source should subsequently suppress detection of the limiting footnote). In contrast to specific training, general training should not evoke goal shielding given it activates the higher-level vigilance goal, which should generally increase detection of both lower-level tactical goals (i.e., increased detection of both illegitimate source and footnote tactics). This led to the following mediational predictions:

H₃: The benefits of specific training in better detecting the focal (illegitimate endorser) tactic should lower identification of the nonfocal (footnote) tactic, implying a significant specific-training + source-detection - footnote-detection path. In contrast, the specific-training + source-detection - footnote-detection path should not hold in the case of general training.

Summary of Experiments

Two experiments tested the predicted effects of specific and general training interventions on the detection of misleading advertising tactics. In particular, Exp 1 manipulated specific

training (vs. no training), as well as the legitimacy of the endorser (legitimate vs. illegitimate) shown in a test ad after the training. This test ad always included a limiting footnote as well. Detection of the footnote served as the main dependent measure. The goal shielding prediction here was that specific training would lead to the successful detection of the illegitimate endorser but ironically lower detection of the footnote (H₁). Exp 2 then replicated this ironic goal shielding effect (H₁), and further tested whether general training would improve detection of both the endorser and footnote tactics, as predicted by H₂. This study also tested the goal shielding mediational prediction (H₃). The results of both studies were generally consistent with predictions.

Exp 1: Ironic Effects of Specific Training on a Non-focal Tactic

Method

Participants and Design

Ninety undergraduate students were randomly assigned to a 2 (Training: specific vs. control) × 2 (Endorser: legitimate vs. illegitimate) between-participants design. The sample was 61.1% female with an average age of 20.74 years, and consisted of all available participants in the departmental research pool. The specific training intervention, which was adapted from Sagarin et al. (2002), instructed participants to be vigilant to illegitimate endorsers. A footnote that meaningfully qualified the advertised offer served as the non-focal tactic, and was included in the ad for all conditions. Such qualifications are considered misleading when they are less conspicuous than the more prominent offer they contradict (Fair, 2014), and consumers readily identify such footnote tactics as misleading (Herbst et al., 2012; Johar & Simmons, 2000). Detection of this qualifying footnote served as the main dependent measure.

Procedure

On arrival, participants in the training condition were given a booklet containing detailed instructions adapted from Sagarin et al., (2002; Supplementary Material Appendix A). This explained that while advertisements often use spokespeople to endorse products, the use of illegitimate endorsers has been deemed unethical by regulators, and that an endorser must pass two tests to be legitimate: (1) they must be a real authority, and not just someone dressed to look like one; and (2) they must be an expert on the product being sold. Control participants completed an unrelated filler task instead.

Finally, all participants were given a separate booklet to assess detection of the illegitimate endorser. Ace Hardware was used as the retailer because it did not exist in the market where the study was completed. Importantly, the test ad included a photograph and product endorsement (“This one has all the tools you need for fixing things up around the house. These high-quality tools will last for years.”), where the endorser was manipulated according to condition (Supplementary Material Appendix B). In the *illegitimate condition*, the endorser was Tim Allen, a famous actor/comedian who starred in the program *Home Improvement*. Bob Vila, a well-known home improvement expert, served as the *legitimate endorser*. The programs of each endorser aired in the local market, and a pretest confirmed Allen was viewed as less expert than Vila. In addition, a qualifying footnote was included in the test ad to serve as the non-focal tactic. The test ad included a savings of 50% (“You pay only \$34.99”) in large type at the top; but a substantial qualification at the bottom stated in small type that the offer was available only to Ace Club members, whereas nonmembers would pay the full price of \$69.99. As all participants were by definition nonmembers, the subtle footnote meant they would not receive the discount.

After evaluating the ad, the main dependent measure was assessed through two items that tapped whether the qualifying footnote was processed. Participants first identified which of the following responses was true: (a) no special price offer was presented, (b) the special price offer was limited to new customers, (c)...limited to club members (correct), (d)...limited to the first 50 customers, or (e)...not limited; and then what price they would pay: \$99.99, \$69.99 (correct), or \$34.99. Total correct answers were summed to create the main dependent measure (range = 0–2; see Bendig, 1954).

While this study was focused on the ironic effects that endorser training had on detection of the non-focal footnote tactic, a secondary dependent variable was also included to enable a test of whether the training led to better detection of the focal endorser tactic, as intended. For this purpose, we simply asked participants to assume they were only interested in buying products from ethical companies, and indicate how likely they were to purchase the Ace toolkit; rated from 1 (extremely unlikely) to 7 (extremely likely). This served as a measure of ethical purchase intentions. Note that purchase intentions are typically measured using a single item scale (Morwitz and Munz, 2021), and such measures have been shown to have similar validity to multi-item scales when testing specific constructs (Bergkvist & Rosser, 2007). Finally, a check for the endorser manipulation asked participants to identify the spokesperson in the test ad from a list of four names that included both Tim Allen and Bob Vila.

Results and Discussion

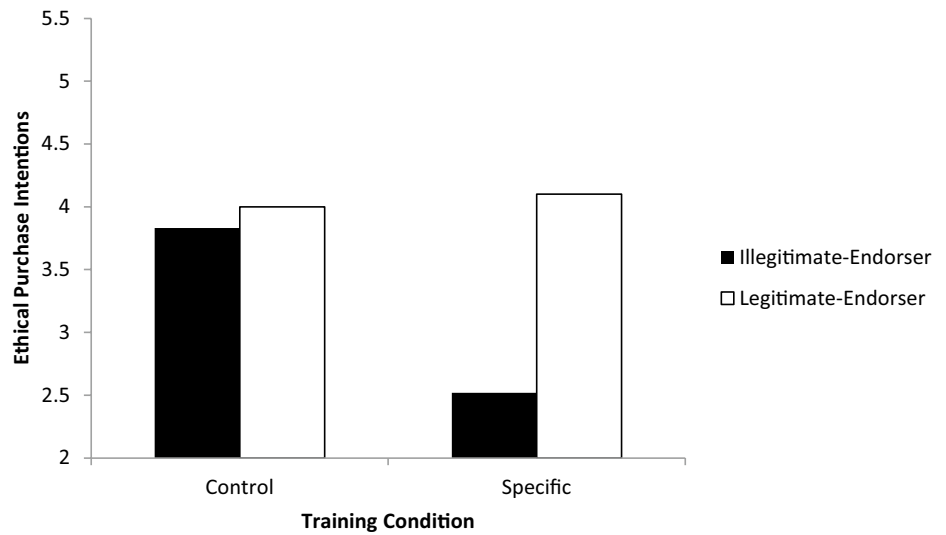
A cross-tab analysis of endorser condition by endorser identified in the manipulation check was conducted to determine whether participants were able to correctly identify the endorser used in the test ad for their particular experimental condition, which revealed they were able to do so ($X^2(4) = 86.09, p < 0.001$). All participants in the illegitimate condition correctly identified Tim Allen as the spokesperson, whereas 88.89% of those in the legitimate endorser condition correctly identified Bob Vila. Only those who correctly identified the appropriate endorser were further analyzed for purposes of internal validity ($N = 85$). Retaining these participants does not qualitatively alter the main conclusions of this study.

A 2×2 ANOVA on ethical purchase intentions showed main effects of both training ($F(1, 81) = 3.43, p = 0.07$) and endorser ($F(1, 81) = 7.13, p < 0.01$), qualified by a training \times endorser interaction ($F(1, 81) = 4.66, p < 0.05$). Within the specific training condition, ethical purchase intentions were lower for the illegitimate versus legitimate endorser ($M_s = 2.52$ vs. 4.10, $p = 0.001$), whereas no such difference was observed for the no-training condition ($M_s = 3.83$ vs. 4.00, $F < 1$, NS; see Fig. 2). Within illegitimate endorser conditions, trained participants were also less likely to purchase from Ace than their untrained counterparts ($M_s = 2.52$ vs. 3.83, $p = 0.005$), but not in the legitimate endorser conditions ($M_s = 4.10$ vs. 4.00, $F < 1$). Thus, specific training improved the ability to detect the illegitimate tactic as intended. Overall, these findings replicate the basic training effects observed by Sagarin et al. (2002), in this case using our own specific training procedure and test ad; and further verify that specific training successfully increased activation of the focal vigilance goal for illegitimate sources and their detection.

The next set of analyses tested the predicted ironic effects of endorser training on detection of the non-focal footnote tactic. According to the goal shielding hypothesis (H_1), we expected those in the training/illegitimate-endorser condition to be the least accurate in detecting the footnote tactic, because the activation and fulfillment of the focal illegitimate endorser goal should inhibit other lower-level detection goals, in this case the nonfocal goal of footnote detection (Förster et al., 2005; Shah et al., 2003). The ANOVA for detection of the non-focal footnote qualification showed a training main effect ($F(1, 81) = 3.16, p = 0.08$), and the predicted training \times endorser interaction ($F(1, 81) = 3.16, p = 0.08$). Within the illegitimate endorser condition, endorser training led to lower detection of the nonfocal footnote tactic compared to no-training ($M_s = 0.81$ vs. 1.38, $p = 0.01$); whereas endorser training had no effect on footnote detection in the legitimate endorser condition ($M_s = 1.20$ vs. 1.20; $F < 1$; Table 2 for a summary of hypothesis testing).

Fig. 2 Specific endorser training led to increased detection for illegitimate endorser tactic (i.e., decreased purchase intentions), but also lowered detection of the footnote tactic (Exp 1)

(A) Intended effect of training: enhanced detection of focal illegitimate endorser tactic



(B) Unintended side effect of training: lower detection of non-focal footnote tactic

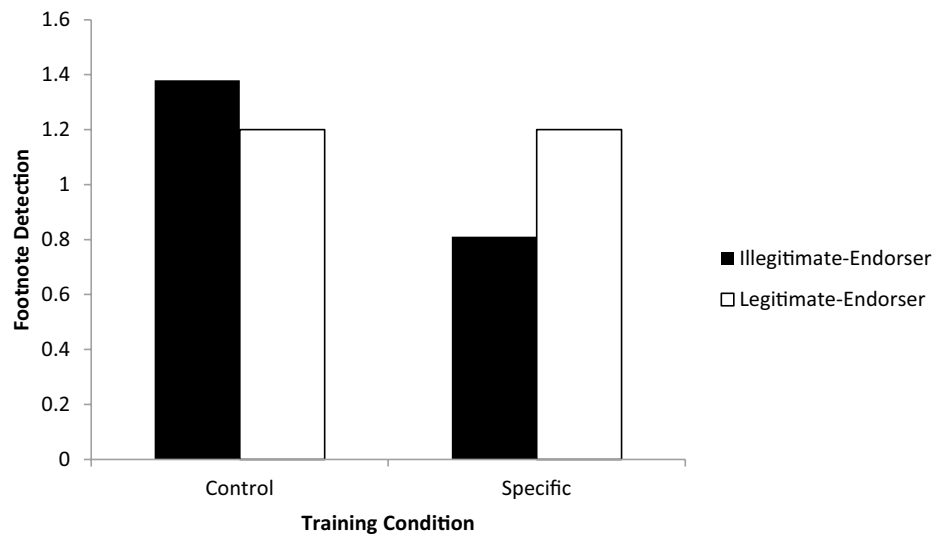


Table 2 Summary of hypothesis testing

| Hypothesis | Exp | Analysis | Key result |
|----------------|-------|--|---|
| H ₁ | Exp 1 | ANOVA of footnote detection by training and endorser legitimacy, with post-hoc comparisons | Within illegitimate endorser conditions, trained participants had lower footnote detection than untrained controls ($M_s = .81$ vs. 1.38 , $p = .01$) |
| H ₁ | Exp 2 | ANOVA of footnote detection by training condition | Specific training ($M = .07$) decreased footnote detection, relatively to both control ($M = .24$, $p = .05$) and general training ($M = .42$, $p < .001$) |
| H ₂ | Exp 2 | ANOVA of source ethicality by training condition | General training led to lower perceived ethicality of the source ($M_s = 4.73$ vs. 5.27 , $p = .07$) |
| H ₃ | Exp 2 | PROCESS Macro regression analysis, where Y = footnote detection, X = training condition, and M = endorser ethicality | Indirect effect was significant for the specific training ($\beta = .047$, $CI_{95} = .0436$ to $.2286$), but not for general training ($\beta = .047$, $CI_{95} = -.0019$ to $.1122$). Path coefficients indicated specific-training increased identification of the endorser as unethical ($a = -1.51$, $p < .0001$), which in turn lowered footnote detection ($b = -.082$, $p = .0001$) |

In summary, Exp 1 provided evidence for the effectiveness of specifically training consumers to detect illegitimate endorsers. However, this same endorser training also suppressed detection of the non-focal footnote tactic when the test ad included an illegitimate endorser. Though ironic, this suppression effect is consistent with our goal systems model for the activation of PK (H_1), and more specifically the idea that activating and fulfilling the focal tactical goal should suppress other tactical goals serving the same higher-level vigilance goal.

Exp 2: Specific Versus General Training

Our second study further tested our goal systems approach for the activation of PK by using a theory-based alternative intervention aimed at helping recognize a broader set of misleading tactics. Remember that the same goal system model also predicts that activation of the higher-level vigilance goal itself should facilitate all of the tactical subgoals that serve it. Exp 2 added a more general form of training aimed at activating the broader vigilance goal to test whether this would simultaneously activate both the subgoal of guarding against illegitimate endorsers and the subgoal of detecting restrictive footnotes, thereby improving the detection of both tactics (H_2).

This study again included a specific endorser training condition to replicate its ironic effects on footnote detection in Exp 1, as well as for comparison purposes with the new general training procedure. This experiment was also designed to improve on the methodology used in the first study in a number of ways. In particular, the sample used here was larger and broader, in order to better establish the validity of the findings. The generalizability of the results was also tested by changing a number of aspects of the specific training procedure and the illegitimate endorser used in the test ad. Finally, we included additional measures that more directly captured detection of the illegitimate source, as well as a standardized measure of manipulative intent (Campbell, 1995) that was expected to be sensitive to the detection of both the illegitimate endorser and the limiting footnote tactic.

Method

Participants and Design

Participants were randomly assigned to either a General-Training, Specific-Training, or No-Training condition in a between-participants design. The test ad always included both the illegitimate endorser and the restrictive footnote. The legitimate endorser condition was not included in this study. The main dependent measures were detection of the

restrictive footnote tactic and perceived ethicality of the endorser tactic. The sample of respondents ($N=297$) was recruited through the CloudResearch Panel (Litman et al., 2016). The average age of the sample was 44.43 years (range = 18–89) and 61.60% were female. The highest level of education attained by participants was: less than high school (3.70%), high school (25.25%), some college or 2 year degree (27.95%), 4 year degree (23.57%), professional/doctorate degree (19.19%), or unreported (0.34%). All participants were located in the United States.

Procedure

The procedure was similar to Exp 1, with the changes noted here. A *general training condition* was added to the design using instructions that encouraged vigilance towards advertising in general (Supplementary Material Appendix C). These instructions informed participants that advertisers have been increasingly accused of certain unethical practices in their ads, and that these tactics rely on the fact consumers do not typically examine advertising with a careful and critical mindset. Participants were also told research had shown that examining ads more carefully and rigorously was usually all that was required to identifying the misleading tactics advertisers typically use and make more accurate judgments. Importantly, these instructions did not specifically refer to the illegitimate use of endorsers or qualifying footnotes.

The *specific (endorser) training condition* was updated and simplified somewhat to make it more suitable for completion online (Supplementary Material Appendix D). This specific training procedure again explained that regulators have deemed the use of illegitimate endorsers to be unethical, and described the same two tests used to determine the legitimacy of an endorser. Participants were then exposed to four contemporary training ads, which were used to illustrate the legitimacy tests. In the *no training condition*, participants again completed an unrelated task, which in this case was a short article concerning technological innovations involving consumer products. After reading each training advertisement, participants in the training conditions responded to a brief check to determine if they had carefully read the training materials. Any participants who answered this check incorrectly were required to complete the relevant training phase a second time. Control participants similarly responded to check items to ensure they had understood the unrelated article.

After completing the training phase, all participants were shown the test ad. This was similar to the test ad in Exp 1, with a number of changes (Supplementary Material Appendix E). First, the retail brand was changed to Home Hardware, which is a Canadian brand that does not exist in the US retail space. Therefore, the members of our US based sample were not club members by definition and would have

to pay the full price according to the limiting footnote. In addition, we changed the illegitimate endorser in the test ad to *Nick Offerman*, who is an actor who played a character known for woodworking in his spare time in the recent *Parks and Recreation* comedy series. Otherwise, the test ad was the same as before. Of note, the test ad always included an illegitimate endorser and a limiting footnote tactic in this study. Remember that it was only the illegitimate endorser condition that was impacted by specific training in our previous study.

After examining the ad, the main dependent measure for footnote detection was assessed. In this case, participants completed a standard thought listing task where they were given 2 min to list up to five thoughts or feelings they had about the ad (Petty & Cacioppo, 1986). Each thought/feeling was listed in a separate box and later coded as either not mentioning the footnote (0) or mentioning the footnote (1) by two coders who were blind to condition. Coder agreement was 94.3% ($r=0.88$, $p<0.01$). Disagreements were settled through discussion. Examples of responses coded as mentioning the footnote included: "Fine print you have to be a member" and "Appears to be a good deal but you have to be a member to get it." These coded responses were summed for each participant to create a footnote detection index (ranging from 0 to 5). In comparison to the aided recall measure used to assess footnote detection in Exp 1, the open-ended thought listing technique used here assesses spontaneous recognition, comprehension, and elaboration of the limiting footnote (Petty & Cacioppo, 1986), and has been used in previous studies examining footnote detection (Johar & Simmons, 2000).

A measure of endorser ethicality was also included to assess the extent to which participants were specifically able to detect the use of an illegitimate spokesperson. This included two items where participants indicated whether they thought the endorsement of the toolkit by the spokesperson in the ad was unfair/fair and unethical/ethical, using ratings scales ranging from 1 to 7 ($\alpha=0.87$). This measure again essentially served as a check for the specific endorser training condition, but also helped determine whether general training would improve detection of the illegitimate source. Finally, participants completed a standardized measure concerning the inference of manipulative intent of the tactics used in the ad (Campbell, 1995). This included five items (e.g., "The way this ad tries to persuade people seems acceptable to me" and "The advertiser tried to manipulate the audience in ways that I don't like") rated from strongly disagree (1) to strongly agree (7). Items were combined into a manipulative intent index where higher scores indicated the ad's tactics were more manipulative ($\alpha=0.77$). We expected that both specific and general training would lead to perceptions the ad used manipulative tactics relative to no training.

Results and Discussion

An initial examination of whether participants recognized that the endorser used in the test ad was Nick Offerman rather than three other options showed that 72.3% correctly identified him. Once again only those who correctly identified the endorser were further analyzed. The results did not qualitatively differ if these participants were retained.

The ANOVA for the source ethicality measure was significant ($F(2, 210)=14.91$, $p<0.001$). Specific training ($M=3.76$) lowered perceptions of source ethicality relative to both no-training ($M=5.27$, $p<0.001$) and general training conditions ($M=4.73$, $p=0.001$). General training also lowered perceptions of source ethicality relative to no training ($M_s=4.73$ vs. 5.27 , $p=0.07$). These findings confirmed that specifically training participants to detect an illegitimate endorser had the obvious effect of improving this form of detection (i.e., of the focal tactic). The results further suggest that general training also aided the detection of an illegitimate endorser, consistent with H_2 (Fig. 3A).

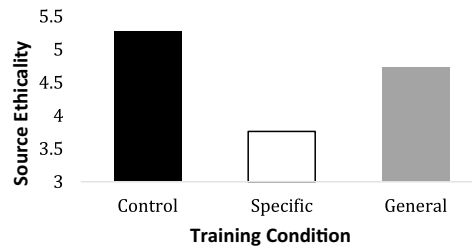
The ANOVA for the footnote detection measure was also significant ($F(2, 211)=7.33$, $p=0.001$). As expected, specific training ($M=0.07$) again ironically decreased detection of the limiting footnote (i.e., the nonfocal tactic) relative to both no training ($M=0.24$, $p=0.05$) and general training conditions ($M=0.42$, $p<0.001$). In contrast, general training improved footnote detection relative to no training ($M_s=0.42$ vs. 0.24 , $p=0.05$). These results replicate those of Exp 1 by showing that specific training again ironically lowered detection of the (nonfocal) footnote tactic (supporting H_1). In contrast, general training had the benefit of increasing footnote detection relative to both the no training and specific training conditions (supporting H_2 ; Fig. 3B).

The ANOVA for the inferred manipulative intent index was also significant ($F(2, 211)=6.91$, $p<0.001$). As expected, both the specific ($M=3.77$) and general ($M=3.73$) training conditions led participants to view the ad as more manipulative than the no training condition ($M=3.03$, $p_s<0.005$), whereas specific and general training conditions did not differ ($p=0.86$). This suggests that both specific and general training alerted consumers to ethical concerns about the ad (Fig. 3, panel C).

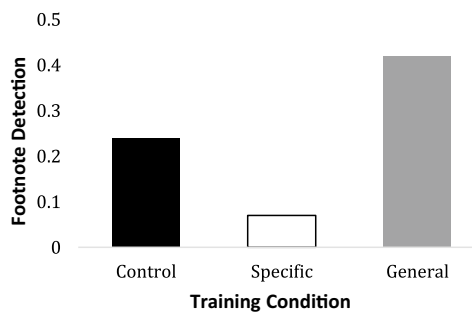
Our conceptual model suggests that the benefits of specific training in identifying the illegitimate endorser (focal tactic) should suppress identification of the footnote (nonfocal tactic) due to goal shielding, whereas general training should have no such effect. This implied endorser detection would mediate the effects of specific training (vs. no-training) on footnote detection but not for general training (H_3). We computed the implied mediational model using the PROCESS macro (Hayes, 2017), where Y =footnote detection index; X =training condition; and M =endorser ethicality (Model 4 with 5000 bootstraps). Training condition

Fig. 3 General training is effective in improving detection of both the illegitimate endorser and footnote tactics, while avoiding the side-effect (Exp 2)

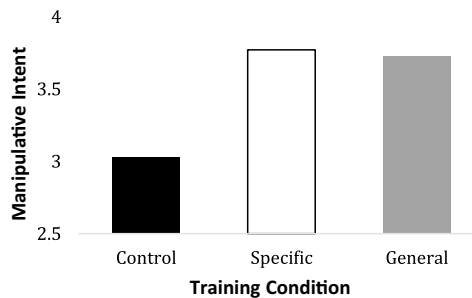
(A) Both specific and general training improve detection of the illegitimate endorser tactic



(B) General training improves footnote detection whereas specific training shows the negative side effect of lowering detection of this (nonfocal) tactic



(C) Both specific and general training cause consumers to view the ad as manipulative



was treated as multicategorical and coded to create a specific training dummy (specific = 1, 0 otherwise) and general training dummy (general = 1, 0 otherwise). As expected, the indirect (mediated) effect was significant for the specific training indicator ($\beta = 0.047$, $CI_{95} = 0.0436-0.2286$). Path coefficients indicated that specific-training increased identification of the endorser as unethical ($a = -1.51$, $p < 0.0001$), which in turn lowered footnote detection ($b = -0.082$, $p = 0.0001$). These mediational findings are therefore consistent with the predicted effects of goal shielding due to specific training. In contrast, identification of the endorser did not suppress detection of footnote in the case of general training ($\beta = 0.047$, $CI_{95} = -0.0019-0.1122$), suggesting that

goal shielding did not occur when general training was used. Overall, these mediational findings are consistent with H_3 .

Summary

The results replicate the Exp 1 finding that specific endorser training was effective in improving detection of the focal endorser tactic, and more importantly, again showed this had the ironic side-effect of lowering detection of the non-focal footnote tactic. The latter evidence further supported the main goal shielding hypothesis (H_1). In contrast, the general training procedure improved detection of both the

illegitimate endorser and limiting footnote tactics. This aspect of the results is consistent with the additional goal systems prediction that activating a higher-level goal should increase vigilance towards a broad range of lower-level tactical goals (H_2). Also, the mediational analyses confirmed the prediction that goal shielding should lead to lower detection of the nonfocal footnote tactic for specific training but not for general training (H_3). Finally, this study provided evidence that both specific and general training helped consumers recognize the target ad as manipulative using a standard PK measure (Campbell, 1995).

General Discussion

Our research shows that well-intentioned consumer interventions that successfully impart specific tactic detection skills can also have the ironic consequence of actually making consumers more vulnerable to other advertising tactics. While specific endorser training improved detection of the illegitimate endorser tactic (Exps 1 and 2), this also had the ironic and unintended consequence of suppressing the detection of a qualifying footnote in the same ad (Exps 1 and 2). This ironic effect was predicted by our goal systems model, in that the goal shielding mechanism suggests that focusing consumers on the goal of detecting a specific tactic (i.e., specific endorser training) and detecting that tactic (i.e., an illegitimate endorser) should suppress detection goals for other tactics (i.e., a limiting footnote) consistent with goal shielding (H_1). Exp 2 also supported the additional prediction that activating a broader vigilance goal using a more general training procedure would facilitate detection of both the illegitimate endorser and footnote tactic sub-goals (H_2). Overall, the results supported both aspects of the proposed goal systems view of PK concerning multiple detection goals.

Contributions

This research offers a number of theoretical contributions to the existing PK literature. First, whereas past studies have examined specific training procedures that essentially teach consumers to detect particular misleading tactics (Sagarin et al. 2002 in particular, but also see Table 1), our research generated the novel finding that this approach can have unfortunate side-effects that undermine the detection of other tactics consumers would otherwise detect. Second, we distinguish between the specific training procedures that have predominated in the existing literature (Table 1), and what we call more general training procedures, which alert consumers to broader concerns about misleading advertising, and urge greater critical thinking and general skepticism when confronted by advertising. To our knowledge, we are

the first to suggest this distinction and to empirically show that it provides important advantages over specific training by causing consumers to better detect a range of unethical advertising tactics, as well as avoid the problematic side effects of specific training. Finally, these findings were predicted on the basis of our theoretical elaboration of the role played by multiple detection goals in the Persuasion Knowledge Model (Boush et al., 2009; Friestad & Wright, 1994). Whereas the existing PK model does not consider the question of concurrent detection goals or the interplay between such goals, our integrated model better specifies how these goals might be cognitively structured and what implications this has for detection when multiple goals are in play. In addition to leading to the novel results mentioned above, the updated model can also serve as a useful basis for future research aimed at better understanding how consumers acquire and use PK to inform judgment (see further discussion below).

Our results also suggest novel practical contributions in terms of their implications for government (e.g., FTC) and consumer advocacy organizations (e.g., Consumer Reports, TINA.org) that seek to alert consumers to the potential use of misleading or manipulative advertising tactics. As mentioned, these groups often rely on specific training procedures that our research shows can have negative side effects in terms of lowering the detection of other misleading tactics. Our research further shows that more general training procedures that encourage the critical evaluation of ads can help consumers recognize a broader range of unethical advertising tactics. Finally, the general training procedure developed here is relatively short and is effective in online contexts involving a broad range of consumers, suggesting it has practical utility. Our results suggest that government agencies and consumer advocacy organizations interested in protecting consumers from deceptive advertising should test their current consumer training materials for goal shielding side effects. Further, if such side effects are found, these materials should be revised along the lines of our general training procedure. That is, the emphasis of training should be on the relatively higher level goal of vigilance and/or critical thinking, rather than emphasizing any individual deceptive tactic.

Implications for Ethics

This research responds to the call for greater consideration of unintended consequences by business ethicists (Koehn, 2011). As has been suggested elsewhere (e.g., Koehn, 2011), well intentioned market interventions should be accountable not only for those good intentions, but also for any unintended consequences, which are often considerable. This is especially the case when the unintended consequences are reasonably foreseeable. We suggest that while the

unintended consequences of consumer education studied here may not be terribly intuitive, our research serves to make such unintended consequences more foreseeable, and should be considered in future educational endeavors.

Limitations

The current research included a number of limitations. First, some of the training results were significant only at the $p < 0.10$ level. It is worth noting that the same was true for some of the illegitimate endorser training effects in the original Sagarin et al. (2002) studies that inspired the current research. This may, in part, reflect the fact that participants received only a single exposure to the training materials in both cases. Both the more general advertising research literature (Pechmann & Stewart, 1992), and research concerned with teaching consumers to better detect misleading claims (Germelmann et al., 2020), suggest that multiple exposures to a message are typically needed to clearly establish the intended communication effects, especially when learning new concepts.

To further address any potential concerns about the reliability of our results, we also conducted a single-paper meta-analysis (SPM; McShane & Bockenholt, 2017) of the two experiments included in this paper, along with a third experiment using a student sample that closely replicated the main findings in Study 2, but was not included in our report due to its redundancy. This meta-analysis showed that each of the key effects in our studies was reliable. In particular, the SPM analysis showed that specific training (vs. control) was effective in improving detection of the illegitimate endorser (estimate = 0.873, SE = 0.126, $z = 6.95$, $p < 0.0001$), and also reliably exhibited the unintended side effect on footnote detection (estimate = 0.632, SE = 0.168, $z = 3.77$, $p < 0.001$). Moreover, analysis of the two experiments that included general training revealed this was effective (relative to control) in enhancing detection of both the illegitimate endorser (estimate = 0.320, SE = 0.143, $z = 2.24$, $p < 0.05$) and the limiting footnote (estimate = 0.350, SE = 0.143, $z = 2.45$, $p < 0.05$).

A second limitation is that our research examined the immediate effects of general and specific training rather than longer-term effects. While there is some reason to believe that PK training can last for some time afterwards (Boerman et al., 2020; Sagarin et al., 2002), it would be of benefit to examine this question for both the ironic effects of specific training and the broader effects of general training shown here. Additional general training (re)exposures may be needed from time-to-time to maintain vigilance towards unethical tactics. Finally, we examined the effects of training in what are traditionally considered low or very low involvement contexts (Aguinis et al., 2021; Sears, 1986). Higher involvement levels may further improve the effectiveness of

general training (Friestad & Wright, 1994), but there is some question as to whether such involvement would attenuate or accentuate the ironic effects of specific tactic training. Future research is needed to examine these remaining issues.

Future Research

As stated above, the theorizing and research reported here suggest a number of practical and theoretical questions that might be addressed in future research. In terms of the practical ethics of training consumers to detect and appropriately respond to unethical or manipulative advertising, our results highlight the need for additional research at the intersection of PK and goal systems, given that knowledge acquisition and goal activation can have competing or complimentary effects on the detection of PK tactics. Clearly, in order for consumers to be capable of detecting deceptive tactics, they must at some point acquire such specific knowledge. However, traditional training approaches that focus on imparting knowledge about specific tactics may unintentionally leave consumers more vulnerable to other (non-focal) tactics. Future research should explore the optimal timing and sequencing of PK acquisition and activation to develop training programs that optimally combine the learning and operation of PK. More generally, future research should consider the more comprehensive effects that training can have, and in particular the potential side effects identified in the research here, rather than exclusively demonstrating their intended effects.

More conceptually, future research should further examine the interplay of goal systems and the acquisition of the PK that serves those goals. Existing research in the area (e.g. Campbell & Kirmani, 2000) largely treats activation of vigilance goals as a consequence of PK activation. The present research, by contrast, views PK activation as an outcome of either specific or general vigilance goal activation. Future research should further explore this apparently bi-directional relationship to develop our understanding of whether and how the causal sequence of effects determines coping with persuasive attempts. Other motives may also play an important role in PK (de)activation. For example, high need for closure can lead to snap judgments based on easily processed information in the choice context (Boush et al., 2009), which may undermine the activation of goals that require more critical thought prior to judgment. Feelings of invulnerability may also suppress detection goals (Chang, 2017), and as a result thwart attempts to activate either specific or general vigilance goals. Finally, increasing the salience of PK tactics after-the-fact (i.e., after consumers have already been tricked by the tactic) is known to produce a defensive form of suspicion that can have broad negative effects on subsequent persuasive attempts (Darke and Ritchie, 2007). This suggests that specific training that

involves teaching consumers about specific tactics after they have been tricked by them may actually increase the detection of other PK tactics in subsequent test ads.

Summary

The current research developed a new goal systems approach to better understand the role of multiple vigilance goals in the detection of misleading advertising tactics. This led to the prediction and finding that training consumers to detect specific tactics can have negative side-effects in terms of limiting the detection of other tactics that are not the focus of the specific training. In contrast, the model also predicted that more general training, focused on higher level vigilance goals, would be more effective in helping consumers identify a broader range of unethical advertising tactics, which was largely supported by the evidence.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10551-021-04937-7>.

Acknowledgements This research was supported by a standard grant from the Social Science and Humanities Research Counsel of Canada to the second author.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aguinis, H., Villamor, I., & Ramani, R. S. (2021). Mturk research: Review and recommendations. *Journal of Management*, *47*(4), 823–837.
- Anderson, K. B. (2017). Mass-market consumer fraud in the United States, A 2017 update. Staff Report of the Bureau of Economics, Federal Trade Commission. Retrieved from <https://www.ftc.gov/system/files/documents/reports/mass-market-consumer-fraud-united-states-2017-update/p105502massmarketconsumerfraud2017report.pdf>
- Attas, D. (1999). What's wrong with "deceptive" advertising? *Journal of Business Ethics*, *21*(1), 49–59.
- Bendig, A. W. (1954). Reliability and the number of rating-scale categories. *Journal of Applied Psychology*, *38*, 38–40.
- Bergkvist, L., & Rossiter, J. R. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research*, *44*(2), 175–184.
- Boerman, S. C., Tessitore, T., & Mueller, C. M. (2020). Long-term effects of brand placement disclosure on persuasion knowledge and brand responses. *International Journal of Advertising*, *40*(1), 26–48.
- Boush, D. M., Friestad, M., & Rose, G. M. (1994). Adolescent skepticism toward T.V. Advertising and knowledge of advertiser tactics. *Journal of Consumer Research*, *21*(1), 165–75.
- Boush, D. M., Friestad, M., & Wright, P. (2009). *Deception in the Marketplace: The Psychology of Deceptive Persuasion and Consumer Self-Protection*. Routledge.
- Brown, C. L., & Krishna, A. (2004). The skeptical shopper: A metacognitive account for the effects of default options on choice. *Journal of Consumer Research*, *31*, 529–539.
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*, 116–131.
- Cain, R. M. (2011). Embedded advertising on television: Disclosure, deception, and free speech rights. *Journal of Public Policy and Marketing*, *30*, 226–238.
- Campbell, M. C. (1995). When attention-getting advertising tactics elicit consumer inferences of manipulative intent: The importance of balancing benefits and investments. *Psychology & Marketing*, *4*(3), 225–254.
- Campbell, M. C., & Kirmani, A. (2000). Consumers' use of persuasion knowledge: The effects of accessibility and cognitive capacity on perceptions of an influence agent. *Journal of Consumer Research*, *27*(1), 69–83.
- Campbell, M. C., & Kirmani, A. (2008). I know what you're doing and why you're doing it: The use of persuasion knowledge in consumer research. In C. Haugtvedt, P. Herr, & F. Kardes (Eds.), *Handbook of Consumer Psychology* (pp. 549–571). Lawrence Erlbaum Associates.
- Chan, F. F. Y. (2020). Prior disclosure of product placement: The more explicit the disclosure, the better the brand recall and brand attitude. *Journal of Business Research*, *120*(November), 31–41.
- Chan, E., & Sengupta, J. (2010). Insincere flattery actually works: A dual attitudes perspective. *Journal of Marketing Research*, *47*(1), 122–133.
- Chang, C. (2017). A metacognitive model of the effects of susceptibility to persuasion self-beliefs on advertising effects. *Journal of Advertising*, *46*(4), 487–502.
- Charlton, A. B., & Cornwell, T. B. (2019). Authenticity in horizontal marketing partnerships: A better measure of brand compatibility. *Journal of Business Research*, *100*(July), 279–298.
- Cho, Y., & Taylor, C. R. (2020). The role of ambiguity and skepticism in the effectiveness of sustainability labeling. *Journal of Business Research*, *120*(November), 379–388.
- Chung, Y. J., & Kim, E. (2020). Predicting consumer avoidance of native advertising on social networking sites: A survey of Facebook users. *Journal of Promotion Management*, *27*(1), 1–26.
- Cialdini, R. B. (2021). *Influence, New and Expanded: The Psychology of Persuasion*. HarperCollins Publishers.
- Cohen, J. (2019). Bringing down the average: The case for a "Less Sophisticated" reasonableness standard in US and EU consumer law. *Loyola Consumer Law Review*, *32*(1), 1–44.
- Darke, P. R., Freedman, J. L., & Chaiken, S. (1995). Percentage discounts, initial price, and bargain hunting: A heuristic-systematic approach to price search behavior. *Journal of Applied Psychology*, *80*(5), 580–586.
- Darke, P. R., & Ritchie, R. B. (2007). The defensive consumer: Advertising deception, defensive processing, and distrust. *Journal of Marketing Research*, *44*(1), 114–127.
- De Kerpel, L., Van Kerckhove, A., and Tessitore, T. (2019). Starting with an apology: paving the way to consumer persuasion? In K. Haws and B. McFerran (Eds.) Society of Consumer Psychology Annual Conference Proceedings, Savannah, GA.

- Eisend, M., & Tarrahi, F. (2021). Persuasion knowledge in the marketplace: A meta-analysis. *Journal of Consumer Psychology, 10*, 2.
- Eisend, M., van Reijmersdal, E. A., Boerman, S. C., & Tarrahi, F. F. (2014). A meta-analysis of the effects of disclosing sponsored content. *Journal of Advertising, 49*(3), 344–366.
- Fair, L. (2014). Full Disclosure. Federal Trade Commission Business Blog. Retrieved from <http://ftc.gov/news-events/blogs/business-blog/2014/09/full-disclosure>
- Fernandez, J. R., & Kruglanski, A. W. (2019). The psychology of multiple goal pursuit: Choices, configurations, and commitments. *Journal of the Association for Consumer Research, 4*(1), 5–12.
- Förster, J., Liberman, N., & Friedman, R. S. (2007). Seven principles of goal activation: A systematic approach to distinguishing goal priming from priming of non-goal constructs. *Personality and Social Psychology Review, 11*(3), 211–233.
- Förster, J., Liberman, N., & Higgins, E. T. (2005). Accessibility from active and fulfilled goals. *Journal of Experimental Social Psychology, 41*(3), 220–239.
- Friestad, M., & Wright, P. (1994). The persuasion knowledge model—How people cope with persuasion attempts. *Journal of Consumer Research, 21*(1), 1–31.
- Ferguson, J. L., Ellen, P. S., & Piscopo, G. H. (2011). Suspicion and perceptions of price fairness in times of crisis. *Journal of Business Ethics, 98*(2), 331–349.
- Germelmann, C. C., Herrmann, J. L., Kacha, M., & Darke, P. R. (2020). Congruence and incongruence in thematic advertisement-medium combinations: Role of awareness, fluency, and persuasion knowledge. *Journal of Advertising, 49*(2), 141–164.
- Grillo, T. L. H., & Pizzutti, C. (2021). Recognizing and trusting persuasion agents: Attitudes bias trustworthiness judgments, but not persuasion detection. *Personality and Social Psychology Bulletin, 47*, 796–809.
- Ham, C. D., & Nelson, M. R. (2019). The reflexive persuasion game: The Persuasion knowledge model (1994–2017). In S. Rodgers & E. Thorson (Eds.), *Advertising Theory* (pp. 124–140). Taylor & Francis.
- Harris, R. J. (1977). Comprehension of pragmatic implications in advertising. *Journal of Applied Psychology, 62*(5), 603–608.
- Hayes, A. F. (2017). *Introduction to Mediation, Moderation, and Conditional Process Analysis, A Regression-Based Approach* (2nd ed.). Guilford Publications. New York.
- Herbst, K. C., Finkel, E. J., Allan, D., & Fitzsimons, G. M. (2012). On the dangers of pulling a fast one: Advertisement disclaimer speed, brand trust, and purchase intention. *Journal of Consumer Research, 38*(5), 909–919.
- Hernandez, J. M. C., Wright, S. A., & Affonso, F. M. (2019). The importance of advertising skepticism for brand extension appeals. *Psychology & Marketing, 36*(7), 687–699.
- Hwang, Y., & Jeong, S. (2019). Editorial content in native advertising: How do brand placement and content quality affect native-advertising effectiveness? *Journal of Advertising Research, 59*(2), 208–218.
- Isaac, M. S., & Grayson, K. (2017). Beyond skepticism: Can accessing persuasion knowledge bolster credibility? *Journal of Consumer Research, 43*(6), 895–912.
- Johar, G. V., & Simmons, C. J. (2000). The use of concurrent disclosures to correct invalid inferences. *Journal of Consumer Research, 26*(4), 307–322.
- Kardes, F. R., Posavac, S. S., Silvera, D., Cronley, M. L., Sanbonmatsu, D. M., Schertzer, S., Miller, F., Herr, P. M., & Chandrashekar, M. (2006). Debiasing omission neglect. *Journal of Business Research, 59*(6), 786–792.
- Kirman, A., & Zhu, R. (2007). Vigilant against manipulation: The effect of regulatory focus on the use of persuasion knowledge. *Journal of Marketing Research, 44*(4), 688–701.
- Koehn, D. (2011). Care ethics and unintended consequences. In M. Sander-Staudt & M. Hamington (Eds.), *Applying Care Ethics to Business, Issues in Business Ethics 34* (pp. 141–153). Springer.
- Kopetz, C. E., Kruglanski, A. W., Arens, Z. G., Etkin, J., & Johnson, H. M. (2012). The dynamics of consumer behavior: A goal systemic perspective. *Journal of Consumer Psychology, 22*(2), 208–223.
- Litman, L., Robinson, J., & Abberbock, T. (2016). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods, 49*, 433–442.
- Main, K. J., Dahl, D. W., & Darke, P. R. (2007). Deliberative and automatic bases of suspicion: Empirical evidence of the sinister attribution error. *Journal of Consumer Psychology, 17*(1), 59–69.
- Main, K. J., Guo, W., & White, K. (2019). Sensitivity to ulterior motives in retail settings: The moderating role of dual-identity versus sole-identity consumers. *Journal of Retailing, 95*(3), 63–75.
- McShane, B. R., & Bockenholt, U. (2017). Single-paper meta-analysis: Benefits for study summary, theory testing, and replicability. *Journal of Consumer Research, 43*, 1048–1063.
- Morwitz, V. G., & Munz, K. P. (2021). Intentions. *Consumer Psychology Review, 4*(1), 26–41.
- Obermiller, C., & Spangenberg, E. R. (1998). Development of a scale to measure consumer skepticism toward advertising. *Journal of Consumer Psychology, 7*(2), 159–186.
- Obermiller, C., Spangenberg, E. R., & MacLachlan, D. L. (2005). Ad skepticism: The consequences of disbelief. *Journal of Advertising, 34*(3), 7–17.
- Pechmann, C., & Stewart, D. W. (1992). Advertising repetition: A critical review of wearin and wearout. *Journal of Current Issues and Research in Advertising, 11*(2), 285–330.
- Peltzman, S. (1981). The effects of FTC advertising regulation. *Journal of Law and Economics, 24*(3), 403–448.
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In R. E. Petty & J. T. Cacioppo (Eds.), *Communication and Persuasion*. Springer, New York: Springer Series in Social Psychology.
- Pratkanis, A. R., & Aronson, E. (2001). *Age of propaganda: The Everyday Use and Abuse of Persuasion*. W.H. Freeman and Company.
- Preston, I. (1994). *The Tangled Web They Weave: Truth, Falsity and Advertisers*. University of Wisconsin Press.
- Russo, J. E., Metcalf, B. L., & Stephens, D. (1981). Identifying misleading advertising. *Journal of Consumer Research, 8*(2), 119–131.
- Sagarin, B. J., Cialdini, R. B., Rice, W. E., & Serna, S. B. (2002). Dispelling the illusion of invulnerability: The motivations and mechanisms of resistance to persuasion. *Journal of Personality and Social Psychology, 83*(3), 526–541.
- Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology, 51*(3), 515–530.
- Shah, J. Y., Friedman, R., & Kruglanski, A. W. (2002). Forgetting all else: On the antecedents and consequences of goal shielding. *Journal of Personality and Social Psychology, 83*(6), 1261–1280.
- Shah, J. Y., Kruglanski, A. W., & Friedman, R. (2003). Goal systems theory: Integrating the cognitive and motivational aspects of self-regulation. In S. J. Spencer, S. Fein, M. P. Zanna, & J. M. Olson (Eds.), *Motivated Social Perception: The Ontario Symposium* (Vol. 9, pp. 247–276). Erlbaum.
- Stearns, J. M., & Borna, S. (1995). The ethics of lottery advertising: Issues and evidence. *Journal of Business Ethics, 14*(1), 43–51.
- Stewart, D. W., & Martin, I. M. (2004). Advertising disclosures: Clear and conspicuous or understood and used? *Journal of Public Policy & Marketing, 23*(2), 183–192.

- Tessitore, T., & Geuens, M. (2019). Arming consumers against product placement: A comparison of factual and evaluative educational interventions. *Journal of Business Research*, 95, 38–48.
- Wei, M., Fischer, E., & Main, K. J. (2008). An examination of the effects of activating persuasion knowledge on consumer response to brands engaging in covert marketing. *Journal of Public Policy & Marketing*, 27, 34–44.
- Wentzel, C., Tomczak, T., & Herrmann, A. (2010). The moderating effect of manipulative intent and cognitive resources on the evaluation of narrative ads. *Psychology & Marketing*, 27(5), 510–530.
- Xie, G. X., Madrigal, R., & Boush, D. M. (2015). Disentangling the effects of perceived deception and anticipated harm on consumer responses to deceptive advertising. *Journal of Business Ethics*, 129(2), 281–293.
- Xie, G. X., Chang, H., & Rank-Christman, T. (2020). Contesting dishonesty: When and why perspective-taking decreases ethical tolerance of marketplace deception. *Journal of Business Ethics*. <https://doi.org/10.1007/s10551-020-04582-6>
- Yang, J., & Mundel, J. (2021). “Are we all in this together?”: Brand opportunism in covid-19 cause related marketing and the moderating role of consumer skepticism. *Journal of Promotion Management*, 27, 77–899.
- Zarouali, B., Poels, K., Ponnet, K., & Walrave, M. (2021). The influence of a descriptive norm label on adolescents’ persuasion knowledge and privacy-protective behavior on social networking sites. *Communication Monographs*, 88(1), 1–21.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.