Warnings of Adverse Side Effects Can Backfire Over Time

Yael Steinhart, Ziv Carmon and Yaacov Trope

Psychological Science 2013 24: 1842 originally published online 2 August 2013
DOI: 10.1177/0956797613478948

Research Report

Psychological Science
24(9) 1842–1847

DOI: 10.1177/0956797613478948
pss.sagepub.com

Warnings of Adverse Side Effects Can Backfire Over Time

Yael Steinhart[1], Ziv Carmon[2], and Yaacov Trope[3]
[1]Faculty of Management, Tel Aviv University; [2]Marketing Department, INSEAD, Singapore; and [3]Department of Psychology, New York University

Abstract

Warnings that a promoted product can have adverse side effects (e.g., smoking cigarettes can cause cancer) should dampen the product's allure. We predicted that with temporal distance (e.g., when an ad relates to future consumption or was viewed some time earlier), this common type of warning can have a worrisome alternative consequence: It can ironically boost the product's appeal. Building on construal-level theory, we argue that this is because temporal distance evokes high-level construal, which deemphasizes side effects and emphasizes message trustworthiness. In four studies, we demonstrated this phenomenon. For example, participants could buy cigarettes or artificial sweeteners after viewing an ad promoting the product. Immediately afterward, the quantity that participants bought predictably decreased if the ad they saw included a warning about adverse side effects. With temporal distance (product to be delivered 3 months later, or 2 weeks after the ad was viewed), however, participants who had seen an ad noting the benefits of the product but warning of risky side effects bought more than those who had seen an ad noting only benefits.

Keywords

decision making, judgment, side effects, warnings

Received 4/22/12; Revision accepted 1/22/13

People often encounter promotional messages that promise seductive benefits, such as social status if one smokes a brand of cigarettes or sexual potency if one consumes an erectile-dysfunction medication, and warn of potential side effects, such as a stroke or cancer (Erceg-Hurn & Steed, 2011; World Health Organization, 2005). Such warnings of adverse side effects buffer firms from liability (Hart, 2010), as disclosing risks to consumers presumably facilitates safer choices and dampens the desirability of risky behaviors. However, there are indications that warnings may not necessarily be effective. For example, although cigarette ads typically include grave warnings (U.S. Food and Drug Administration, FDA, 2012), a meta-analysis (Nelson, 2006) suggests that warnings are not associated with decreased cigarette consumption. More generally, opinions about the effectiveness of warnings are mixed. Some researchers suggest that warnings enhance welfare (Saffer & Chaloupka, 2000). Others claim that warnings can inhibit consideration of risks (Grandpre, Alvaro, Burgoon, Miller, & Hall, 2003; Ruiter & Kok, 2005), lead to overreaction (Griffin & Harris, 2011), or even invoke side effects (Blackman, 2009). Altogether, not enough is known about the effects of warnings (Cecil, Evans, & Stanley, 1996; Main & Argo, 2004; Torres, Sierra, & Heiser, 2007).

Building on construal-level theory (CLT; Liberman & Trope, 2008), we explored the effects of warnings embedded in promotional messages and how those warnings are affected by temporal perspective. Information can be represented at different levels of construal (cf. Strack, Schwarz, & Gschneidinger, 1985; Vallacher & Wegner, 1989): Low-level construal emphasizes concrete, secondary, lower-level features; high-level construal emphasizes the gist—abstract, primary, higher-level features. CLT suggests that what is more psychologically distant (e.g.,

Corresponding Author:
Ziv Carmon, INSEAD, 1 Ayer Rajah Ave., Singapore 138676, Singapore
E-mail: ziv.carmon@insead.edu

spatially, socially, or temporally) is represented at higher construal levels.

We propose that embedding warnings of undesirable side effects in promotional messages has two conflicting outcomes. On the one hand, the appeal of the advertised product is decreased because of the side effects specified in the warning. On the other hand, message trustworthiness is increased, as the presence of a warning connotes that the communication is two-sided (presents both pros and cons), and this increases the message recipient's sense of knowledge and confidence about the product (Eisend, 2006; Petty, Wegener, & Fabrigar, 1997; Rucker, Petty, & Briñol, 2008; Schwarz, Sanna, Skurnik, & Yoon, 2007). Building on CLT, we suggest that temporal distance deemphasizes the potential side effects specified in warnings, because side effects are inherently secondary in nature and thus reflect low-level construal. Conversely, temporal distance emphasizes the trustworthiness of the message (driven by the presence of a warning), because it reflects high-level construal, as it is more abstract, general, and primary in nature. The pilot studies we conducted in support of the present research further supported classifying side effects as low-level construal and trustworthiness as high-level construal.

To summarize, we predicted that from a proximal time perspective, a warning will have the predictable effect of reducing the appeal of the product. However, from a distal time perspective, as information is construed at a higher level (Liberman & Trope, 2008), we expected the presence of a warning to ironically boost the appeal of the risky behavior, a phenomenon we term *ironic effect of warnings*, our central prediction in this article.

This predicted effect relates to the *sleeper effect*, which states that a message's negative impact decreases over time (Pratkanis, Greenwald, Leippe, & Baumgardner, 1988). We expected that, as predicted by the sleeper effect, negative effects of warnings would fade over time. However, we also expected that, unlike predictions of the sleeper effect, delay would emphasize the positive connotation of warnings and boost the appeal of the promoted behavior. Moreover, the sleeper effect concerns past temporal distance, whereas we expected our predictions to hold for both past and future.

In four studies, we examined ad perceptions and tested the ironic effect of warnings on actual purchase decisions and evaluations of different products. Results of these main studies (supplemented by several pilot studies conducted with participants drawn from the same population used for the main studies) illustrate the robustness of the effect. All studies were conducted in Israel in Hebrew. Experimental materials presented in this article were translated into English and sums were converted into U.S. dollars.

Study 1: Effects of Warning Labels on Delayed Cigarette Acquisition

Method

Seventy-one people (20% women, 80% men; mean age = 29 years) were randomly assigned to four conditions in a 2 (with warning vs. without warning) × 2 (near-future purchase vs. distant-future purchase) design. Potential participants were screened to ensure eliminating nonsmokers as well as those already familiar with the brand used. Participants viewed an ad promoting a cigarette brand that was (supposedly) going to be sold in their region soon.[1] Half the participants saw a version of the ad that also warned of smoking risks. The other half saw the same ad without the warning. In a pilot study (N = 25), participants were asked to rate on separate 7-point scales the degree to which message trustworthiness and potential side effects are central to product judgments. The results confirmed that for this study's ad, message trustworthiness was perceived as more central (M = 4.08, SD = 2.02) than the potential side effects (M = 3.40, SD = 2.12), $t(24) = 2.36$, $p < .05$.

After viewing the ad, all participants had an opportunity to purchase the product—they saw a pack of the advertised brand, were told that each pack costs $2.00, and were asked whether and how many packs they wished to buy. Half the participants were told that they would receive any packs they ordered within 24 hr (near-future condition); the other half were told that they would receive the packs 3 months later (distant-future condition). To avoid distributing potentially harmful products, we apologized to participants after the study concluded and explained that we would not provide the product.

Results

Supporting our predictions, an analysis of variance (ANOVA) on the number of packs participants chose to buy revealed a significant two-way interaction between the presence or absence of a warning and the timing of receiving the packs, $F(1, 67) = 8.93$, $p < .05$; $\eta^2 = .12$. In the distant-future condition, participants who saw the warning label chose to buy more packs (M = 5.16, SD = 7.38) than participants who did not see the warning label ($M = 0.87$, $SD = 0.99$), $t(32) = 2.24$, $p < .05$. Conversely, in the near-future condition, participants who did not see the warning label chose to buy more packs (M = 3.00, SD = 5.04) than participants who did see the warning label ($M = 0.75$, $SD = 1.02$), $t(35) = 1.96$, $p = .06$. Thus, although smoking risks seem ubiquitous, participants were affected by the warning.

Study 2: Effects of Warning Labels on Delayed Purchase of Artificial Sweetener

In Study 2, we conceptually replicated Study 1 by manipulating past rather than future temporal distance with a different product category.

Method

Seventy-four women (mean age = 35 years) were randomly assigned to four conditions in a 2 (with warning vs. without warning) × 2 (immediate evaluation vs. delayed evaluation) design. Potential participants were screened for product relevance (i.e., whether they consumed artificial sweeteners). Half the participants saw an ad promoting an artificial sweetener that included a warning about risks associated with the product, whereas the other half saw the same ad but without the warning. A pilot study (N = 25) confirmed that for this study's ad, message trustworthiness was perceived as more central to product judgments (M = 6.44, SD = 1.00) than potential side effects (M = 5.84, SD = 1.70), $t(24) = 2.08$, $p < .05$.

After viewing the ad, half the participants were offered an opportunity to purchase the product (immediate-choice condition). They saw a package costing $0.50 that contained 25 packets and were asked whether and how many packages they wished to buy. We told the remaining participants that we would contact them again 2 weeks later (delayed-choice condition). Two weeks later, we indeed contacted them, offering an opportunity to buy the advertised product. We apologized to participants after the study concluded and explained that, for logistical reasons, we would not sell the product.

Results

Supporting our predictions, an ANOVA on the number of packages participants chose to buy revealed a significant two-way interaction between the presence or absence of a warning label and the time at which participants made a buying decision, $F(1, 70) = 11.21$, $p < .05$; $\eta^2 = .13$. In the immediate-choice condition, participants who saw the ad without the warning label chose to buy more packages (M = 5.50, SD = 7.58) than participants who saw the ad with the warning label (M = 0.33, SD = 0.82), $t(33) = 2.64$, $p < .05$. In the delayed-choice condition, however, participants who saw the ad with the warning label chose to buy more packages (M = 2.92, SD = 5.04) than participants who saw the ad without the warning label (M = 0.80, SD = 1.35), $t(37) = 2.03$, $p = .05$.

Study 3: Effects of Warning Labels on Delayed Perceptions of an Erectile-Dysfunction Medication

In this study, we further explored the process underlying the ironic effect of warnings by examining how perceptions of the different meanings of the warning change with temporal perspective (distal vs. proximal).

Method

Thirty-four men (mean age = 58 years) were randomly assigned to two conditions: near-future event and distant-future event. All participants viewed an ad promoting erectile-dysfunction medication that featured a warning of potential side effects, such as heart disease. Half the participants were told that the product would soon be launched (near-future condition); the other half were told that the product would be launched the following year (distant-future condition). All participants were asked to rate on separate 7-point scales the product's attractiveness and their agreement with the following statements: "The warning increases my concerns about potential product side effects," and "the warning increases the trustworthiness of the ad." As with the earlier studies, a pilot study (N = 25) showed that for this ad, message trustworthiness was perceived as more central to product judgments (M = 5.44, SD = 1.65) than the potential side effects of the medication (M = 4.84, SD = 1.86), $t(24) = 2.32$, $p < .05$.

Results

Consistent with our predictions and with our earlier results, the findings in Study 3 revealed that participants in the distant-future condition rated the product as more attractive (M = 4.67, SD = 0.72) than did those in the near-future condition (M = 3.74, SD = 1.59), $t(32) = 2.09$, $p < .05$. Analyzing ratings as a function of temporal perspective (product-launch timing: near future vs. distant future) revealed that when product-launch timing was distant, the ad was rated as more trustworthy in the distant-future condition (M = 5.13, SD = 1.24) than in the near-future condition (M = 4.11, SD = 1.63), $t(32) = 2.02$, $p = .052$. Conversely, concern with the potential side effects (lower-level feature) was rated higher in the near-future condition (M = 5.42, SD = 1.12) than in the distant-future condition (M = 4.47, SD = 1.51), $t(32) = 2.21$, $p < .05$. Mediation analyses confirmed the underlying role of higher-level features—the effect of temporal distance on product evaluations was mediated by message trustworthiness—95% confidence interval (CI) = [0.01, 0.90].

Study 4: Effects of Warning Labels on Delayed Evaluations of Hair-Loss Medication

In this study, we further examined our claim that the perceived trustworthiness of the ad drives the ironic effect of warnings.

Method

Seventy-six men (mean age = 54 years) were randomly assigned to four conditions in a 2 (with warning vs. without warning) × 2 (immediate evaluation vs. delayed evaluation) design. Half the participants saw an ad for a hair-loss medication that also warned that weight gain was a common side effect. Other participants saw the same ad without the warning. A pilot study ($N = 25$) confirmed that for this study's ad, message trustworthiness was perceived as more central to product judgments ($M = 5.80$, $SD = 1.65$) than potential side effects ($M = 5.29$, $SD = 1.87$), $t(24) = 2.52$, $p < .05$.

Immediately after seeing the ad, half the participants evaluated the product (immediate-evaluation condition), whereas the remaining participants were contacted 2 weeks later and asked to evaluate the product then (delayed-evaluation condition). Participants rated product attractiveness and ad trustworthiness on a 7-point scale.

Results

The results consistently supported our predictions. An ANOVA on product ratings revealed a significant two-way interaction between the presence or absence of the warning and the time of the evaluation, $F(1, 71) = 14.42$, $p < .05$; $\eta^2 = .16$. In the immediate-evaluation condition, product ratings were more favorable when a warning label was absent ($M = 4.29$, $SD = 1.53$) than when a warning label was present ($M = 3.17$, $SD = 1.75$), $t(33) = 2.02$, $p = .05$. However, in the delayed-evaluation condition, ratings were higher when the warning label was present ($M = 4.94$, $SD = 1.19$) than when it was absent ($M = 3.35$, $SD = 1.58$), $t(38) = 3.47$, $p < .05$. An ANOVA on ad trustworthiness also revealed a significant two-way interaction, $F(1, 71) = 5.11$, $p < .05$; $\eta^2 = .07$. In the immediate-evaluation condition, ad trustworthiness did not differ regardless of whether the warning was present ($M = 3.56$, $SD = 1.72$) or absent ($M = 3.65$, $SD = 1.45$), $t(33) = 0.17$, $p > .1$. However, in the delayed-evaluation condition, participants evaluated the ad with the warning as more trustworthy ($M = 4.88$, $SD = 1.57$) than did participants who evaluated the ad without the warning ($M = 3.35$, $SD = 1.43$), $t(38) = 3.20$, $p < .05$. Moderated mediation analyses using bootstrapping mediation tests (e.g., Preacher, Rucker, & Hayes, 2007) with 5,000 replications revealed that, as predicted, ad trustworthiness mediated the temporal effect when the warning was present (95% CI = [0.13, 1.03]) but not when the warning was absent (95% CI = [–0.49, 0.20]).

General Discussion

Messages that promote potentially risky behaviors, such as smoking or taking a medication, frequently warn of potential adverse side effects. Such warnings are often provided under the guidance of well-meaning organizations, such as the U.S. FDA. An underlying assumption (Wogalter & Laughery, 1996) is that such warnings reduce the appeal of the risky behavior. The current research shows how and why this assumption can be misguided—warnings about detrimental side effects can ironically backfire in common situations when time separates the message from the related behavior. This is because warnings can have two conflicting effects: (a) decreasing the appeal of the object of the warning and (b) increasing the apparent trustworthiness of the ad because of the two-sided nature of the communication. As time separates the message from the behavior of which it warns, the prominence of side effects is attenuated, and the trustworthiness of the ad rises.

The ironic effect of warnings documented in this article is counterintuitive. Marketing experts we surveyed, for example, mispredicted it—23 marketing professors who examined the ad used in Study 1 expected the likelihood of evaluations becoming more favorable over time to be lower ($M = 2.09$, $SD = 1.62$) than the likelihood of evaluations becoming less favorable over time ($M = 3.52$, $SD = 2.27$), $t(22) = 2.33$, $p < .05$.

This research offers additional theoretical contributions. By identifying conditions under which warnings can backfire, our research informs the debate on the effects of warning of unintended consequences (cf. Skurnik, Yoon, Park, & Schwarz, 2005). Our work also extends CLT research to settings in which information is used some time after it is encountered rather than immediately (past temporal CLT work focused on currently presented information for decisions about the future; Trope & Liberman, 2010).

Future research could extend the understanding of the ironic effect of warnings by exploring characteristics of the warning, message, and context that may also mediate or moderate this phenomenon. For example, it seems that whether the warning was mandated or made voluntarily (Eisend, 2006) might matter (note that pilot studies in which we explored this reasonable possibility failed; manipulation checks revealed that most participants did not believe the warning was made voluntarily). Whether the warning seemed difficult to notice (e.g., displayed

only briefly; Herbst, Finkel, Allan, & Fitzsimons, 2012) could also be significant. Features of the message such as whether no product benefit is communicated (e.g., as on cigarette packs in some countries) may also matter.

Because using information encountered days earlier (Studies 2 and 4) relies on long-term memory, it may involve consolidation and retrieval processes in addition to encoding. Deep examination of these subtle processes is generally beyond the scope of this research. That said, memory follows the pattern predicted by CLT—compared with immediate recall, delayed recall is typically more abstract and schematic (Alba & Hasher, 1983; Bartlett, 1932; Hintzman, 1986) and tends to reflect metacognitive knowledge (Strack & Förster, 1998). Further support for the present findings comes from follow-up studies in which we explored recollection of the ads used in the main studies. Independent judges coded the information recalled as high level (e.g., "I saw a warning") or low level (e.g., "you may experience weakness"). All initial disagreements (6%–8%) were resolved via discussion. In one follow-up, within either minutes or days of seeing the ad used in Study 2, 100 women (mean age = 34 years) were asked to recall the content of the ad. As expected, recollections were judged as less high level in the immediate-recall condition (20.9%) than in the delayed-recall condition (63.2%), $\chi^2(2, N = 100) = 52.09$, $p < .001$. The same pattern emerged in a follow-up experiment to Study 4 with 30 men (mean age = 60 years): Recollections were judged as less high level in the immediate-recall condition (28.6%) than in the delayed-recall condition (66.7%), $\chi^2(2, N = 30) = 7.31$, $p < .05$. The recall pattern emerged again in a follow-up experiment to Study 3 with 30 men (mean age = 61 years), in which temporal distance was future-rather than past-related. Recollections were judged less high level in the proximate-recall condition (23.1%) than in the distal-recall condition (58.8%), $\chi^2(2, N = 30) = 7.03$, $p < .05$. Future research could seek further insights, for example, by exploring how psychological distance affects different aspects of memory (e.g., information encoding and retrieval).

In conclusion, our findings are disturbing—including a warning (e.g., risk of emphysema) in promotional messages (e.g., a cigarette ad) can stimulate rather than curtail the very behavior the warnings aim to discourage (e.g., smoking). Worse, this effect can "fly under the radar," as warnings may backfire only some time after the message is viewed (when consumption decisions are likely) rather than soon after the message is viewed (when regulators examine message impact). More broadly, that warnings of adverse side effects can backfire has significant implications in such important domains as health care, finance, and law. Depending on the timing, informing people of risks associated with a medical procedure or an investment option, for example, may ironically increase the likelihood that they adopt the risky behavior.

Author Contributions

Y. Steinhart, Z. Carmon, and Y. Trope developed the experimental concept, contributed to the design, and approved the final version of the manuscript for submission.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Note

1. Reproductions of the ads used in all four studies can be found online at http://www.insead.edu/facultyresearch/study-figures.html.

References

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin*, *93*, 203–231.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, England: Cambridge University Press.

Blackman, S. (2009, April 29). Why health warnings can be bad. *Financial Times*. Retrieved from http://www.ft.com/cms/s/0/0d9499ec-2d75-11de-9eba-00144feabdc0.html#axzz2WrGH3nGH

Cecil, H., Evans, R. I., & Stanley, M. A. (1996). Perceived believability among adolescents of health warning labels on cigarette packs. *Journal of Applied Social Psychology*, *26*, 502–519.

Eisend, M. (2006). Two-sided advertising: A meta-analysis. *International Journal of Research in Marketing*, *23*, 187–198.

Erceg-Hurn, D. M., & Steed, L. G. (2011). Does exposure to cigarette warnings elicit psychological reactance in smokers? *Journal of Applied Social Psychology*, *41*, 219–237.

Grandpre, J., Alvaro, E. M., Burgoon, M., Miller, C., & Hall, J. R. (2003). Adolescent reactance and anti-smoking campaigns: A theoretical approach. *Health Communication*, *15*, 349–366.

Griffin, D., & Harris, P. R. (2011). Calibrating the responses to health warnings: Limiting both overreaction and underreaction with self-affirmation. *Psychological Science*, *22*, 572–578.

Hart, A. N. (2010). Federal preemption of state-law failure-to-warn claims: Has the presumption against preemption gone too far? *Seventh Circuit Review*, *6*, 308–336.

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*, 909–919.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*, 411–428.

Liberman, N., & Trope, Y. (2008). The psychology of transcending the here and now. *Science*, *322*, 1201–1205.

Main, K. J., & Argo, J. J. (2004). A meta-analysis of the effectiveness of warning labels. *Journal of Public Policy and Marketing*, *23*, 193–208.

Nelson, J. P. (2006). Cigarette advertising regulation: A meta-analysis. *International Review of Law and Economics*, *26*, 195–226.

Petty, R. E., Wegener, D. T., & Fabrigar, L. R. (1997). Attitudes and attitude change. *Annual Review of Psychology*, *48*, 609–647.

Pratkanis, A. R., Greenwald, A. G., Leippe, M. R., & Baumgardner, M. H. (1988). In search of reliable persuasion effects: III. The sleeper effect is dead: Long live the sleeper effect. *Journal of Personality and Social Psychology*, *54*, 203–218.

Preacher, K. J., Rucker, D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, *42*, 185–227.

Rucker, D. D., Petty, R. E., & Briñol, P. (2008). What's in a frame anyway? A meta-cognitive analysis of the impact of one versus two sided message framing on attitude certainty. *Journal of Consumer Psychology*, *18*, 137–139.

Ruiter, R. A. C., & Kok, G. (2005). Saying is not (always) doing: Cigarette warning labels are useless. *European Journal of Public Health*, *15*, 329.

Saffer, H., & Chaloupka, F. (2000). The effect of tobacco advertising bans on tobacco consumption. *Journal of Health Economics*, *19*, 1117–1137.

Schwarz, N., Sanna, L. J., Skurnik, I., & Yoon, C. (2007). Metacognitive experiences and the intricacies of setting people straight: Implications for debiasing and public information campaigns. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 39, pp. 127–161). San Diego, CA: Academic Press.

Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research*, *31*, 713–724.

Strack, F., & Förster, J. (1998). Self-reflection and recognition: The role of metacognitive knowledge in the attribution of recollective experience. *Personality and Social Psychology Review*, *2*, 111–123.

Strack, F., Schwarz, N., & Gschneidinger, E. (1985). Happiness and reminiscing: The role of time perspective, affect, and mode of thinking. *Journal of Personality and Social Psychology*, *49*, 1460–1469.

Torres, I. M., Sierra, J. J., & Heiser, R. S. (2007). The effects of warning-label placement in print ads: A social contract perspective. *Journal of Advertising*, *36*, 49–62.

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*, 440–463.

U.S. Food and Drug Administration. (2012). *Cigarette health warnings*. Retrieved from http://www.fda.gov/TobaccoProducts/Labeling/ucm259214.htm

Vallacher, R. R., & Wegner, D. M. (1989). Levels of personal agency: Individual variation in action identification. *Journal of Personality and Social Psychology*, *57*, 660–671.

Wogalter, M. S., & Laughery, K. R. (1996). WARNING! Sign and label effectiveness. *Current Directions in Psychological Science*, *5*, 33–37.

World Health Organization. (2005). *WHO Framework Convention on Tobacco Control*. Retrieved from www.who.int/tobacco/framework/WHO_FCTC_english.pdf