

When is an image a health claim? A false-recollection method to detect implicit inferences about products' health benefits

Naomi A. Klepacz¹, Robert A. Nash^{2*}, M. Bernadette Egan¹, Charo E. Hodgkins¹, & Monique M. Raats¹

¹ School of Psychology, University of Surrey

² School of Life and Health Sciences, Aston University

Author note. This research was supported in part by funding from the European Commission's Seventh Framework Programme, within the CLYMBOL (Role of Health-Related Claims and Symbols in Consumer Behaviour; www.clymbol.eu) and PlantLIBRA (Plant food supplements: Levels of Intake, Benefit and Risk Assessment; www.plantlibra.eu) projects (Contract N°s: 311963 and 245199). The authors are grateful to project colleagues for their input on study materials; in particular to Azucena Gracia, Andrea Groeppel-Klein, Jure Pohar, and Evelien van de Veer for their assistance with translation in Experiment 3, and Erica van Herpen for her valuable comments on an earlier draft of this paper. The European Commission had no role in the experimental design, analysis or writing of this paper. The content of the paper reflects only the views of the authors; the European Commission is not liable for any use that may be made of the information contained in this paper. N.A.K., M.B.E., C.E.H. and M.M.R.'s research center provided consultancy to, and received travel funds to present research results from, organisations supported by food and drink companies. R.A.N. has no conflicts of interest to declare.

***Corresponding author**

Robert A. Nash

School of Life and Health Sciences

Aston University

Birmingham, B4 7ET

Tel: +44 (0)121 204 4522

R.Nash1@aston.ac.uk

Objective: Images on food and dietary supplement packaging might lead people to infer (appropriate or inappropriate) health benefits of those products. Research on this issue largely involves direct questions, which could (a) elicit inferences that would not be made unprompted, and (b) fail to capture inferences made implicitly. Using a novel memory-based method, the present research explored whether packaging imagery elicits health inferences without prompting, and the extent to which these inferences are made implicitly.

Methods: In three experiments, participants saw fictional product packages accompanied by written claims. Some packages contained an image that implied a health-related function (e.g., a brain), and some contained no image. Participants studied these packages and claims, and subsequently their memory for seen and unseen claims was tested.

Results: When a health image featured on a package, participants often subsequently recognized health claims that—despite being implied by the image—were not truly presented. In Experiment 2, these recognition errors persisted despite an explicit warning against treating the images as informative. In Experiment 3 these findings were replicated in a large consumer sample from five European countries, and with a cued recall test.

Conclusions: These findings confirm that images can act as health claims, by leading people to infer health benefits without prompting. These inferences appear to often be implicit, and could therefore be highly pervasive. The data underscore the importance of regulating imagery on product packaging; memory-based methods represent innovative ways to measure how leading (or misleading) are specific images.

Keywords: health claims; imagery; memory; false recognition; inference

What should consumers infer about the health benefits of a breakfast cereal, whose packaging pictures it inside a heart-shaped bowl? Or about a new dietary supplement, whose packaging bears a symbol of a human brain? Regulators and marketers have debated the likelihood that such imagery influences consumers' product understanding, yet there remains minimal empirical evidence on this issue. Here we report a novel method for assessing whether imagery on products' packaging leads consumers to draw implicit and explicit inferences about health benefits.

The role of front-of-pack imagery in shaping consumers' cognition about health is a timely and important concern. In recent years, legislation has been developed in many countries that guides manufacturers and marketers on the exact health claims that they are—and are not—authorized to make about products (e.g., European Commission, 2006). Importantly, this legislation applies to images as well as to text, because it assumes that images can lead consumers to make health-related inferences about the product inside the package (Wartella, Lichtenstein, Yaktine, & Nathan, 2011). But how valid is this assumption? Do consumers, consciously or unconsciously, treat packaging imagery as offering informational value? Clearly, the answer would be of direct legislative importance, and would shine a light on broader questions about people's understanding of health-related information. However, different methods for answering this question offer different kinds of information.

Measuring inferences

One obvious way to determine whether people make inferences about products is to ask them. Many studies adopt this approach through direct questioning methods such qualitative interviews and questionnaires (e.g., Lähteenmäki et al., 2010; Wansink, 2003; see Leathwood, Richardson, Sträter, Todd, & van Trijp, 2007; Williams, 2005 for overviews of methods). For instance, Grunert, Scholderer, and

Rogeaux (2011) showed participants a commercial for a probiotic product, then asked “After seeing this pack and commercial, if you had to tell a friend what [the product] does, what would you say?” (p. 270). Participants’ responses were content analyzed, to capture whether their understanding of the health claims accorded with the scientific basis for those claims. On health images in particular, Saba et al. (2010) showed that adding simple symbols to a product’s package—either a ‘natural’ symbol of a plant leaf, or a ‘medical’ cross symbol—led participants to rate the product as more healthy. Likewise, Carrillo, Fiszman, Lähteenmäki, and Varela (2014) showed participants images such as a person running toward the sun, some of which were taken from genuine food products. Using a word association task, the authors showed that participants often associated these images with general health concepts (e.g., well-being, healthy) and specific health functions (e.g., strength, good for the heart).

Studies such as these offer support for the notion that images lead people to infer health properties, but they also raise questions about the nature and origin of these inferences. Johnson-Laird (1982) differentiated two distinct types of inferences: explicit, and implicit. Explicit inferences, he argues, are made deliberately and consciously, by systematically considering and evaluating the available evidence. Implicit inferences, by contrast, are made spontaneously and without conscious awareness, and often go beyond the available evidence. Knowing which types of inferences are evoked by packaging imagery is important, not least because errors based on false implicit inferences are notoriously difficult to counter with corrective information (e.g., Guillory & Geracy, 2010).

In all of the above studies, participants were asked directly about their beliefs and inferences. Whereas this approach offers valuable insights, it also suffers some limitations. First, questioning people about their inferences is often likely to be what

actually prompts them to draw those inferences. These methods might therefore overestimate the influence of packaging imagery, because the average consumer might rarely infer information from this imagery unless they are overtly prompted to do so. A second reason is that direct questioning relies on people having conscious awareness of—and thus being able to report—their cognitions (Sheeran, Gollwitzer, & Bargh, 2013). This reliance is not always warranted: myriad studies show that persuasive influence can occur without the recipient's conscious awareness (Fitzsimmons et al., 2002; Strahan, Spencer, & Zanna, 2002), and that implicit associations can influence people's behavior despite going unreported in explicit tests (Maison, Greenwald, & Bruin, 2004). These studies might therefore *underestimate* the influence of packaging imagery, by failing to adequately capture participants' implicit inferences. To better capture such inferences, one needs different methods.

Memory as a window on implicit inferences

Because one characteristic of implicit inferences is that they frequently go beyond the available evidence (Johnson-Laird, 1982), people's implicit inferences often lead them to remember information that they spontaneously inferred but never truly saw (e.g., Barclay, 1973; Brewer, 1977). Memory measures can therefore offer valuable information about people's implicit and explicit inferences, without directly asking them to form or to report such inferences. Indeed, advertising researchers have recognized memory errors as a useful index of the subtle effects of persuasive influence (Braun-LaTour & Zaltman, 1998), and such indices might illuminate the issue of how packaging imagery affects consumers' inferences.

Relevant to the present context, memory errors arise not only when people make inferences from text, but also from images (e.g., Henkel, 2012). For instance, Garry, Strange, Bernstein, and Kinzett's (2007) participants read a newspaper article

about a devastating hurricane, which was accompanied by a photo of a village taken either before or after the hurricane struck. On a subsequent memory test, those in the ‘after’ condition were more likely to remember reading about injuries and deaths, even though no such detail was evidenced in the article or in the photo. According to the source monitoring framework (Johnson, Hashtroudi, & Lindsay, 1993), inferences can distort memory because they promote thoughts and mental images which, when later retrieved, feel like memories of real experiences. For instance, if a product’s packaging makes a person think about heart health, then when they later attempt to recall the claims they saw, claims about heart function should come to mind easily and clearly, and feel familiar. These memory-like characteristics might then lead them to believe they saw those claims before, rather than having only thought about them.

The present research applied a novel memory-based method to explore how packaging imagery affects people’s inferences about health benefits. Participants viewed fictional dietary supplements (Experiments 1 and 2) or food packages (Experiment 3) with accompanying written claims, and were subsequently tested on their ability to recognize which claims they saw (and to recall the claims, in Experiment 3). We predicted that if images lead participants to make health-related inferences, then participants would falsely recognize unseen claims that follow from those inferences. Furthermore, across these studies we took several approaches to gather information on the extent to which these inferences were implicit vs. explicit.

Experiment 1

Method

The protocol for all studies described in this paper was approved by the university’s Ethics Committee, and participants gave informed consent to participate.

Participants. A total of 36 students at a UK university (31 females and 5 males, mean age = 19.60, $SD = 1.74$, range = 18-28) took part either in exchange for course credit or without compensation.

Materials. In the present study, participants saw a series of product package images, alongside which they read a number of claims for each product and were later tested on their recognition of these claims. To this end, we developed a stimulus set of package images and written claims.¹

Supplement packages. To begin, we obtained digital images of front-of-pack labels for genuine but little-known dietary supplements, representing six health categories (women's health; memory and cognitive function; sleep; bones and joints; cold and flu; heart function). Using Photoshop, we digitally copied various elements of each package including designs and symbols. For each of the six health categories, we combined these elements to create a fictional dietary supplement package comprising a fictional brand name, a visually-appealing design, a brief description of the active ingredient, some peripheral text (e.g., the number of capsules inside the packet), and generic symbols (e.g., a manufacturer's logo stating "Established 1828").

For each fictional product, we also identified an image from the packaging of one genuine dietary supplement that represented a particular health-function (i.e., rather than depicting the product itself). For example, to represent 'sleep' an image of a moon surrounded by clouds was obtained from the label of a supplement for treating sleep problems. For each stimulus package, we created a secondary version onto which the appropriate function image was added. That is to say, we created two versions of each package, one with the function image present (hereafter, the image-present packages), and one without the image (the image-absent packages).

¹ The images used in these studies can be obtained from the corresponding author

Written claims. To accompany each of the six fictional products, we developed a stimulus set of 14 written claims. Six claims for each product made reference to some detail on the package (e.g., “Contains 150mg of valerian root extract”), but never referred to the health-function that was implied by the image on image-present packages. These six claims were the ‘related claims’. A further six claims in each set were ‘unrelated claims’, which made no reference to any aspect of the package or the image (e.g., “Provides nutritional support for those aged 50 years+”). Together the 12 related and unrelated claims will be referred to as the *noncritical* claims. The final two claims were our ‘critical’ claims; that is, those of central interest to our hypotheses. These claims made direct reference to the health function implied by the image on image-present packages (e.g., “Supports mental performance and cognitive function”), and are listed in Table S1 (supplementary material). All claims were found on the packaging of genuine dietary supplements, but adapted to ensure approximate consistency in length.

A small pilot study verified our assumption that the function images would shape people’s explicit inferences about these products when they were prompted to draw such inferences. Specifically, 25 undergraduate students saw the packages and rated the likelihood that various claims were true of those products (1 = Entirely false; 7 = entirely true). They judged the critical claims as more likely to be true when the function images were present on the packages ($M = 4.98$, $SD = 0.92$) than when they were absent ($M = 3.77$, $SD = 0.74$), $t(24) = 7.06$, $p < .001$, $d = 1.45$.

Design and procedure. This was a single-session laboratory experiment with two phases: encoding and recognition (Figure S1 in the online supplementary materials summarizes the procedure). The dependent variable was the proportion of critical claims that participants claimed to recognize. The independent variable was

the presence vs. absence of a function image on each package, manipulated within-subjects. Because our hypotheses relate only to (false) recognition of the critical claims, our analyses focus on these rather than on recognition of noncritical claims.

Encoding phase. Participants sat at a computer for the entire experimental session, and after consenting to participate they received written instructions on the screen. To begin, a random exemplar of the dietary supplement packages appeared on the screen for 10 s. Participants were simply asked to study the package during this time, and told that they would then see a series of claims presented underneath the package. Next, eight written claims supposedly pertaining to this product (4 related, 4 unrelated) appeared sequentially and in a random order underneath the package. Each claim was displayed for 4 s before being replaced by the next, and participants were asked to study these claims carefully. After the eighth written claim disappeared, participants completed a 1-min filler task – solving anagrams on a separate worksheet.

After 1 min, participants were instructed to refocus on the screen. At this point a new supplement package appeared, and the encoding procedure described above was repeated for the remaining five packages. During this phase, each participant saw three image-present packages and three image-absent packages in a random order; the assignment of packages to image condition was counterbalanced across participants.

Recognition phase. After the encoding phase was complete, the recognition phase began. On-screen instructions explained that participants would now see the same products again, paired with more claims. They were told that some of these would be pairs they saw previously, and some they did not see. Participants were once again shown the same six packages that they saw in the encoding phase, one by one and in a random order. This time, each package was again presented alongside a series of written claims. As each claim appeared, participants were asked “Have you

previously seen this product and claim pair?”. Participants responded by pressing the ‘Y’ key for ‘yes’ and ‘N’ for ‘no’. Whenever participants pressed ‘Y’, indicating that they had seen the claim paired with the particular product, they were asked to make a Remember/Know/Guess (R/K/G) judgment (Gardiner, Ramponi, & Richardson-Klavehn, 1998). Specifically, they pressed ‘1’ if they actually remembered seeing the claim paired with that product; ‘2’ if they knew the claim was paired with that product but did not explicitly remember any details; or ‘3’ if they were guessing.

Recognition judgments were made for 10 health claims paired with each of the six products (i.e., 60 judgments in total). Four of these were noncritical ‘related’ claims—two old (i.e., genuinely seen during encoding) and two new (i.e., not seen during encoding). Four were noncritical unrelated claims, again two old and two new. The remaining two were our critical claims, that is, those that were never presented during encoding, but were related to the function image that was included on image-present packages. All 10 claims were presented in a random order for each product.

On completion of the recognition phase, participants were shown the six function images in isolation from any packaging context, and asked whether they recognized any from a genuine product. This measure ensured that participants’ responses were not based on prior knowledge about the real products from which the images were taken. No participant recognized any of the images. Finally, participants recorded their age and gender before being thanked and debriefed.

Results

Recognition of critical claims. Our main interest was in the extent to which adding function images to the packages led participants to believe they read health claims that were potentially implied by these images (i.e., critical claims). To this end, for each participant we calculated the proportion of critical claims that they falsely

recognized (proportions out of 6 per condition – 3 packages x 2 claims). We focus on parametric analyses here, however, given the small frequencies in some of these analyses, we confirmed the pattern and significance of all t -tests using nonparametric Wilcoxon tests. A paired t -test confirmed that participants were more than twice as likely to falsely recognize critical claims in the image-present ($M = .29, SD = .26$) compared to the image-absent condition ($M = .13, SD = .17$), $t(35) = 3.54, p = .001, d = 0.61$, 95% CI of the difference in means [.07, .25]. In short, when an image that implied a health function was present on a product package, this frequently led participants to falsely recognize health claims consistent with the implied function.

Subjective judgments for critical claims. We reasoned that if the overall main effect of image were driven by explicit inferences, then it should be attributable to increases in ‘guess’ responding. However, if the effect were driven by implicit inferences, then it should be attributable to increases in ‘remember’ and/or ‘know’ responding. To address this issue, we calculated the proportion of remember, know, and guess responses given by each participant to critical claims in both the image-absent and image-present conditions. The effect of image on participants’ false recognition of critical claims was not driven significantly by increases in guessing ($M_{\text{present}} = .10, SD = .15; M_{\text{absent}} = .05, SD = .10$), $t(35) = 1.68, p = .10, d = 0.28$, 95% CI of the difference in means [-.01, .11]. We found that the presence of a function image increased the proportion of know responses ($M_{\text{present}} = .15, SD = .18; M_{\text{absent}} = .06, SD = .13$), $t(35) = 2.62, p = .01, d = 0.45$, 95% CI of the difference in means [.02, .16], but did not significantly increase the proportion of remember responses ($M_{\text{present}} = .05, SD = .10; M_{\text{absent}} = .02, SD = .07$), $t(35) = 1.22, p = .23, d = 0.26$, 95% CI of the difference in means [-.02, .06].

Recognition of noncritical claims. The subjective judgment data refute the possibility that the false recognition effects were driven purely by patterns of guessing, suggesting that implicit inferences played a primary role. Yet one counter-explanation is that participants were reluctant to admit too often that they were guessing, and so reported many of their guesses as ‘know’ or ‘remember’ responses. Recognition accuracy data for the noncritical claims help to tackle this explanation. If accuracy for these claims were at chance levels (i.e., 50%), this would indicate that participants remembered the claims very poorly, and were therefore adopting a pure guessing strategy even if not admitting so. In this study, recognition accuracy for noncritical claims was 73%, indicating that participants remembered the claims reasonably well and were not consistently guessing.

Experiment 2

The findings of Experiment 1 suggest that function images on product packaging can, in some cases, prime people to draw inferences about the health properties of those products. Our R/K/G data suggest that these inferences were largely implicit, insofar as the effect was not solely driven by guessing. However, to more confidently conclude that participants’ inferences were indeed largely implicit, this assumption can be addressed more directly. In Experiment 2 we did so through the use of an explicit warning.

In social and cognitive psychology, researchers have used warnings to test whether people can consciously avoid being persuaded or misled, and thus to infer how automatic (implicit) or controlled (explicit) are the processes underlying social influence. For instance, Petty and Cacioppo (1977) found that participants were less easily persuaded if they had been forewarned prior to receiving a persuasive message, suggesting that they had a degree of conscious control over this social influence. The

authors showed that a forewarning motivated participants to consider more fully their own positions, thus generating arguments to defend against the impending influence attempt. The effects of warnings on people's resistance to persuasion are mirrored in studies that explore people's resistance to making memory errors (Nash, Wheeler, & Hope, 2015). For instance, prior to giving participants misinformation about a witnessed event, Greene, Flynn, and Loftus (1982), warned them that the information they were about to receive 'may be inaccurate'. This warning significantly reduced subsequent memory errors, again indicating a degree of conscious control.

The aims of Experiment 2 were twofold. The first aim was to test the replicability of the effect of function images on false recognition, shown in Experiment 1. The second aim was to examine the extent to which forewarning participants—instructing them to avoid being influenced by the function images—would assist them in avoiding these recognition errors. This approach would provide further evidence on the extent to which implicit vs. explicit inferences drive these errors. If the recognition errors were due solely to explicit inferences, then people should resist making these errors when warned to treat function images as uninformative. In contrast, if the errors were at least partly due to implicit inferences, then the effect of function images should persist despite the warning.

Method

Participants. A total of 54 students from a UK university (44 females and 10 males, mean age = 20.15, $SD = 2.88$, range = 18-34) took part, either in exchange for course credit or without compensation, and were randomly allocated to either the warning or the no-warning group. A priori power analysis indicated that this sample size would permit detection of a medium-sized interaction effect (Cohen's $f = .25$,

given $\alpha = .05$, power = .80, and correlation between repeated measures of $r = .20$, approximated from the Experiment 1 data).

Materials. We used the same materials as in Experiment 1.

Design and Procedure. This study was identical to Experiment 1 except that half of participants received an explicit warning prior to the encoding phase. The warning was given in the written instructions and also verbally by the experimenter:

Some of the labels you will see have pictures or symbols on them.

These pictures and symbols have been placed onto the labels entirely at random. Because of this randomness, the pictures and symbols you will see actually provide no information about the products' real function.

All participants who received the warning confirmed that they understood it before commencing the study. For those in the no-warning group, the study was identical to Experiment 1. Again, no participants reported at the end of the study that they recognized any of the function images from genuine products.

Results

Recognition of critical claims. As in Experiment 1, we began by calculating for each participant the proportion of critical claims falsely recognized. The bottom section of Table 1 shows these proportions. All of the main effects described below were mirrored in nonparametric Wilcoxon tests, which reached identical conclusions. A 2 (image-present vs. image-absent) x 2 (warning vs. no-warning) mixed-factor ANOVA confirmed the overall finding of Experiment 1 via a significant main effect of image. Specifically, participants falsely recognized a greater proportion of critical claims in the image-present than in the image-absent condition, $F(1, 52) = 8.87$, $p < .01$, $\eta^2_p = .15$, 95% CI of the difference in means [.03, .18]. Of central importance to this study, we next examined whether the warning enabled participants to avoid

making recognition errors. These analyses suggested not – the main effect of warning was not significant, $F(1, 52) = 0.09, p = .77, \eta^2_p < .01$, 95% CI of the difference in means $[-.09, .12]$, and nor was the interaction, $F(1, 52) = 2.12, p = .15, \eta^2_p = .04$.

[TABLE 1 ABOUT HERE]

Subjective judgments for critical claims. We next looked to participants' R/K/G data, conducting new mixed-factor ANOVAs separately for remember, know, and guess responses respectively (see upper sections of Table 1). First, the main effects of image in these analyses confirmed that these data replicated the findings of Experiment 1: function images significantly increased the proportion of know responses, $F(1, 52) = 8.61, p < .01, \eta^2_p = .14$, 95% CI of the difference in means $[.02, .12]$, without significant increases in guessing and remembering (Guess, $F(1, 52) = 0.59, p = .45, \eta^2_p = .01$, 95% CI of the difference in means $[-.03, .06]$; Remember, $F(1, 52) = 2.35, p = .13, \eta^2_p = .04$, 95% CI of the difference in means $[-.01, .05]$.

The interaction effects—which index the effect of warnings on each type of subjective judgment—were more revealing. We found that the warning had almost no effect on participants' remember or know responses for critical claims (Remember, $F(1, 52) = 0.05, p = .83, \eta^2_p < .01$; Know, $F(1, 52) = 0.02, p = .90, \eta^2_p < .001$), therefore indicating further evidence of implicit inferences. However, the warning did moderate the effect of images on the proportion of guess responses, $F(1, 52) = 6.76, p = .01, \eta^2_p = .12$. In sum, the small and nonsignificant drop in recognition errors as a result of the warning appears, if anything, to reflect strategic shifts in guessing rather than a reduction in confident errors. This result may suggest that explicit inferences added to the effect of function images, albeit cannot explain the entire effect.

Recognition of noncritical claims. Participants' overall recognition accuracy for noncritical claims was 73%, identical to Experiment 1, again supporting the notion that participants generally had a reasonably good memory for the claims they saw.

Experiment 3

The data from Experiments 1 and 2 indicate that function images on product packaging can lead people to infer health claims, and that these inferences occur at least partly outside of conscious control. In our final experiment we had three further aims. The first was to see whether these effects extend to a recall task – that is, when people attempt to reconstruct the claims from memory, rather than simply making old/new judgments. If consumers really do form these implicit or explicit inferences without prompting, then they should generate the false information themselves rather than only affirming false information that is suggested to them at test. Therefore, if the effects transpire in recall as well as recognition memory, then this would add greater weight to the conclusion that function images promote unprompted inferences about health, as well as to the external validity and robustness of the effects.

The second aim of Experiment 3 was to compare the effect of function images on memory across different types of claims. Legislation distinguishes between nutrition claims—which simply make reference to ingredients or constituents of a product—and health claims, which make assertions about a specific health benefit provided through consuming that product or its constituents (European Commission, 2006). We predicted that when products carry health claims, the presence of function images would facilitate correct recollection of these claims. However, we also predicted that when products carry nutrition claims, function images would increase the likelihood that these claims would be 'upgraded' to health claims in memory. In other words, people would use the image to infer how the particular ingredient or

constituent should benefit them. In contrast, we predicted a more modest effect of images when products only carry generic claims with no reference to ingredients.

The final aim was to extend our findings to a more diverse and representative participant sample. To this end instead of sampling only university students, we sampled community members across a range of ages in five European countries.

Method

Participants. A total of 410 participants were recruited using a UK-based online panel provider, and completed the study in full. Due to a technical error, data for 38 participants were improperly recorded and so the final sample comprised 372 participants (185 females and 187 males, mean age = 45.07, $SD = 14.53$, range = 18-75), each of whom was remunerated in the form of points that could be converted to cash. All participants were permanent residents of the UK ($N = 81$), Germany ($N = 79$), Netherlands ($N = 71$), Slovenia ($N = 71$), or Spain ($N = 70$), and within each nationality we obtained a stratified sample of males and females across a breadth of age groups. Participants completed the study in their own language.

Materials. We developed a new stimulus set of fictional product packages and associated claims.

Food packages. To begin, we chose six health functions, and selected two foodstuffs containing nutrients that support each of those functions (e.g., two foods containing vitamin D for healthy teeth; see Table S2 in supplementary materials for a full list). For each of these 12 foodstuffs we designed a fictional product package, taking generic packaging images from the Internet, and using Photoshop to remove much of the detail from these to create basic templates. To each product we added a fictional brand name, a description of the foodstuff (e.g., “wholegrain bread”), a picture of the foodstuff, and other generic information (e.g., the product’s weight).

Next, we identified images that would represent each of the six health functions. For example, to represent ‘muscles and energy,’ we chose an image of a running man. These images were taken from the packaging of foodstuffs available for purchase in at least one of the countries in the sample. For each of the 12 foodstuffs we created a parallel version of the package onto which the appropriate function image was digitally added (the same image was added to both foodstuffs representing each of the six health-function categories). These two versions of each package constitute the basis of our ‘image-present’ versus ‘image-absent’ manipulation.

Health claims on packages. Rather than presenting written claims alongside the packages as in Experiments 1 and 2, this time we also added a written claim to each food package. Specifically, we added one claim representing one of three different claim-types: either a generic claim (e.g., “Fantastic new taste”), a nutrition claim (e.g., “Source of Zinc”), or a health claim (e.g., “Zinc contributes to normal cognitive function”). The nutrition and health claims were selected and worded based on the guidance and approved claims in the ‘EU Register of Nutrition and Health Claims Made on Food’ (European Commission, 2013). The generic claims mostly pertained to taste and flavour, mirroring the types of non-regulated generic claims typically displayed on foodstuff packaging in the sampled countries (see Table S2).

In sum, for each of the 12 foodstuffs we created six different versions of the package (see Figure 1 for an example). Finally, we created further parallel versions of every package image, translating all text from English into Dutch, German, Slovenian, and Spanish. The entire stimulus set therefore comprised 360 different packages [6 different health functions x 2 different food exemplars x 2 (image-present vs. image-absent) x 3 (generic vs. nutrition vs. health claim) x 5 languages].

[FIGURE 1 ABOUT HERE]

Design and procedure. Participants completed the study online and received on-screen instructions in their own language. The study comprised two phases: an encoding phase, and a recall/recognition phase. This study used a 2 (image-present vs. image-absent) x 3 (generic vs. nutrition vs. health claim) within-subject design.

Encoding phase. To begin, a randomly selected food package appeared on the participant's screen for 20 s. Participants were asked to study the package during this time and to remember as much as possible about it. Next, the package disappeared and was replaced with another package for a further 20 s. This process was repeated until the participant had seen one variant of all 12 foodstuffs. All foods were displayed in a random order; every participant saw six image-present products and six image-absent products, and within each of these image conditions they saw two with a generic claim, two with a nutrition claim, and two with a health claim. The assignment of products to image and claim-type conditions was fully counterbalanced across participants. After seeing all 12 products, participants solved logic puzzles on the screen for 3 minutes, after which they were moved on automatically.

Recall/recognition phase. After the encoding phase was complete, the recall phase began. On-screen instructions outlined the task for participants in their own language. Participants were once again sequentially shown the same 12 packages that they saw in the encoding phase, in a new random order. The only difference was that this time, a black panel obscured the written claim on each package. For each product, participants were asked to recall as closely as possible the claim that had appeared on the package, and to type this into a text-box provided. After submitting their response for the first of 12 packages, participants were next shown a list of six claims, from which they were asked to select the one they had seen. These six claims included the corresponding generic, nutrition, and health claims that we used in the encoding phase

(for each participant, one of these three would be the correct answer), plus three fillers (e.g., “Suitable for vegetarians”), all presented in a random order. After completing both the recall and recognition tasks for the first package, participants repeated this process for the remaining 11 packages. Finally, participants recorded their age and gender before being provided a written debriefing and the researchers’ contact details.

Results

Coding of recall data. Given that the recall task was very difficult, we expected that participants would rarely if ever recall the claims verbatim. Rather than focusing on absolute memory accuracy, then, we instead coded each response in terms of its level of specificity; that is, whether the participant recalled each claim as a generic claim, nutrition claim, or health claim. To this end, our team first translated every response, blind to condition, into English. Next, one researcher coded each response, again blind to condition, as a generic, nutrition, or health claim, or as an omission. Responses coded as omissions were those indicating “don’t know”, or that were otherwise irrelevant to the task. Across the sample, participants gave claim-like responses (i.e., not omissions) in 71.1% of cases. A second researcher also blind to condition coded 22% of responses; inter-rater agreement was 98.9% (Cohen’s $\kappa = .99$), therefore we used the first researcher’s coding for analyses of recall.

Recall analysis. Overall, 58.9% of responses (excluding omissions) were at the correct level of specificity, significantly above the 33.3% chance-level. We were especially interested in whether the presence of function images would lead people to recall claims (correctly or incorrectly) as health claims. To address this question, we calculated the proportion of each claim-type that were recalled as health claims; these data are illustrated in the left-half of Figure 2. A 2 (image-present vs. image-absent) x 3 (claim-type) repeated-measures ANOVA revealed that the presence of a function

image did indeed lead to a significant increase in recalling claims as health claims, $F(1, 371) = 30.50, p < .001, \eta^2_p = .08$, 95% CI of the difference in means [.04, .08]. However, we also found a significant interaction effect, $F(2, 742) = 3.30, p = .04, \eta^2_p = .01$. As predicted, follow-up t -tests showed that function images significantly increased the (appropriate) recall of health claims as health claims, $t(371) = 3.85, p < .001$. Images also significantly increased the (false) recall of nutrition claims as health claims, $t(371) = 4.86, p < .001$. However, images did not significantly increase the (false) recall of generic claims as health claims, $t(371) = 1.58, p = .12$.²

If these recall effects were solely due to participants drawing explicit inferences in the absence of memory (i.e., educated guessing), then arguably they should occur primarily among those participants who performed worst on the recall test. To test this possibility, we calculated the total number of claims that each participant recalled at the correct level of specificity (excluding omissions), and used these scores to conduct a median split of the dataset. The main effect of function images was comparable in the low-accuracy ($p < .001, \eta^2_p = .09$) and high-accuracy groups ($p < .001, \eta^2_p = .07$), thus suggesting that implicit inferences played a key role.

Recognition analysis. We finished by conducting a 2 (image-present vs. image-absent) x 3 (claim-type) repeated-measures ANOVA, examining the proportion of cases in which participants (correctly or incorrectly) picked the health claim from the list of six options as the one they saw on the package. As the right-half of Figure 2 shows, once again these data revealed a main effect of image, $F(1, 371) = 29.15, p < .001, \eta^2_p = .07$, 95% CI of the difference in means [.04, .09] – with the function

² For both the recall and recognition data, we also ran analyses that included Country as a between-subjects factor. No main effect or interaction involving this variable was significant (all $p > .13$), therefore the analyses reported here collapse data across the five countries.

image present, people were more likely to choose the health claim. There was no significant image x claim-type interaction, $F(2, 742) = 1.23, p = .29, \eta^2_p < .01$. Unlike in Experiments 1 and 2, participants here were actually presented with the correct answer as one of the multiple-choice options; in the generic and nutrition claim conditions, function images therefore led participants away from this correct answer. The main effect of image held in the low-accuracy ($p = .001, \eta^2_p = .07$) and high-accuracy ($p < .001, \eta^2_p = .08$) groups, indicating that it was not due solely to guessing.

[FIGURE 2 ABOUT HERE]

Discussion

Together, the results of these studies lend good support for the notion that packaging imagery can indeed lead people to infer health properties of products. In all three experiments, participants falsely recognized health claims that they had not truly read, and these recognition errors increased significantly when images that imply specific health functions were present on the products' packaging.

In showing that images affect people's inferences about health, our data fit with the small empirical literature that has used more direct methods (e.g., Carrillo et al., 2014; Saba et al., 2010). However, by virtue of using a memory task that involved no direct questioning, the data extend those prior findings in several ways. First, it is clear that these inferences occur without direct prompting. Second, and of particular importance, several elements of the data indicate that participants' inferences from the images were often implicit, occurring spontaneously and outside of conscious control rather than through deliberate reasoning. In Experiments 1 and 2 we showed that the errors were not due solely to educated guesswork; rather, participants were confident that they read these claims, most frequently making errors with 'know' rather than 'guess' responses. In Experiment 2 the errors were not prevented when participants

were explicitly warned that the images were meaningless. In Experiment 3, those participants whose recall-test responses were more accurate—suggesting a lesser tendency to guess—were just as influenced by function images as those participants who were less accurate. Together, these findings indicate that participants were not always aware of, nor able to avoid, forming inferences when viewing the products.

Beyond the theoretical implications of these findings, we propose that our novel experimental paradigm represents an example of how memory-based methods might help manufacturers and regulators quantify the extent to which specific packages and advertisements lead or mislead consumers. The importance of this quantification is especially salient when considering the minimal effect of forewarning participants, because this finding indicates that interventions involving consumer education alone might not necessarily offer immunity to the misleading and suggestive power of images. Methods such as ours are undoubtedly less straightforward than is direct questioning, yet it is clear that the latter approach can underestimate the implicit potency of imagery. The value of uncovering implicit inferences may depend on the likelihood that they would influence people's behavior; indeed, one question not addressed here is whether images lead consumers only to infer that they saw health claims, or also to actually believe those claims. The latter type of belief, we presume, would be an important precondition if the inferences were to translate into behavior, such as purchasing the product in reaction to specific health concerns. Future research should address this important question.

Whereas we have used terms such as 'misleading' and 'suggestive' to describe the potential effect of images, one might construe these findings differently. If, for instance, an image of bones led people to misremember "a source of calcium" (a nutrition claim) instead as "with calcium for strong bones" (a health claim), this error

might indicate that the image facilitates consumer understanding. In this sense the bone image serves an educational function—helping people understand and remember the function of calcium—which might otherwise be less well-served by a complex written claim (Wansink, Sonka, & Hasler, 2004). Indeed, in many cases marketers may use nutrition claims intentionally to lead consumers to infer health benefits. From a legislative perspective this educational function of packaging imagery, albeit positive at face value, could be troublesome. Approved health claims are often lengthy and unwieldy because they communicate nuance about the limits of scientific consensus and probable size of any benefit (to illustrate, consider the EU-approved health claim “*carbohydrate-electrolyte solution contributes to the maintenance of endurance performance during prolonged endurance exercise*”; European Commission, 2013). Images, in contrast, rarely communicate such nuance, can be nonspecific or ambiguous, and might therefore lead consumers to infer health benefits more numerous than or different from those supported by scientific evidence. This problem is potentially greater when images are used in tandem with nutrition claims rather than health claims, as the former afford greater scope for consumers to infer health benefits that are unfounded. Whether images educate or miseducate will undoubtedly often depend on broader contextual factors such as these.

Images offer many advantages to consumers and retailers—ease-of-processing, visual appeal, memorability, and so forth (Winkielman, Schwarz, Reber, & Fazendeiro, 2003)—but there is clearly much to learn about the diverse health inferences that specific images elicit, explicitly and implicitly, and about the potency of images relative to written claims. The spontaneity and persistence of participants’ recognition errors nevertheless implies that function imagery such as heart-shaped cereal bowls could have subtle yet pervasive effects on consumers’ cognition.

References

- Barclay, J. R. (1973). The role of comprehension in remembering sentences. *Cognitive Psychology, 4*, 229-254.
- Braun-Latour, K. A., & Zaltman, G. (2006). Memory change: An intimate measure of persuasion. *Journal of Advertising Research, 46*, 57-72.
- Brewer, W. F. (1977). Memory for the pragmatic implications of sentences. *Memory & Cognition, 5*, 673-678.
- Carrillo, E., Fiszman, S., Lähteenmäki, L., & Varela, P. (2014). Consumers' perception of symbols and health claims as health-related label messages. A cross-cultural study. *Food Research International, 62*, 653-661.
- European Commission (2006). Regulation (EC) No. 1924/2006 of the European Parliament and of the Council of 20 December 2006 on nutrition and health claims made on foods. *Official Journal of the European Union, L404*, 9–25.
- European Commission (2013). *EU register of nutrition and health claims made on food*. Retrieved from <http://ec.europa.eu/nuhclaims/>.
- Fitzsimmons, G. J., Hutchinson, J. W., Williams, P., Alba, J. W., Chartrand, T. L., Huber, J., et al. (2002). Non-conscious influences on consumer choice. *Marketing Letters, 13*, 269-279.
- Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (1998). Experiences of remembering, knowing, and guessing. *Consciousness and Cognition, 7*, 1-26.
- Garry, M., Strange, D., Bernstein, D. M., & Kinzett, T. (2007). Photographs can distort memory for the news. *Applied Cognitive Psychology, 21*, 995-1004.
- Greene, E., Flynn, M. S., & Loftus, E. F. (1982). Inducing resistance to misleading information. *Journal of Verbal Learning and Verbal Behavior, 21*, 207-219.

- Grunert, K. G., Scholderer, J., & Rogeaux, M. (2011). Determinants of consumer understanding of health claims. *Appetite, 56*, 269-277.
- Guillory, J. J., & Geraci, L. (2010). The persistence of inferences in memory for younger and older adults: Remembering facts and believing inferences. *Psychonomic Bulletin & Review, 17*, 73-81.
- Henkel, L. (2012). Seeing photos makes us read between the lines: The influence of photos on memory for inferences. *Quarterly Journal of Experimental Psychology, 65*, 773-795.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3-28.
- Johnson-Laird, P. N. (1982). Ninth Bartlett memorial lecture. Thinking as a skill. *Quarterly Journal of Experimental Psychology, 34A*, 1-29.
- Lähteenmäki, L., Lampila, P., Grunert, K., Boztug, Y., Ueland, Ø., Åström, A., & Martinsdóttir, E. (2010). Impact of health-related claims on the perception of other product attributes. *Food Policy, 35*, 230-239.
- Leathwood, P. D., Richardson, D. P., Sträter, P., Todd, P. M., & van Trijp, H. (2007). Consumer understanding of nutrition and health claims: sources of evidence. *British Journal of Nutrition, 98*, 474-484.
- Maison, D., Greenwald, A. G., & Bruin, R. H. (2004). Predictive validity of the Implicit Association Test in studies of brands, consumer attitudes, and behavior. *Journal of Consumer Psychology, 14*, 405-415.
- Nash, R. A., Wheeler, R. L., & Hope, L. (2015). On the persuadability of memory: Is changing people's memories no more than changing their minds? *British Journal of Psychology, 106*, 308-326.

- Petty, R. E., & Cacioppo, J. T. (1977). Forewarning, cognitive responding, and resistance to persuasion. *Journal of Personality and Social Psychology*, *35*, 645-655.
- Saba, A., Vassallo, M., Shepherd, R., Lampila, P., Arvola, A., Dean, M., [...] & Lähteenmäki, L. (2010). Country-wise differences in perception of health-related messages in cereal-based food products. *Food Quality and Preference*, *21*, 385-393.
- Sheeran, P., Gollwitzer, P. M., & Bargh, J. A. (2013). Nonconscious processes and health. *Health Psychology*, *32*, 460-473.
- Strahan, E. J., Spencer, S. J., & Zanna, M. P. (2002). Subliminal priming and persuasion: Striking while the iron is hot. *Journal of Experimental Social Psychology*, *38*, 556-568.
- Wansink, B. (2003). How do front and back package labels influence beliefs about health claims? *Journal of Consumer Affairs*, *37*, 305-316.
- Wansink, B., Sonka, S. T., & Hasler, C. M. (2004). Front-label health claims: when less is more. *Food Policy*, *29*, 659-667.
- Wartella, E. A., Lichtenstein, A. H., Yaktine, A., & Nathan, R. (Eds.) (2011). *Front-of-package nutrition rating systems and symbols: Promoting healthier choices*. Washington, DC: National Academies Press.
- Williams, P. (2005). Consumer understanding and use of health claims for foods. *Nutrition Reviews*, *63*, 256-264.
- Winkielman, P., Schwarz, N., Reber, R., & Fazendeiro, T. A. (2003). Cognitive and affective consequences of visual fluency: When seeing is easy on the mind. In L. M. Scott, & R. Batra (Eds.). *Persuasive imagery: A consumer response perspective* (pp.75-90). Mahwah, NJ: Lawrence Erlbaum.

Table 1. *Proportion of recognition errors made by participants in Experiment 2, overall and split by subjective R/K/G response (standard deviations in parentheses).*

Condition		Image-absent	Image-present
No-warning	Overall false recognition	.12 (.17)	.28 (.27)
	Remember	.02 (.05)	.04 (.07)
	Know	.04 (.09)	.11 (.17)
	Guess	.06 (.11)	.12 (.15)
Warning	Overall false recognition	.19 (.22)	.24 (.24)
	Remember	.02 (.07)	.04 (.08)
	Know	.03 (.14)	.10 (.17)
	Guess	.14 (.19)	.10 (.19)
Total	Overall false recognition	.15 (.20)	.26 (.25)
	Remember	.02 (.06)	.04 (.08)
	Know	.04 (.08)	.11 (.18)
	Guess	.10 (.16)	.15 (.20)



Figure 1. Examples of packaging used in Experiment 3, with generic claims (panels a and b), nutrition claims (c and d), and health claims (e and f). Exemplars in the left column represent the image-absent condition; those in the right column the image-present condition. In this case the corresponding function image is a stretching human figure with bone illustration.

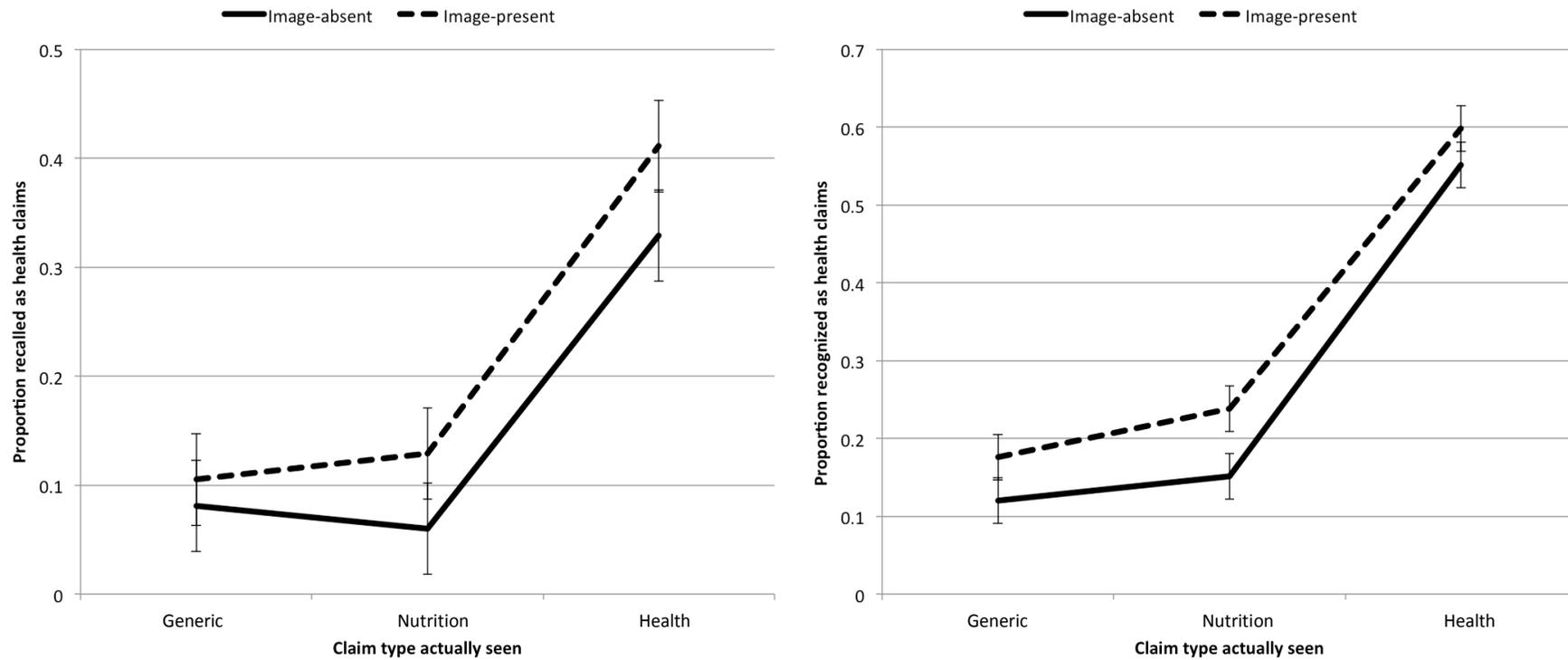


Figure 2. Proportion of cases in which participants in Experiment 3 recalled claims as being health claims (left panel), and chose the health claim from the recognition test (right panel), as a function of the actual claim-type seen, and function image presence vs. absence. Error bars are 95% within-subject confidence intervals.