ORIGINAL PAPER

When is Sustainability a Liability, and When Is It an Asset? Quality Inferences for Core and Peripheral Attributes

Siv Skard[1] · Sveinung Jørgensen[2] · Lars Jacob Tynes Pedersen[1]

Abstract

Sustainable products offered in today's marketplace are labelled with product-related green attributes (i.e. green *core* attributes) or non-product-related green attributes (i.e. green *peripheral* attributes). The current research investigates consumers' inferences about a product's functional quality when its core attributes are green (e.g. the ingredients) and when its peripheral attributes are green (e.g. the product packaging). Four experimental studies and an internal meta-analysis show that there is a sustainability liability effect in strength-dependent categories (for both core and peripheral attributes), and a sustainability asset effect in gentleness-dependent categories (for core attributes only). Our research contributes to the current understanding of how consumers make inferences about product quality when contemplating different types of green attributes. The findings have implications for how strength-dependent and gentleness-dependent products should be labelled as green.

Keywords Sustainability liability · Functional quality · Green product attributes

Introduction

In many markets, there is a rise of conscious consumption, characterized by individual consumers who select green products in order to minimize their negative impact on the environment (e.g., Sachdeva et al. 2015). In line with such global sustainable consumption trends, companies are increasingly launching green alternatives to their non-green counterparts (Olsen et al. 2014). Recent market research shows that as much as a third of all consumers claim to prefer sustainable brands (Unilever 2017). However, the gap between consumers' explicit attitudes towards sustainable products and their actual purchase behavior is still evident (Auger and Devinney 2007; Luchs et al. 2010), suggesting

that barriers against sustainable consumption prevail in many consumer decisions.

A key barrier against sustainable consumption is the perceived trade-off between sustainability and functional product quality (Luchs and Kumar 2017; Luchs et al. 2012). Prior research shows that for product categories where strength-related attributes are valued (henceforth referred to as strength-dependent product categories), consumers may prefer less sustainable products because they are perceived as more effective than sustainable alternatives (Lin and Chang 2012; Luchs et al. 2010; Pancer et al. 2017). This effect has become known as the 'sustainability liability effect' (Luchs et al. 2010).

A question that has not been addressed in prior research is whether this liability effect will occur for any type of green product attribute. Some green attributes are product-related (Keller 1993), which means that they are necessary for the product's core functions. For example, Unilever's detergent OMO EcoActive is labelled with the green claim '70% plant-based cleaning ingredients'. The plant-based ingredients can be defined as a product-related, or core, green attribute. Other green attributes are non-product-related, meaning that they are only peripherally linked to the core product. Coca Cola's 'PlantBottle' is an example of a product with a non-product-related green attribute. In this case, the beverage itself remains the same, and the sustainability is manifested

✉ Siv Skard
siv.skard@nhh.no

Sveinung Jørgensen
sveinung.jorgensen@inn.no

Lars Jacob Tynes Pedersen
lars.pedersen@nhh.no

[1] NHH Norwegian School of Economics, Helleveien 30, 5045 Bergen, Norway

[2] Inland Norway University of Applied Sciences, Gudbrandsdalsvegen 350, 2624 Lillehammer, Norway

in an attribute that is peripheral to the core product (i.e. the beverage). The question is whether consumers make similar quality inferences for core and peripheral green attributes.

Given that there is a perceived negative correlation between sustainability (i.e. the observable attribute) and quality (i.e. the unobservable attribute) in strength-dependent product categories, consumers may use lay theories to infer the level of quality according to its correlation with sustainability (Chernev and Carpenter 2001). In the current research, we argue that this type of process will make consumers infer lower quality of strength-dependent products even when the green attribute is non-product-related (e.g. packaging). However, since the inter-attribute correlation will be more salient when the green attribute is product-related (e.g. core ingredients), we predict a stronger impact on quality inferences for such attributes.

A second gap in the literature is the relationship between green product attribute information and quality inferences in categories where gentleness-related attributes (e.g. mild, soft, safe) are valued (henceforth referred to as gentleness-dependent product categories). Prior research shows that consumers prefer sustainable products in categories that appeal to gentleness (Luchs et al. 2010), but less is known about the quality inferences consumers make when they encounter green attributes in this category. As noted by Lin and Chang (2012), product effectiveness may be a valid concern also in categories in which non-strength-related attributes (e.g. gentleness) are important for consumer choice. Because consumers associate sustainability with gentleness-related attributes (Luchs et al. 2010), it is likely that they will use information about green attributes to infer higher functional product quality in categories where gentleness is valued (i.e. the sustainability asset effect). As for the liability effect, we argue that a green attribute that is directly (vs. peripherally) linked to the core functions of a gentleness-dependent product will lead to more positive quality inferences.

Based on the discussion above, we address the following research question in this paper: *Will consumers infer lower (higher) functional quality for strength-dependent (gentleness-dependent) products both when the green attribute is product-related (i.e. core) and non-product-related (i.e. peripheral)?* We address this question in four experimental studies testing sustainability effects in strength-dependent and gentleness-dependent categories when the green attribute is either core or peripheral.

The remainder of this paper is structured as follows. First, we account for the theoretical background. Second, we develop hypotheses. Thereafter, we sequentially report Study 1, 2, 3 and 4, before we report internal meta-analyses of the findings from the four studies. Finally, we discuss our findings and outline contributions and avenues for future research.

Theoretical Background

Sustainability and Quality Inferences

There is some evidence that a company's social responsibility profile may foster consumer beliefs that the brand is able to deliver functional benefits (Du et al. 2007). However, with regards to environmental sustainability, most consumer studies point to a trade-off evaluation between product sustainability and functional product quality, especially for utilitarian product categories (Luchs and Kumar 2017). Work by Luchs and colleagues (Luchs and Kumar 2017; Luchs et al. 2012) demonstrates that when consumers are presented with a choice situation with a trade-off between sustainability and functional performance, they tend to choose products with superior functional performance over products with superior sustainability characteristics. Other studies document a direct negative relationship between sustainability and functional quality perceptions. For example, using a projective technique, Luchs et al. (2010) demonstrate that sustainable products in strength-dependent categories are perceived as less durable than non-sustainable products are. Pancer et al. (2017) document that a single environmental packaging cue (the color green or an eco-label) may reduce perceived product efficacy. Wood et al. (2018) show that mainstream brands (but not their niche competitors) are perceived as less effective when presented with a green label. Lin and Chang (2012) demonstrate that consumers use a greater amount of environmentally friendly cleaning products (hand sanitizers and detergents) compared to regular cleaning products. This likely reflects an attempt to compensate for perceived lower product effectiveness.

The sustainability-quality trade-off has been explained by lay theories of consumer decision-making (Lin and Chang 2012; Newman et al. 2014). These theories suggest that trade-off evaluations are not likely to be based on actual knowledge about how the sacrifice of quality for the benefit of sustainability plays out in the production of a product. Rather, consumers use heuristics, or simple inferences, regarding the relationship between product sustainability and product quality. Some research argues that consumers use a zero-sum heuristic, assuming that a company's effort to make a product more sustainable means that the company must allocate resources away from product quality (Newman et al. 2014). However, this type of compensatory inference-making (Chernev and Carpenter 2001) can not explain why sustainability is an asset in gentleness-dependent categories (Luchs et al. 2010). If negative quality inferences for sustainable products is contingent on product category, consumers are more likely to apply a correlation-based inference strategy,

where the value of an unobservable attribute (e.g. quality) is inferred according to its perceived correlation with the observable attribute (e.g. sustainability). This strategy is referred to as "probabilistic consistency inferences" (Dick et al. 1990). For gentleness-dependent product categories, the inter-correlation between sustainability and quality is likely to be positive. Indeed, Luchs et al. (2010) show that there is a strong association between sustainability and gentleness-related attributes in the consumers' mind. Therefore, if using a probabilistic consistency strategy, consumers are likely to infer higher quality when facing sustainable product information in gentleness-dependent categories. In strength-dependent product categories, the inter-correlation between sustainability and quality is likely to be negative, and a probabilistic consistency strategy will therefore lead to negative quality inferences when consumers encounter information about product sustainability.

Type of Green Attribute: Quality Inferences for Core and Peripheral Green Attributes

There is a growing literature on the effectiveness of different types of sustainability labels, often referred to as 'eco-labelling' (e.g., Cho 2015; Cho and Baskin 2018; Gosselt et al. 2019; Pancer et al. 2017; Vanclay et al. 2011). However, research on how green labelling influences consumers' inferences about product quality is scarce. Apart from the studies on how sustainability cues in general might influence perceived product performance (Lin and Chang 2012; Luchs et al. 2010; Wood et al. 2018), there has been little attempt to theorize and test the effect of specific green labels on quality inferences.

Most studies on sustainability-quality trade-off effects use generic sustainability scores/descriptions to inform consumers about the product's sustainability characteristics (e.g., Luchs and Kumar 2017; Luchs et al. 2010, 2012, Newman et al. 2014) However, some studies use more specific green claims when testing trade-off effects. For example, Luchs et al. (2010) described the sustainability of the production methods and materials. Pancer et al. (2017) tested two different packaging cues (color and eco-label) on product quality perceptions. They showed that the cues in isolation had a negative effect on perceived product quality, but when used together, the negative effect was mitigated.

A product can be environmentally sustainable in many ways. For example, the production process can be sustainable, or the physical product attributes can be made of sustainable material. Concerning the latter, there is an important distinction between product-related and non-product-related attributes (Keller 1993). Whereas product-related attributes are the ingredients, or physical composition, of a product that are necessary for the product's core functions,

non-product-related attributes are peripheral to the core product and do not directly influence product performance (Keller 1993). Product packaging is an example of a non-product-related attribute.

The distinction between product-related and non-product-related attributes is relevant in a green labelling perspective for two main reasons. First, sustainable ingredients (i.e. a product-related attribute) and sustainable packaging (i.e. a non-product-related attribute) are two of the most common ways that packaged goods are labelled green. Second, the distinction is theoretically applicable due to its relationship with the functional performance of the products. Since extensive research shows that consumers are making trade-offs between quality and sustainability (Luchs and Kumar 2017), it is plausible that attributes that vary according to their link to functionality will have different effect on quality inferences. Based on rational arguments, one could expect that consumers will not conclude that a non-product-related green attribute, such as recycled packaging, would influence the functionality of the content of the product. For example, a drain opener contained by recycled plastic should not affect the drain opener's ability to unblock pipes. However, the aforementioned lay theories of consumer decision-making and inference-making would suggest that any green cue may evoke negative inferences about functional quality.

In the next sections, we outline predictions for how product-related (hereafter referred to as 'core') and non-product-related (hereafter referred to as 'peripheral') green attributes influence consumers' inferences about functional quality in strength-dependent and gentleness-dependent product categories. We use the following terminology throughout our presentation of hypotheses and results: 'Green core attribute' refers to products with green ingredients. 'Green peripheral attribute' refers to products with green packaging. 'No green attribute' refers to regular products with no green attributes in either ingredients or packaging.

Hypotheses

Quality Inferences in Strength-Dependent Product Categories

The issue of how consumers make inferences about unavailable or missing information has received much attention in existing consumer research (Chernev and Carpenter 2001). When consumers assume a causal relationship between the value of a missing attribute (e.g. quality) and the value of a known attribute (e.g. sustainability), they apply the so-called 'probabilistic consistency' strategy (Dick et al. 1990). Applied to the current study, we may expect that consumers infer a causal relationship between environmentally harmful ingredients (e.g. chemicals) and product performance

in strength-dependent categories. Although green innovations in the chemical industry make it possible to offer more sustainable ingredients in strength-dependent categories, the learned association between harmful chemicals and effectiveness is likely to dominate consumers' judgments. Therefore, information about a green core attribute will be perceived as incongruent with beliefs about functional quality and cause a negative quality inference effect.

A non-product-related attribute, on the other hand, is not directly related to the core function of the product. However, theory about consumer inference-making will nevertheless predict that consumers may use a green peripheral attribute to make negative inferences about functional product quality. Consumers may apply the aforementioned probabilistic consistency strategy, assuming a causal relationship between the observed attribute ("green" packaging) and the unobserved attribute (functional quality). This process may occur regardless of the missing link between the green attribute and the core product, since any green information is inconsistent with dominant expectations in the category (strong, environmentally harmful chemicals). Since most fast moving consumer goods (FMCG) decisions are made under low effort, consumers may not reflect on the lack of a true causal relationship between a peripheral attribute and the core functioning of the product.

An alternative process is the compensatory inference strategy, in which consumers use intuitive theories about market efficiency (Chernev and Carpenter 2001). When consumers encounter information about a green product benefit, they may infer that the company has diverted resources away from developing the product's quality (Newman et al. 2014). This mechanism may produce the same negative quality inference effect, regardless of whether the green attribute is core or peripheral, and regardless of how important strength-related attributes are in the category. However, compensatory inferences represent a more complex, two-staged process, which is likely to be superseded by the probabilistic consistency strategy when the attributes are correlated (Chernev and Carpenter 2001). Since there is likely to be a negative inter-attribute correlation between sustainability and functional quality in strength-dependent categories (see Luchs et al. 2010), we believe that consumers are more likely to use the probabilistic consistency strategy when inferring functional product quality for sustainable products. We expect that this type of inferences will occur for both core and peripheral green attributes, albeit to a lesser degree when the green attribute is only peripherally linked to the product's functionality (e.g. "green" packaging). A core green attribute is directly linked to the functionality of the product, creating a strong and salient inter-attribute correlation between sustainability and functionality. For the green peripheral attribute, there is no direct link between the green attribute and the functionality of the product, making

the inter-attribute correlation weaker and less salient. Some consumers may even be able to correct their heuristic judgments of green products as less effective when they realize that there is no link between a peripheral attribute and the core product. However, due to bounded rationality and the general human tendency to base judgments on stereotypes and heuristics, we expect that even a green peripheral attribute will make consumers infer lower functional quality.

Based on the discussions above, we propose the following hypotheses:

H1 In strength-dependent product categories, consumers infer lower functional product quality (a) when the product has a green core attribute (vs. no green attribute), and (b) when the product has a green peripheral attribute (vs. no green attribute)

H2 The sustainability liability effect on functional product quality will be stronger for the green core (vs green peripheral) attribute

Prior research suggests that consumers sometimes prefer sustainable products despite a perceived trade-off with functional performance. However, preference for a sustainable alternative when contemplating such a trade-off requires that the sustainable option meets a minimum threshold level of functional performance (Auger and Devinney 2007; Luchs et al. 2012). Choosing the sustainable alternative may be perceived as the morally superior option, and consumers may choose this in an attempt to reduce feelings of guilt (Luchs et al. 2012). Consumers who identify as environmentally conscious may be especially prone to choose sustainable alternatives, despite acknowledging a quality trade-off (Luchs et al. 2012). Moreover, when asked to make moral choices in a laboratory setting, there is a high risk of social desirability in responding. In fact, Luchs et al. (2010) found no liability effect when asking participants in lab studies to indicate their personal preferences, whereas projective questions asking participants to take other consumers' perspective revealed a significant liability effect. In a real choice setting, Luchs et al. (2010) found significant preferences for non-sustainable products, but only when participants felt unobserved, fueling the social desirability explanation.

In sum, prior research gives reasons to expect that under certain circumstances, consumers may prefer a sustainable option even when contemplating a quality trade-off. Nevertheless, the negative inferences about functional quality when presented with a green attribute (core or peripheral) are likely to influence product preference negatively. Therefore, we predict a negative indirect effect of sustainability on product preferences through perceived functional quality. In line with H1, we expect this indirect effect for both green core and green peripheral attributes.

H3 There is a negative indirect effect of (a) green core attribute and (b) green peripheral attribute (vs. no green attribute) on product preference through perceived functional product quality

Quality Inferences in Gentleness-Dependent Product Categories

Luchs et al. (2010) explain preferences for sustainable products in gentleness-dependent categories by the positive association consumers hold between sustainability and gentleness. A relevant question is whether this effect merely is a halo effect of such benefit congruity, or if consumers are actually inferring higher quality when gentleness-dependent products contain a green attribute. While several studies have aimed to study sustainability-quality trade-off-effects in strength-dependent product categories, less is known about possible positive quality inferences in gentleness-dependent product categories.

One possible inference process that would cause positive evaluations of perceived functional quality is the 'evaluative consistency strategy' (Chernev and Carpenter 2001). Applying this strategy, consumers would infer that a product that scores high on sustainability also would be superior on functional quality. In such a case, consumers would infer higher functional quality for sustainable products in both strength-dependent and gentleness-dependent categories. We contend that consumers will rather apply the correlation-based probabilistic inference strategy and form evaluations based on assumptions of a positive causal relationship between sustainability and functional quality in gentleness-dependent categories. In consumers' minds, sustainability is conceptually linked to important gentle product benefits (Luchs et al. 2010). We contend that this link is causal in nature, whereby consumers believe that a sustainable attribute increases the functional performance of the product. For example, a consumer may believe that natural ingredients in a baby shampoo improves the product's ability to make the child's hair clean.

When a green attribute is only peripherally linked to the key functions of the product, the assumed causal relationship between the two attributes will be weaker and less salient compared to when the green attribute is core. Some consumers may even be able to correct their stereotypical view of product greenness as a functional quality indicator in gentleness-dependent categories, and thus rely less on heuristic judgments. Nevertheless, the learned correlational relationship between sustainability and quality in gentleness-dependent product categories will make consumers use even peripheral green attributes to infer higher functional quality. Therefore, we formulate the following hypotheses about a quality asset effect for sustainability in gentle product categories:

H4 In gentleness-dependent product categories, consumers infer higher functional product quality (a) when the product has a core green attribute (vs. no green attribute), and (b) when the product has a peripheral green attribute (vs. no green attribute)

H5 The sustainability asset effect on functional product quality will be stronger for the green core (vs green peripheral) attribute

Product effectiveness is a relevant consideration also in categories where gentleness-related attributes are valued (Lin and Chang 2012). Therefore, the extent to which green attributes lead to positive quality inferences is likely to increase product preferences. It follows from this that perceived functional quality is a mediator mechanism between sustainability and product preferences also for gentleness-dependent product categories. Formally, we propose the following hypothesis:

H6 There is a positive indirect effect of (a) green core attribute and (b) green peripheral attribute (vs. no green attribute) on product preference through perceived functional product quality

In the following, we report four experimental studies. The first study tests our hypotheses on a student sample (N = 436), using an online survey-based experimental design. The second study is a framed field-experiment testing the hypotheses on a more representative consumer sample (N = 181). This design allowed us to use physical products and capture perceived functional quality through a practical measurement task. The third study (N = 164) only tests the hypotheses pertaining to the gentleness-dependent category (H4–H6). This study was conducted to test an alternative labelling strategy due to some unexpected findings in the two first studies. Finally, while the three first studies test type of green attribute (core and peripheral) as a within-subjects factor, the fourth study (N = 407) replicates the findings in a between-subjects design. Across all studies, we prioritized a large sample size to be able to identify a true effect. Large replication projects in social sciences have failed to demonstrate as large effect sizes as in the original published studies, suggesting that effect sizes in existing literature tend to be inflated. Thus, our large sample size (N = 1188 across four studies) should be able to detect true effects with effect sizes that may be more realistic.

Study 1

The purpose of Study 1 was to test whether the use of a green core attribute or a green peripheral attribute influence consumers' inferences about the functional quality of the

green products compared to an identical product without a green attribute. We tested the prediction that consumers would infer *lower* functional quality for the two green attributes in a strength-dependent product category (H1) and *higher* functional quality for the identical attributes in a gentleness-dependent product category (H4). Moreover, we tested whether using a green core (vs green peripheral) attribute would amplify the sustainability liability and asset effects predicted for strength-dependent and gentleness-dependent categories (H2 and H5). Finally, the study tested indirect effects on product preferences through the quality inferences (H3 and H6).

Pretest of Product Category

Thirty-three business-students participated in a pretest conducted to assure proper manipulation of strength-dependent and gentleness-dependent product categories. Following the procedures in Luchs et al. (2010), we tested the degree to which gentleness and strength were valued in three different product categories: shampoo, body lotion, and drain opener. We expected the participants to value gentleness to a larger extent in the shampoo and body lotion categories, and strength to a larger extent in the drain opener category. We asked participants to imagine that they were going to purchase a product in the three categories. The order of the categories was fully rotated. Then, we asked the participants to rate the importance of four gentleness attributes ("gentle", "mild", "soft", and "kind") and four strength attributes ("intense", "aggressive", "strong", and "tough") on a 7-point Likert scale (1 = not important at all, 7 = very important).[1]

A factor analysis of the attributes revealed two factors, strong and gentle. We calculated average measures of the four gentleness items (Cronbach's $\alpha = .87$) and the four strength items (Cronbach's $\alpha = .91$). An analysis of the attribute importance rating for the three categories confirmed that gentleness was significantly more important for body lotion ($M = 5.5$) than for drain opener ($M = 2.5$, $p < .0001$), and that strength was significantly more important for drain opener ($M = 5.5$) than for body lotion ($M = 2.4$, $p < .0001$). The shampoo was also significantly more gentleness-oriented ($M = 4.5$, $p < .0001$) and less strength-oriented ($M = 2.7$, $p < .0001$) than the drain opener. However, since body lotion rated significantly higher on gentleness compared to shampoo ($p < .0001$), we selected body lotion as the product category for further testing.

Stimuli, Procedures and Measures

Four-hundred and thirty-six students (61% male) enrolled in a business graduate school participated in an online survey-experiment set up in Qualtrics. By completing the survey, they had the opportunity to win a Bose QC35 II headset at a value of approximately $400. The experiment was a mixed within-between-subjects design with product category (gentleness-dependent vs. strength-dependent) as the between-subjects factor and type of green attribute (core vs. peripheral vs. regular baseline) as the within-subjects factor. Participants were randomly assigned to one of the product categories (body lotion or drain opener) and asked to imagine that they were going to buy a product in the given category. Then, they were presented with three different alternatives: a product with a green core attribute, a product with a green peripheral attribute, and a no green attribute product. In accordance with Keller's (1993) definition of product-related and non-product-related attributes, we manipulated the green core attribute using the description "100% natural ingredients" and the green peripheral attribute using the description "100% recycled packaging material".

To measure perceived functional quality, the respondents were asked to answer "how would you rate the ability of these products to moisturize dry skin?" (Body lotion)/"How would you rate the ability of these products to open clogged pipes?" (Drain opener) using a 7-point Likert scale anchored by "low ability and "high ability". This measure is based on the measure of product quality in Newman et al. (2014).

We measured product preference by asking participants to rate the likelihood that they would choose each of the different alternatives if they were in need for a body lotion/drain opener. We used a 7-point Likert scale anchored by "not likely at all" and "very likely". A similar measure was used by Newman et al. (2014). Due to the risk of social desirability in answering questions involving socially and environmentally relevant issues, we included a second measure of product evaluation, asking participants to rate the likelihood that each alternative will be a success in the market. We used a 7-point Likert scale to measure anticipated market success, anchored by "not a success at all" and "major success". A similar projective measurement approach was used by Luchs et al. (2010).

As a manipulation check, we included a measure of product greenness, using two items from Gershoff and Frels (2015). Participants were asked to indicate on a 7-point Likert scale their level of agreement with the following statements: "Buying this product is a good environmental choice" and "a person who cares about the environment would buy this product". Since consumers' self-identity as green consumers may influence their preferences for sustainable products (Lin and Chang 2012; Luchs et al. 2012; Olson 2013), we included a control

---

[1] The choice of attributes is partly based on Luchs et al. (2010), but some items were altered to better fit the participants' language (Norwegian).

question capturing the respondents' "green profile". We asked participants to evaluate their level of agreement with the following statement: "It is important to me that the products I purchase are environmentally friendly".

Results

Our manipulation of green products was successful across both product categories; Consumers perceived the two green alternatives to be more sustainable than the regular product (see Table 2 in the Appendix for mean scores).

Strength-Dependent Product Category

A one-way repeated measures ANOVA showed a significant effect of product type on functional quality ($F(2, 210) = 87.00$, $p < .0001$). In support of H1, pairwise comparisons showed that the no green attribute product in the strength-dependent category scored significantly higher on functional quality ($M_{\text{No-green}} = 5.86$, SD = 1.26) compared to both the green core attribute ($M_{\text{Core}} = 4.25$, SD = 1.36, $p < .0001$) and the green peripheral attribute ($M_{\text{Peripheral}} = 5.33$, SD = 1.32, $p < .0001$).

To test whether the green core attribute had a stronger negative effect on functional quality perceptions than the green peripheral attribute, we created a quality difference score for the non-green product compared to the two green products. The hypothesis (H2) was supported if the difference between the green and non-green product was significantly larger for the core (vs peripheral) green attribute. In accordance with our prediction, results showed that the negative difference between the no green attribute and green core attribute on perceived functional quality was significantly greater than the difference score between the no green attribute and green peripheral attribute ($M_{\text{difference score for Core}} = -1.62$, $M_{\text{difference score for Peripheral}} = -.53$, $F(1, 210) = 109.28$, $p < .0001$).

To test whether differences on perceived functional quality mediated effects on product preference, we applied the MEMORE macro for SPSS (Montoya and Hayes 2017), which estimates indirect effects for within-subjects designs. Bootstrap analyses with 5000 samples (Preacher and Hayes 2008) showed a significant negative indirect effect of product greenness on product preferences through functional quality for both the core attribute ($\beta = -.962$, SE = .16, 95%CI -1.281, $-.666$) and the peripheral attribute ($\beta = -.240$, SE = .14, 95%CI $-.397$, $-.117$). Therefore, our findings lend support to H3. Figure 1 displays the statistical mediation diagrams for the effects.

Gentleness-Dependent Product Category

In the gentleness-dependent product category, we did not find the expected difference between the non-green and green products on perceived functional quality (see Table 2 in the Appendix for mean scores and standard deviations). Hence, H4 was not supported by the data. However, the quality difference score between green and non-green product was significantly more positive for the green core (vs green peripheral) attribute ($M_{\text{difference score for Core}} = .14$, $M_{\text{difference score for Peripheral}} = -.53$, $F(1, 210) = 109.28$, $p < .0001$), lending support to H5. Since the green attributes did not enhance functional quality perceptions, there were no indirect effects of the green attributes on product preference through functional quality. Therefore, the data does not support H6.
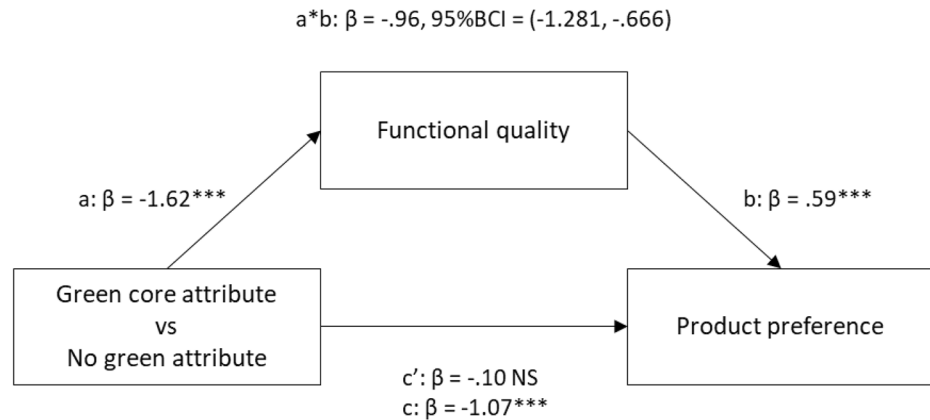
Discussion

Strength-Dependent Product Category

Study 1 provided empirical support for the predicted negative effect of green core and green peripheral attributes on quality inferences in the strength-dependent product category. Although both attributes produced negative quality inferences, this effect was stronger for the core attribute, as predicted. In addition to the expected indirect effects on product preference through functional quality, we observed significant total effects of the green attributes on product preference. Participants had higher preferences for the non-green product ($M_{\text{No-green}} = 5.30$, SD = 1.37) compared to both the green core attribute ($M_{\text{Core}} = 4.24$, SD = 1.56, $p < .0001$) and the green peripheral attribute ($M_{\text{Peripheral}} = 5.02$, SD = 1.39, $p < .05$). Thus, whereas Luchs et al. (2010) in their lab studies only found evidence for a sustainability liability effect when using projective measurement techniques, our study demonstrated a strong sustainability liability effect on personal product preferences.
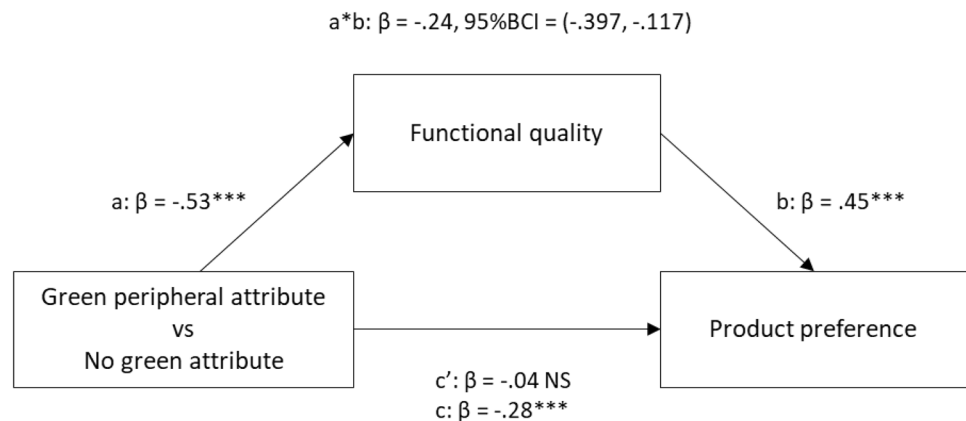
Since consumers with a strong concern for the environment (i.e. high green profile) may prefer green products regardless of a perceived quality trade-off, we tested the potential moderating role of green profile. Results from interaction analyses, using the Johnson-Neyman procedures for probing, showed that participants across all levels of green profile perceived the non-green product to have higher functional quality than the two green products. All participants, regardless of green profile, had stronger preferences for the non-green product compared to the green core attribute. However, when comparing the non-green alternative to the green peripheral attribute, participants with a green profile ($M \geq 5.23$; 19% of the sample) had stronger preferences for the green product. This result suggests that the sustainability liability effect can be attenuated by a green

Fig. 1 Study 1: Statistical mediation diagrams, strength-dependent category. **a** Green core attribute. **b** Green peripheral attribute. *Notes* results are based on MEMORE macro for SPSS. $a*b$ = the indirect effect. 95%CI95% Bootstrapped confidence intervals. The indirect effect is significant when the confidence interval does not include zero. c' = the direct effect. c = the total effect

A Green core attribute

a*b: β = -.96, 95%BCI = (-1.281, -.666)

Functional quality

a: β = -1.62***

b: β = .59***

Green core attribute vs No green attribute

Product preference

c': β = -.10 NS
c: β = -1.07***

B Green peripheral attribute

a*b: β = -.24, 95%BCI = (-.397, -.117)

Functional quality

a: β = -.53***

b: β = .45***

Green peripheral attribute vs No green attribute

Product preference

c': β = -.04 NS
c: β = -.28***

peripheral (vs green core) attribute for green consumers. However, since the functional quality inferences for the peripheral green attribute were negative across all levels of green profile, it is possible that green consumers in a real choice setting would also choose the non-green alternative.

Gentleness-Dependent Product Category

For the gentleness-dependent category, the results differed from our predictions: The green alternatives did not make consumers infer higher functional product quality. Notably, the green peripheral attribute scored significantly *lower* than the non-green product on both functional quality and preference. When including gender as a control variable, however, results showed that type of green attribute had a significant effect also for the gentleness-dependent category ($F(2, 216) = 6.79$, $p < .01$). In order to understand this finding better, we investigated the effects for male and female participants separately. Results showed that the predicted

asset effect for the green core attribute (vs no green attribute) was present for female, but not male, consumers. However, the negative effect of the green peripheral attribute prevailed for both genders. An explanation may be that the body lotion category is more relevant for female consumers, and male consumers may have responded negatively due to the green-feminine stereotype (Brough et al. 2016).

As for the strength-dependent category, we used the Johnson-Neyman probing technique to inspect the nature of the interaction between product type and green consumer profile on functional quality and preference. When comparing the green products to the non-green baseline, the interaction analyses showed that consumers with a low green profile had significantly higher preferences for the non-green product, while consumers with a high green profile had a significantly higher preference for the green products. Concerning functional quality perceptions, green consumers ($M \geq 4.34$, 55% of the sample) perceived the green core attribute (vs. non-green baseline) to have higher functional quality, whereas

the green peripheral attribute had a negative effect on functional quality inferences for all participants, regardless of green profile.

In conclusion, Study 1 suggested that the green peripheral attribute (vs regular baseline) did not lead to improved functional quality inferences for gentleness-dependent products, but may nevertheless be preferred to non-green products among green consumers. The predicted asset effect for the green core attribute was contingent on green consumer profile.

A limitation of the first study was the lack of representativeness of the student sample (cf. Harrison and List 2004), which is particularly important in the investigation of social preferences (Falk et al. 2013). Therefore, it is important to test the robustness of the sustainability liability effect on a more representative sample. Moreover, the asset effect of sustainability in the gentleness-dependent category seemed to be contingent on individual consumer factors, which also warranted further testing on a representative sample.

Study 2

The first purpose of Study 2 was to test the robustness of the effects documented for the strength-dependent product category in Study 1 on a representative sample. Second, the unpredicted findings in the gentleness-dependent category warranted further testing of this category. Third, we wanted to test the research hypotheses in a setting closer to an actual purchase decision, while securing experimental control. The design in Study 1 had some limitations since it represented a hypothetical consumer decision with rather abstract notions of objects to be evaluated. Therefore, in Study 2, we set up the experiment closer to a real shopping situation, using physical products. Thus, the study can be considered a so-called framed field experiment; that is, a randomized experiment on a representative sample of subjects, which has field context in either the commodity, task, or information set that the subjects use (Harrison and List 2004). In collaboration with the largest Norwegian producer of FMCGs, we created a fictitious brand name and used physical bottles with different labels corresponding to the experimental conditions. This also gave us the opportunity to include a physical measure of perceived functional quality; the specific product amount required to solve a functional problem (i.e. open clogged pipes).

Recruitment, Procedures and Material

One-hundred and eighty-one customers recruited at a large shopping mall completed the survey-based experiment. Participants were rewarded a gift card at the value of approximately $7 upon completing the survey. The participants' age ranged from 15 to 78, with an average of 36 years old. 66.3% were female. 40% of the sample reported high school as their highest level of education, 35.4% had a bachelor degree, and 13.8% had a master degree.

Separate stalls were set up, preventing anyone from observing the participants while they filled out the survey on a laptop computer. Two boxes, labelled '1' and '2', were placed inside each stall. The boxes contained three product versions (green core attribute, green peripheral attribute, and no green attribute) either in the gentleness-dependent category (body lotion) or the strength-dependent category (drain opener). Participants were randomly assigned to one of the two product categories by the online survey program. The products were real bottles, and their shape were typical of the product categories. We relabeled the bottles for the purpose of the experiment. All bottles were white, and since color may affect consumers' evaluation of green products (Pancer et al. 2017; Seo and Scammon 2017), we kept the labels black and white. All products were given the fictitious brand name 'Sera'.

We used the same descriptions as in Study 1 to manipulate the type of green attribute ("100% natural ingredients" and "100% recycled packaging material", respectively). To ensure realism, we included standard visual identifiers of the product category on the bottles. The drain openers showed a pipe, and the body lotions had an abstract pattern (see Figs. 8, 9 in the Appendix). In addition, we included a verbal message, indicating the functional benefit of the product: "Unclogs clogged pipes" (strength-dependent category) and "Body lotion for dry skin" (gentleness-dependent category). This serves two purposes. First, such sentences are typical for the category labelling, thus increasing the realism of the product design. Second, since we are testing the effect of green attributes on functionality, it is a stronger test of the effect if we include the information that all products serve this functional benefit. To ensure realism of the design, the bottles were designed in collaboration with product managers at the FMCG company.

After the participants had given their consent to participate in the study, they were given a short introduction to the survey. Since the visual appearance of the labels deviated somewhat from the design of real products, we informed participants that they were going to evaluate beta versions, and that the design would be further developed prior to launching them in stores. On the first page of the survey, participants were instructed to open one of the two boxes that they were randomly assigned to (the product categories). Then, they were asked to imagine that they were going to buy a product in the target category condition, and that they could choose between the three alternatives presented.

Measures

In addition to the measures from Study 1,[2] Study 2 included a measurement task, where participants were asked to indicate how much of each product they thought would be necessary to solve a specific functional problem. The task for the strength-dependent category condition (i.e. drain opener) was framed as follows: "Laboratory tests have been conducted of the products in order to reveal the exact amount of each product that is needed to open completely clogged pipes within 15 min. Therefore, we know how much of each product is needed, and we would like you to guess this amount. The person who comes closest to the correct amount for all three products will win two cinema tickets." Then, we instructed participants to pour drain opener into measuring cups that were included in the box. After completing the task, they were told to use a sliding scale from zero to five hundred milliliters to indicate the amount in each measuring cup. A low product amount indicates perceptions of high functional quality (Lin and Chang 2012).

The measurement task for the gentleness-dependent category (body lotion) was set up differently. We gave the same information about laboratory testing and informed that we knew how much of each product was necessary to effectively moisturize dry skin. Pouring content into a measuring cup is not a meaningful task for this category. Therefore, we first asked participants to answer 'yes' or 'no' to the question of whether they thought the lab tests had shown different amounts for the three products. If they answered 'yes', a new question asked them to rank the three versions of the products from 1 (=the least amount needed) to 3 (=the largest amount needed). Participants were given the option to include their email address at the end of the survey if they wanted to participate in the competition for cinema tickets.[3]

The reason for including the measurement task was twofold. First, since functional quality is the key mechanism explored in this research, it is important to capture multiple facets of the construct. Second, the new measure was an attempt to control for potential social desirability effects for quality inferences, in the same way as a projective technique for product preference (Luchs et al. 2010).

Since consumers tend to perceive sustainable products as more expensive compared to regular products (Bonini and Oppenheim 2008), status motives are a possible driver for green product preferences (Griskevicius et al. 2010).

Therefore, we included price perception as a control measure.[4]

Results

Strength-Dependent Category

A repeated measures ANOVA showed that product type had a significant effect on perceived functional quality ($F(2, 88) = 25.14$, $p < .0001$). As predicted (H1), the no green attribute product was rated higher on functional quality ($M_{\text{No-green}} = 5.64$, SD = 1.28) compared to both the green core attribute ($M_{\text{Core}} = 4.35$, SD = 1.43, $p < .0001$) and the green peripheral attribute ($M_{\text{Peripheral}} = 4.79$, SD = 1.50, $p < .0001$). In this study, we included a second measure of perceived functional quality, by asking respondents to indicate the exact amount of the product they thought was needed to solve a functional problem. A repeated measures ANOVA showed a significant effect of product type on amount indicated as the optimal level for effectiveness ($F(2, 88) = 7.70$, $p < .01$). Mean comparisons showed that participants guessed a significantly higher amount for the green core attribute ($M_{\text{Core}} = 234.7$ ml) compared to the no-green attribute ($M_{\text{No-green}} = 205.9$ ml; $p < .05$). However, contrary to the first quality measure, the green peripheral attribute product was not evaluated as less effective on this functional quality measure ($M_{\text{Peripheral}} = 200.8$ ml).

Analyses of the difference scores on functional quality showed that the difference between the no green attribute and green core attribute on functional quality was significantly greater than the difference score between the no green attribute and green peripheral attribute ($M_{\text{difference score for Core}} = -1.29$, $M_{\text{difference score for Peripheral}} = -.86$, $F(1, 89) = 7.23$, $p < .01$), supporting H2.

Bootstrap analyses with 5000 samples revealed negative indirect effects of sustainability through functional quality on preference for both the green core attribute ($\beta = -1.261$, SE = .25, 95%CI − 1.778, − .823) and the green peripheral attribute ($\beta = -.653$, SE = .19, 95%CI = − 1.084, − .310), lending support to H3. We performed a mediation analysis to test whether there was an indirect effect on product preference also through the new measure of quality. A bootstrap analysis confirmed that the green core (vs no green) attribute
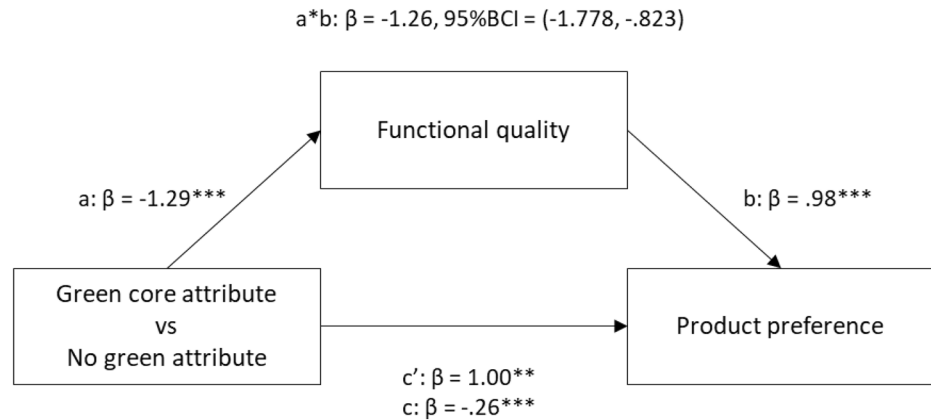

Fig. 2 Study 2: Statistical mediation diagrams, strength-dependent category. **a** Green core attribute. **b** Green peripheral attribute. *Notes* results are based on MEMORE macro for SPSS. $a*b$ = the indirect effect. 95%CI 95% Bootstrapped confidence intervals. The indirect effect is significant when the confidence interval does not include zero. c' = the direct effect. c = the total effect

A Green core attribute

a*b: β = -1.26, 95%BCI = (-1.778, -.823)

Functional quality

a: β = -1.29***

b: β = .98***

Green core attribute vs No green attribute

Product preference

c': β = 1.00**
c: β = -.26***

B Green peripheral attribute

a*b: β = -.65, 95%BCI = (-1.084, -.310)

Functional quality

a: β = -.86***

b: β = .76***

Green peripheral attribute vs No green attribute

Product preference

c': β = .42 NS
c: β = -.23***

had a negative effect on product preference through the perceived amount of product needed to solve the functional problem ($\beta = -.210$, SE $= .14$, 95% CI $-.521$, $-.002$). Figure 2 illustrates the statistical mediation diagrams for the effects.
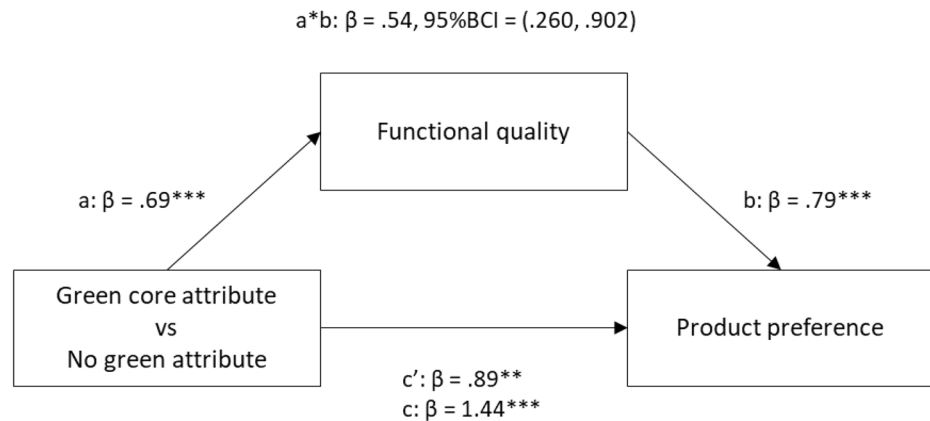
Gentleness-Dependent Category

In the gentleness-dependent category, there was also a significant effect of product type on functional quality ($F(2, 89) = 18.19$, $p < .0001$). Pairwise comparisons showed that the green core attribute was perceived as significantly more effective compared to the no green attribute ($M_{\text{Core}} = 5.21$, SD $= 1.26$, $M_{\text{No-green}} = 4.52$, SD $= 1.46$, $p < .0001$). Contrary to our expectations, but consistent with findings in Study 1, the green peripheral attribute was regarded as significantly *less* effective than the no green attribute ($M_{\text{Peripheral}} = 4.25$, SE $= 1.50$, $p < .05$). Hence, H4 was only partly supported. H5 was supported, as the quality difference score between

non-green and green products was higher for the core than the peripheral attribute ($M_{\text{difference score for Core}} = .692$, $M_{\text{difference score for Peripheral}} = -.264$, $F(1, 90) = 36.73$, $p < .0001$).
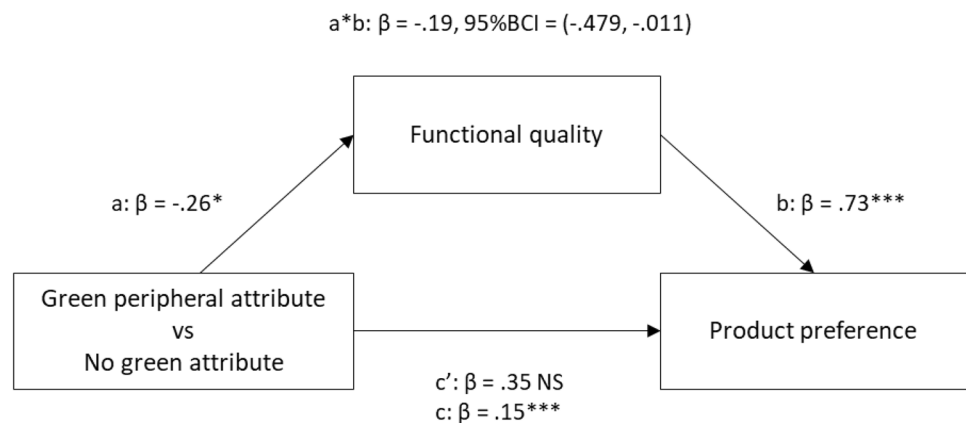
The amount-measurement task had to be performed differently for the gentleness-dependent category. The product that was most frequently rated as the one that needs the *least* amount to moisturize dry skin is regarded as the most effective. 44% of the participants responded that they did not think there was a difference between the products with respect to amount needed to moisturize dry skin. Of the 56% who responded 'yes' to a difference, 60% rated the product with a green core attribute as the one for which the least amount was needed (i.e. it was judged as most effective). This is significantly higher than the null-hypothesis value of 33% ($z = 4.45$, $p < .0001$). 15% rated the product with the green peripheral attribute as the one for which the least amount was needed, which is significantly *lower* than the null-hypothesis value ($z = 2.97$, $p < .001$). The product

Fig. 3 Study 2: Statistical mediation diagrams, gentleness-dependent category. **a** Green core attribute. **b** Green peripheral attribute. *Notes* results are based on MEMORE macro for SPSS. $a*b$ = the indirect effect. 95%CI 95% Bootstrapped confidence intervals. The indirect effect is significant when the confidence interval does not include zero. c' = the direct effect. c = the total effect

A Green core attribute

a*b: β = .54, 95%BCI = (.260, .902)

Functional quality

a: β = .69***

b: β = .79***

Green core attribute
vs
No green attribute

Product preference

c': β = .89**
c: β = 1.44***

B Green peripheral attribute

a*b: β = -.19, 95%BCI = (-.479, -.011)

Functional quality

a: β = -.26*

b: β = .73***

Green peripheral attribute
vs
No green attribute

Product preference

c': β = .35 NS
c: β = .15***

with no green attribute was rated as the one for which the least amount was needed among 25% of the participants, which is *not* significantly different from the null-hypothesis value ($z = 1.34$, $p = .19$). These results are consistent with the explicit measure of perceived functional quality: The product with the green core attribute was perceived as the most effective product, whereas the product with the green peripheral attribute was seen as the least effective product.

Mediation analysis using bootstrapped confidence intervals showed that functional quality significantly mediated the preference for the green core attribute (vs non-green) product ($\beta = .545$, SE $= .16$, 95%CI .260, .902). Since the green peripheral attribute did not improve functional quality perceptions, H6 is only partially supported. Figure 3 illustrates the statistical mediation diagrams for the effects in the gentleness-dependent category. As shown in diagram B, there is a significant *negative* indirect effect of the green peripheral attribute on product preference through functional quality.

In this study, we included price perceptions as a control variable. As expected, the vast majority of the participants expected a price difference across the products in each category (94.4% in the strength-dependent category and 90.1% in the gentleness-dependent category), and the green products were expected to be more expensive than the non-green product. Controlling for price perceptions did not alter the results from the hypotheses testing.

Discussion

Strength-Dependent Product Category

The findings from our second study showed that participants considered both green alternatives in the strength-dependent category to be less effective than the non-green product, representing an indirect negative effect on consumer preferences. As for Study 1, the negative effect on functional quality perceptions was significantly lower when the green attribute was peripheral rather than core. Interaction analyses

using Johnson-Neyman probing demonstrated that all participants, regardless of green profile, evaluated the green products as less functional than the regular one. However, when analyzing product preference, we found a significant preference for the non-green product only for consumers with a low score on green profile ($M \leq 4.09$; 40% of the sample). This supports the idea that green consumers may prefer a green product despite of being conscious about a trade-off with quality.

Gentleness-Dependent Product Category

Whereas in Study 1, the green core attribute improved functional quality perceptions only among green consumers, Study 2 showed that the green core attribute led to improved functional quality inferences regardless of green consumer profile. There was a positive indirect effect of the green core attribute on product preference through functional quality. Therefore, Study 2 supported the predicted sustainability asset effects when the green attribute was related to the core functions of the product. However, Study 2 replicated the unexpected negative effect for the green peripheral attribute. Similarly to Study 1, moderation analyses showed that green consumers ($M \geq 5.14$; 30% of the sample) preferred the green peripheral attribute to the no-green attribute. However, this seems to happen only because of their concern for the environment, as also the green consumers perceive the green peripheral product to hold lower functional quality compared to the non-green product. We designed a third study to address this particular finding.

Study 3

Studies 1 and 2 both showed that the green peripheral attribute (vs no green attribute) was a liability rather than an asset in the gentleness-dependent category. In an attempt to understand this unexpected finding better, we designed a third study only for the gentleness-dependent category. The aim of this study was to test a different way of communicating the green peripheral attribute for the gentleness-dependent category. A consistent finding in the two first studies was that "recycled packaging material" as a green label had a negative effect on functional quality inferences, also among participants with a high green profile. Hence, the theoretical assumption of a correlation-based inference process seems to have worked in the opposite direction of our predictions for the gentleness-dependent category: consumers use the green peripheral attribute to infer lower functional product quality. We speculated whether participants may have associated the term 'recycled packaging material' with second-hand material (i.e. waste), which is likely to be negatively evaluated in a personal care category. Therefore, we wanted

to test the asset prediction using the term "100% plant-based packaging material". This is a proper manipulation of a peripheral attribute, as the focus is on the packaging rather than the core product.

Experimental Design and Procedure

To achieve a representative sample and the benefits of physical products, we conducted this study in the field, similarly to Study 2. We recruited one-hundred and sixty-four customers (38% male, average age: 35 years old) at a shopping mall to participate in an experiment with type of green peripheral attribute label as the between-subjects factor ("100% recycled packaging material" vs "100% plant-based packaging material"), and the type of green attribute as the within-subjects factor (core, peripheral, non-green). Participants were randomly assigned to the between-subjects factor. We used the same measures as in the previous studies.

Results

Results showed that for the peripheral label used in the two first studies ("100% recycled packaging material"), the green peripheral attribute scored significantly lower on functional quality compared to the non-green product ($M_{\text{Peripheral}} = 4.09$, SD = 1.70, $M_{\text{No-green}} = 4.47$, SD = 1.68, $p < .05$). The new alternative peripheral label ("100% plant-based packaging material") scored higher on functional quality than the non-green product, but the difference was not statistically significant ($M_{\text{Peripheral}} = 4.48$, SD = 1.66, $M_{\text{No-green}} = 4.25$, SD = 1.54, $p = .14$). Similarly to Study 2, the green peripheral attribute (in both label conditions) was significantly preferred among green consumers, suggesting that the green peripheral attribute can produce an asset effect in gentleness-dependent categories, but not through functional quality inferences.

In support of H4a, the green core attribute improved functional quality perceptions of the gentleness-dependent product ($M_{\text{Core}} = 5.05$ SD = 1.48, $M_{\text{No-green}} = 4.25$, SD = 1.54, $p < .0001$), which had a significant indirect effect on product preference ($\beta = .550$, SE = .11, 95%CI.349, .770). Therefore, H4a and H6a were supported.

Study 4

The three first studies tested type of green attribute as a within-subjects factor. Although such design reflects how consumers are making choices in stores for these types of products (i.e. high ecological validity), we cannot rule out the possibility that the effects of attribute type are a result of contrast- or spillover-effects between the green attributes in the choice set. For example, when simultaneously presented

with a green core attribute, the information about the packaging being recycled may evoke expectations that also the core attribute should be green. Consequently, when contemplating both a core and peripheral attribute that is green, the lack of core attribute information may have created a negative contrast effect for the green peripheral product.

We designed the fourth study using product type as a between-subjects factor to test whether the green peripheral attribute could improve functional quality perceptions of the gentleness-dependent product when participants were not reminded of a core attribute. The between-subjects design also allowed us to test whether the liability effect documented for the green peripheral attribute in the strength-dependent category had been merely a spillover effect due to the within-subjects design in the first two studies. When presented with information about the core attribute, participants may have assumed that the product with recycled packaging material (peripheral attribute) also contained natural ingredients (core attribute). A between-subjects design would rule out such an explanation.

Experimental Design and Procedure

Four-hundred and seven Norwegian residents were recruited from a large online panel to participate in a 2 (product category: strength-dependent vs gentleness-dependent) × 2 (green attribute type: core vs peripheral) full factorial experiment. When recruiting participants to the gentleness-dependent category, we used a screening question to secure category relevance ("have you ever bought cleaning products for small children, or do you plan to do so in the near future"). This left us with a sample of two-hundred and three participants in the gentleness-dependent category, of which 53% was male, and the average age was 36. For the strength-dependent product category, the initial recruitment process provided a sample with an average age of 47, which was considerably higher than in the other studies. Therefore, we made an effort to recruit more respondents into this category, targeting younger residents specifically. We ended up with a total sample of $N = 565$ (47% male), with an average age of 31 in the strength-dependent category.

Consistent with manipulations used in prior research (Luchs et al. 2010; Lin and Chang, 2012), we used hand sanitizer to represent the strength-dependent category and baby shampoo for the gentleness-dependent category. We collaborated with the same FMCG company in the development of realistic product labels. For this study, we developed two different brand names ('Sera' and "Aveno"). Thus, participants were presented with two brands with different names, and we rotated which brand was presented with a green attribute. The purpose of using two different brands was to increase the realism in the judgment task,

and to reduce the possibility that consumers infer the same attributes across the products due to the same brand name. For the green peripheral attribute, we used the new claim developed for study 3 ("plant-based packaging material"), and we kept the same label for the green core attribute ("100% natural ingredients"). The materials used for Study 4 can be found in the appendix.

All participants were asked to evaluate a green product against a non-green counterpart, before they were instructed to rate the products' functional performance ("How would you rate the ability of this product to kill bacteria/clean the child's hair") on a 7-point Likert scale, anchored by "low ability" and "high ability". We measured product preference, product greenness and green consumer profile in the same way as in the previous studies.

## Results

Strength-Dependent Product Category

Consistent with the two first studies, participants rated the green core attribute significantly lower on functional quality compared to the non-green product ($M_{Core} = 5.34$, SD $= 1.38$, $M_{No\text{-}green} = 5.53$, SD $= 1.34$, $p < .0001$). Results show that also the green peripheral attribute received significantly lower functional quality ratings than the non-green product ($M_{Peripheral} = 5.55$, SD $= 1.32$, $M_{No\text{-}green} = 5.65$, SD $= 1.24$, $p < .05$). Hence, the data supports H1. The quality difference scores were not significantly different from each other, meaning that H2 is not supported by the data. Mediation analysis showed that there were significant negative effects of sustainability on product preferences through functional product quality, both when the green attribute was core ($\beta = -.153$, SE $= .05$, 95%CI $-.275$, $-.060$) and peripheral ($\beta = -.090$, SE $= .04$, 95%CI $-.181$, $-.010$), supporting H3.

As for the other studies, we tested the moderating role of green consumer profile. Johnson-Neyman probing showed that green consumers had significant preferences for both green attributes (vs no-green attribute), despite their negative perceptions of functional product quality.

Gentleness-Dependent Product Category

For the gentleness-dependent category, the green core attribute (vs no green attribute) had a significant positive effect on functional quality ($M_{Core} = 4.72$, SD $= 1.37$, $M_{No\text{-}green} = 4.39$, SD $= 1.40$, $p < .05$). However, similar to the three first studies, there was no positive effect of the green peripheral attribute ($M_{Peripheral} = 4.45$, SD $= 1.22$, $M_{No\text{-}green} = 4.42$, SD $= 1.33$), providing only partial support for H4. H5 was supported since the quality difference score was significantly

Fig. 4 Meta-analytic forest plot of the green core attribute effect in strength-dependent categories. *Notes* error bars indicate 95% confidence intervals. The confidence intervals are numerically stated in parentheses after each effect size. The meta-analytic effect across all studies is placed in the bottom right corner ($d = -.57$)
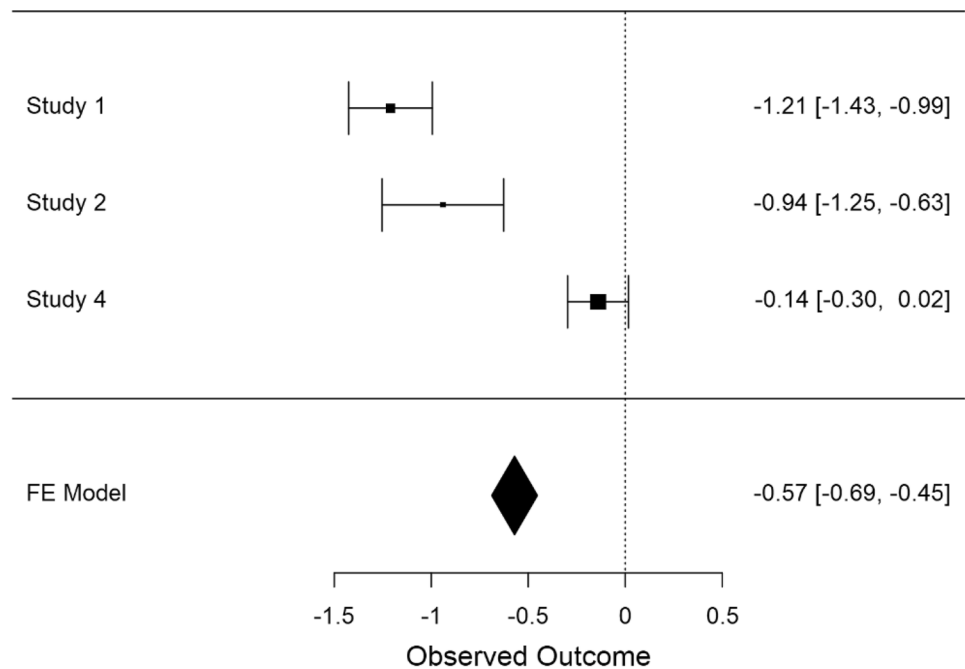
more positive for the core (vs. peripheral) attribute ($M_{\text{difference score for Core}} = .32., M_{\text{difference score for Peripheral}} = -.04$, $p < .05$). As predicted by H6a, there was a significant positive effect of the core attribute on product preference through functional quality ($\beta = .247$, SE $= .11$, 95%CI.062, .472).

Internal Meta-analyses

We have reported results from four experimental studies, of which all studies tested the asset effect for the gentleness-dependent category, and three studies (studies 1, 2 and 4) tested the liability effect for the strength-dependent category. Results showed three consistent patterns of findings. (1) Consumers infer lower functional quality for both green core and green peripheral attributes compared to no green attribute in the strength-dependent product category (i.e. the sustainability liability effect). (2) Consumers infer higher functional quality for products with a green core attribute compared to no green attribute in the gentleness-dependent category (i.e. the sustainability asset effect). (3) There is either a negative or no effect of the green peripheral attribute (vs no green attribute) on functional quality inferences in the gentleness-dependent category.

In order to evaluate the strength of these effects, we conducted internal meta-analyses. This allowed us to estimate a single meta-analytic effect across the studies. The analyses were conducted in JASP (2019) using fixed effects. The estimations were based on standardized effect sizes (Cohen's d) and the standard error of each effect.

The Sustainability Liability Effect in Strength-Dependent Categories

The meta-analytical result showed that the average sustainability liability effect on functional quality for the green core attribute was $d = -.57$, which indicates a medium effect size, that is highly significant (Z $= -.9.51$, $p < .001$). Thus, in line with our prediction in H1a, consumers infer lower functional quality when a product in a strength-dependent category has a green core attribute (vs. no green attribute). The meta-analytic effect is illustrated in Fig. 4.

The meta-analysis for the green peripheral attribute showed an effect size of $d = -.35$, which indicates a moderate effect size, that is significant (Z $= -.2.27$, $p < .05$). Therefore, our research shows that there is a liability effect even for a non-product-related attribute, such as packaging, in strength-dependent product categories. However, the effect is weaker than the liability effect documented for the green core attribute, which is also in line with our predictions (H2). The meta-analytic effect for the green peripheral attribute is illustrated in Fig. 5.

The Sustainability Asset Effect in Gentleness-Dependent Categories

Concerning the predicted asset effect for the green core attribute in gentleness-dependent product categories, the meta-analysis shows a moderate effect size $d = .27$, which is highly significant (Z $= .4.33$, $p < .001$). The weakest effect ($d = .11$) is found in Study 1, where we tested body lotion on a student sample. As discussed under Study 1, we suspect

Fig. 5 Meta-analytic forest plot of the green peripheral attribute effect in strength-dependent categories. *Notes* error bars indicate 95% confidence intervals. The confidence intervals are numerically stated in parentheses after each effect size. The meta-analytic effect across all studies is placed in the bottom right corner ($d = -.35$)
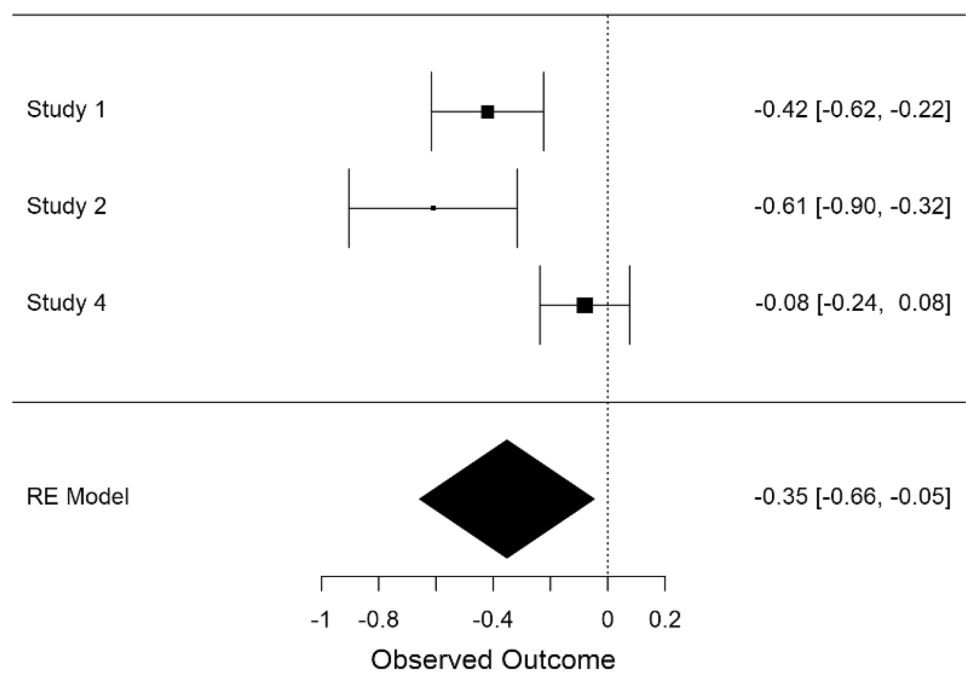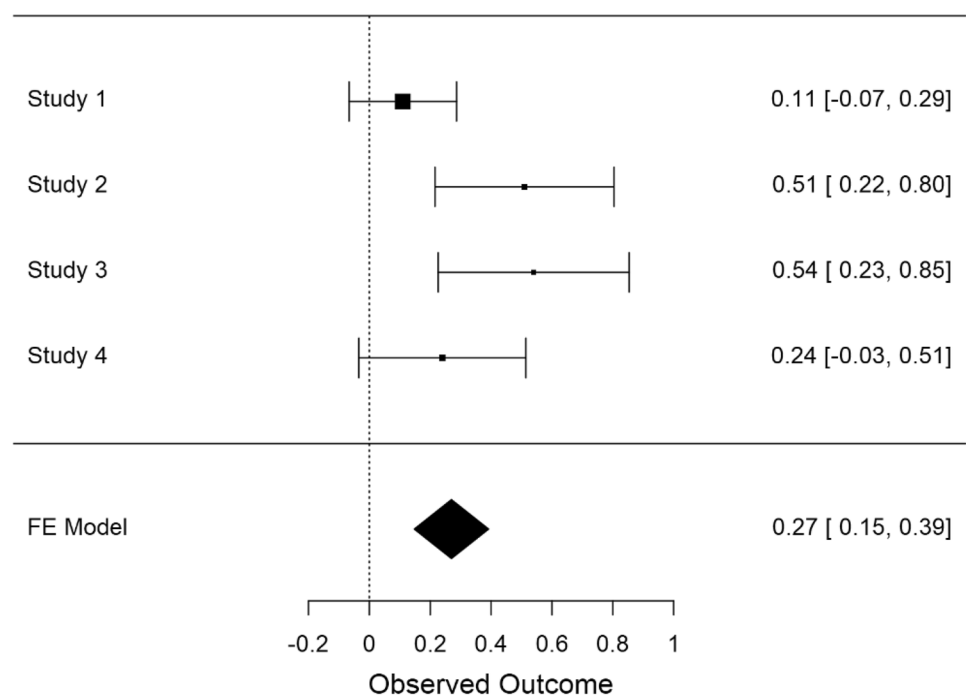

Fig. 6 Meta-analytic forest plot of the green core attribute effect in gentleness-dependent categories. *Notes* error bars indicate 95% confidence intervals. The confidence intervals are numerically stated in parentheses after each effect size. The meta-analytic effect across all studies is placed in the bottom right corner ($d = .27$)


that the lack of an asset effect was due to the low perceived relevance of the category for male students. Therefore, on the basis of the meta-analytical results, our research suggests that a green core attribute makes consumers infer higher functional quality in gentleness-dependent product categories. The meta-analytic effect for the green core attribute in the gentleness-dependent category is illustrated in Fig. 6.

The meta-analytic result for the green peripheral attribute in the gentleness-dependent category showed a small, but significant negative effect ($d = -.19$; $Z = -.319$, $p < .001$) on functional quality. The results are illustrated in Fig. 7, showing that Study 3 was entered as two separate studies, reflecting the two peripheral labels. Study 3a represents the original label ("recycled packaging material") and Study 3b represents the new label ("plant-based packaging material"). The new peripheral label was used also for Study 4. From the forest plot in Fig. 7, it is evident that the negative effects of the green peripheral attribute occurred when the original

Fig. 7 Meta-analytic forest plot of the green peripheral attribute effect in gentleness-dependent categories. *Notes* error bars indicate 95% confidence intervals. The confidence intervals are numerically stated in parentheses after each effect size. The meta-analytic effect across all studies is placed in the bottom right corner ($d = -.19$). Study 3a represents the original peripheral label, and 3b is the new peripheral label
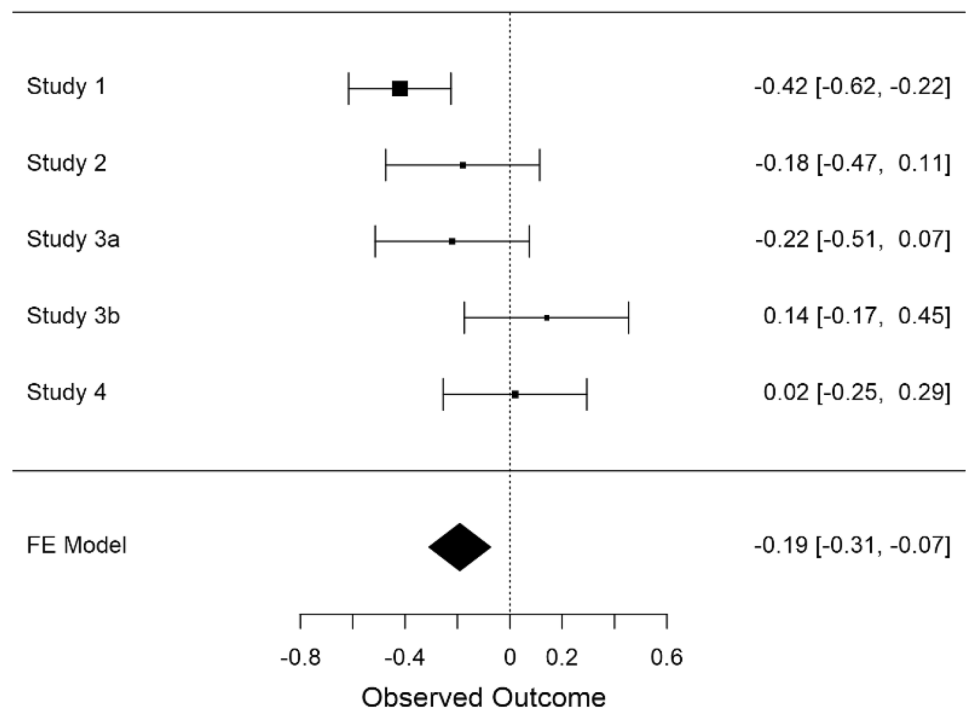
label was used (Studies 1, 2, and 3a), whereas the new green peripheral label had no effect on functional quality perceptions (Studies 3b and 4).

General Discussion

Across three experiments (Studies 1, 2 and 4), we found consistent evidence of a sustainability liability effect on functional quality inferences for both green core and green peripheral attributes in strength-dependent categories. The internal meta-analysis suggests that the liability effect on functional quality inferences is smaller for the green peripheral attribute. In addition to the functional quality inference effects (H1–H2), all studies showed significant indirect effects of the green attributes on product preference through functional quality perceptions (H3). In Study 2 and Study 4, probing-results from the interaction analyses showed that participants with a high green profile had stronger preferences for the green products, despite their negative inferences about green product quality.

These results suggest that green product attributes will make consumers infer lower functional quality for strength-dependent products. This holds whether the attributes are directly linked to the functions of the product (e.g. product ingredients), or only peripherally connected to the product (e.g. product packaging). However, our results suggest that green consumers may accept the quality trade-off and choose the green alternatives. Nevertheless, it should be noted that self-reported preference measures are susceptible to social desirability, and that the liability effect on personal preferences may be present in real choice settings for green consumers (Luchs et al. 2010).

The results for the gentleness-dependent product category largely support the predicted sustainability asset effect on functional quality when the green attribute is core. While prior research on the sustainability-quality relationship has focused mostly on quality trade-off perceptions, our research demonstrates that sustainability can improve perceptions about product performance when gentleness-related attributes are valued. However, for such an effect to occur, our research suggests that the green attribute must have a direct link to the core functions of the product. If the green attribute is only peripherally linked to the core product, the asset effect may be attenuated (Studies 3b and 4) or even create a liability effect (Studies 1, 2 and 3a).

Results for the hypotheses testing across all studies are reported in Table 1.

Theoretical Contributions

This research furthers the current understanding of how consumers evaluate sustainable products in several ways. First, we demonstrate that consumers form negative functional quality inferences in strength-dependent categories, even when the green attribute is only peripherally linked to the core product. To the best of our knowledge, no prior research has made theoretical predictions for how different green attributes or claims can influence functional quality inferences. Given that functional quality will be a key consideration in real choice situations (Luchs et al. 2010), knowledge about the influence of green claims on quality inferences is vital insight.

Table 1 Summary of hypothesis-testing, Study 1–4

Hypotheses	Study 1	Study 2	Study 3	Study 4
H1a: In strength-dependent product categories, consumers infer lower functional product quality when the product has a green core attribute (vs. no green attribute)	Supported	Supported	n.a.	Supported
H1b: In strength-dependent product categories, consumers infer lower functional product quality when the product has a green peripheral attribute (vs. no green attribute)	Supported	Supported	n.a.	Supported
H2: The sustainability liability effect on functional product quality will be stronger for the green core (vs green peripheral) attribute	Supported	Supported	n.a.	Not supported
H3: There is a negative indirect effect of (a) green core attribute and (b) green peripheral attribute (vs. no green attribute) on product preference through perceived functional product quality	Supported	Supported	n.a.	Supported
H4a: In gentleness-dependent product categories, consumers infer higher functional product quality when the product has a core green attribute (vs. no green attribute)	Not supported	Supported	Supported	Supported
H4b: In gentleness-dependent product categories, consumers infer higher functional product quality when the product has a peripheral green attribute (vs. no green attribute)	Not supported	Not supported	Not supported	Not supported
H5: The sustainability asset effect on functional product quality will be stronger for the green core (vs green peripheral) attribute	Supported	Supported	Supported	Supported
H6: There is a positive indirect effect of (a) green core attribute and (b) green peripheral attribute (vs. no green attribute) on product preference through perceived functional product quality	Not supported	(a) Supported (b) Not supported	(a) Supported (b) Not supported	(a) Supported (b) Not supported

Second, our research responds to calls for more research into the role of sustainability in categories that appeal to gentleness (Lin and Chang 2012). Most research inquiries on the topic of quality inferences for sustainable products have focused on trade-off effects (Lin and Chang 2012; Luchs and Kumar 2017; Luchs et al. 2012; Pancer et al. 2017). Our research advances the current understanding of when sustainability may serve as an asset. According to the benefit-congruity principle proposed by Luchs et al. (2010), one should expect a positive effect of any sustainability cue in categories where gentleness-related attributes are valued. In accordance with Luchs et al. (2010), three of our studies (Studies 2, 3b and 4) showed a significant positive total effect of the green peripheral attribute on product preference (see Tables 2, 3, 4, 5, 6, 7 in the Appendix for mean scores). However, the asset effects on preference caused by the peripheral attribute could not be explained by improved functional quality inferences. Across two gentleness-dependent product categories (body lotion and baby shampoo), using two different peripheral labelling strategies ("recycled packaging material" and "plant-based packaging material"), there were no positive effects of such non-product-related green attributes on functional quality inferences. For the green core attribute, on the other hand, positive preferences can be explained by higher functional quality perceptions.

On a more general note, our research contributes to the understanding of what type of inference-strategies consumers use when evaluating the quality of sustainability-labelled products. Consumers may apply different strategies for making inferences about quality when they encounter information about sustainability. Newman et al. (2014) show that consumers use compensatory inferences and assume that (intentional) product sustainability means that the company deters resources from quality to sustainability. However, this result was produced when consumers encountered information about the production processes and were explicitly asked to evaluate resource allocation (which sensitizes consumers to the compensatory process). If consumers had applied the compensatory inference in our study, we expect that there would have been a negative effect of sustainability for both the strength-dependent and the gentleness-dependent category. However, in accordance with our predictions, there is a sustainability asset effect for the gentleness-dependent category, suggesting that consumers engage in correlational inferences, such as the probabilistic inference strategy (Chernev and Carpenter 2001).

Practical Implications

Understanding when sustainability is a liability and when it is an asset is vital information for marketers of sustainable products. The present work supports prior research showing that sustainability can be a disadvantage in categories where strength-related attributes are valued (Luchs et al. 2010). We demonstrate across three studies that even green peripheral attributes lead to negative functional quality inferences, which reduce consumers' preferences for the products. Therefore, contrary to prior advice to use peripheral attributes to attenuate a liability effect (Gershoff

& Frels, 2015), our findings imply that companies should try to overcome the seemingly paradoxical effect that consumers infer poorer functional product quality even when the packaging is sustainable. In this regard, we find the result of the measurement task in Study 2 particularly relevant. The fact that a cognitively more demanding task removed the perceived liability effect from the green peripheral attribute suggests that consumers should be encouraged to reflect upon the type of green attribute and its implications for functional quality. For example, when introducing a new version of a strength-dependent product in green packaging, companies may highlight the fact that the core attribute (the ingredients) is the same.

The results for the gentleness-dependent product category implies that marketers of products where gentleness-related attributes are valued should be careful when using green peripheral attribute labelling in isolation. The current research did not test a joint labelling strategy (both core and peripheral attribute claims), but this could potentially be a way to avoid negative effects of the peripheral attribute claim alone.

Limitations and Future Directions

Our results consistently show that consumers infer lower functional quality for sustainable products in categories where strength-related attributes are valued. The liability effect is present even for an attribute that has no direct link to the core product. Future research should test whether the quality liability effect extends to attributes that have an even weaker link to the core product, such as a social dimension of sustainability (e.g. fair trade). Indeed, Newman et al. (2014) showed that a benefit that is separated from the product (e.g. fair trade) did not lead to compensatory inferences about the product's quality, as did a benefit that was inherent to the product (e.g. the formula for cleaning products).

The liability effect for the green peripheral attribute disappeared when participants in Study 2 were asked to pour content into a measurement cup as an indication of functional product quality. The measurement task is more demanding than answering a single scale-based quality question. Therefore, we argued that the task itself may have encouraged more active reasoning about the association between a sustainable packaging and product performance. Whether higher-level elaboration about green peripheral attributes may attenuate the liability effect should be addressed in future studies. For example, the quality inference effects documented in our research could be tested in contexts that vary according to consumers' cognitive busyness (Gilbert and Hixton 1991). Because people under cognitive load are less able to correct existing perceptions and rely more on stereotypes (Gilbert et al. 1988), one may expect that the liability effect for the peripheral green attribute is stronger when cognitive busyness is high. Future studies should test whether consumers

will be less likely to account for the lack of causal link between the peripheral attribute and product performances in purchase situations that pose fewer constraints on cognitive reasoning. Further, studies may test whether marketing communication can promote active reasoning and thus reduce the negative effect of a peripheral green attribute.

While our findings lend support to the existence of an asset effect in the gentleness-dependent category, further research is needed to shed more light on the effect of the green peripheral attribute in this category. A potential line of inquiry could for instance follow the approach of Pancer et al. (2017), who studied the differential effects of isolated versus combined green cues in packaging design. Future studies may test whether the negative effect of a green peripheral attribute in gentleness-dependent categories may be attenuated by the presence of a green core attribute.

Although we tested the research hypotheses in a setting close to real purchase decisions (i.e. shopping malls), a limitation of our research is the controlled laboratory environment for judgments and decision-making. We suggest that future research should validate our findings using experimental designs in the field, measuring actual purchase decisions.

Compliances with Ethical Standards

Conflict of interest We hereby declare that there is no conflict of interest.

Ethical Approval All individuals who participated in the experimental studies gave their informed consent.

Appendix

See Tables 2, 3, 4, 5, 6 and 7 and Figs. 8, 9, 10, and 11.

Table 2 Mean scores Study 1

Measures	Strength-dependent category (N=212)			Gentleness-dependent category (N=224)		
	Core	Peripheral	Non-green	Core	Peripheral	Non-green
Functional qual.	4.25 (1.36)	5.33 (1.32)	5.86 (1.26)	5.36 (1.31)	4.70 (1.28)	5.22 (1.21)
Product pref.	4.24 (1.56)	5.02 (1.39)	5.30 (1.37)	5.10 (1.55)	4.50 (1.49)	4.98 (1.40)
Market success	4.49 (1.24)	5.04 (1.21)	5.29 (1.20)	5.39 (1.15)	4.75 (1.23)	5.29 (1.20)
Prod. greenness	5.66 (1.24)	5.97 (1.08)	2.81 (1.08)	5.06 (1.24)	6.11 (.94)	3.16 (.99)

Mean scores with SD in parenthesis

Table 3 Mean scores Study 2

Measures	Strength-dependent category (N=90)			Gentleness-dependent category (N=91)		
	Core	Peripheral	Non-green	Core	Peripheral	Non-green
Functional qual.	4.35 (1.43)	4.79 (1.50)	5.64 (1.28)	5.21 (1.26)	4.25 (1.50)	4.52 (1.46)
Amount[a]	234.7 ml (121.3)	200.9 ml (95.0)	206.0 ml (116.4)	60%	15%	25%
Product pref.	4.72 (1.88)	4.74 (1.88)	4.98 (1.80)	5.38 (1.57)	4.10 (1.83)	3.95 (1.70)
Market success	4.83 (1.63)	4.67 (1.65)	4.42 (1.60)	5.46 (1.21)	4.73 (1.56)	3.84 (1.47)
Prod. greenness	5.56 (1.47)	4.70 (1.58)	2.46 (1.55)	5.04 (1.27)	5.78 (1.19)	2.69 (1.39)

Mean scores with SD in parenthesis

[a]For the gentle category, amount is shown as the percentage of participants that indicated the lowest amount needed. Lowest rating indicates highest perceived functional quality

Table 4 Mean scores Study 3 (original peripheral label)

Measures	Original peripheral label (N=85) ("100% recycled packaging material")		
	Core	Peripheral	Non-green
Functional qual.	5.06 (1.54)	4.09 (1.70)	4.47 (1.68)
Product pref.	5.18 (1.78)	4.44 (1.84)	4.47 (1.71)
Market success	5.56 (1.39)	4.88 (1.58)	4.21 (1.73)

Mean scores with SD in parenthesis. Study 3 only tested the gentleness-dependent product category

Table 6 Mean scores Study 4 (strength-dependent product category)

Measures	Core vs. regular (N=283)		Peripheral vs. regular (N=282)	
	Core	Regular	Peripheral	Non-green
Functional qual.	5.34 (1.38)	5.53 (1.34)	5.55 (1.26)	5.65 (1.18)
Product pref.	4.75 (1.73)	4.67 (1.69)	4.67 (1.79)	4.61 (1.75)
Market success	4.66 (1.32)	4.61 (1.33)	4.71 (1.32)	4.67 (1.33)

Mean scores with SD in parenthesis

Table 5 Mean scores Study 3 (new peripheral label)

Measures	New peripheral label (N=79) ("100% plant-based packaging material")		
	Core	Peripheral	Non-green
Functional qual.	5.05 (1.42)	4.48 (1.66)	4.25 (1.54)
Product pref.	5.56 (1.38)	4.66 (1.60)	3.78 (1.66)
Market success	5.56 (1.33)	4.82 (1.43)	4.00 (1.59)

Mean scores with SD in parenthesis. Study 3 only tested the gentleness-dependent product category

Table 7 Mean scores Study 4 (gentleness-dependent product category)

Measures	Core vs. regular (N=102)		Peripheral vs. regular (N=101)	
	Core	Regular	Peripheral	Non-green
Functional qual.	4.72 (1.37)	4.39 (1.40)	4.42 (1.33)	4.46 (1.22)
Product pref.	4.08 (1.68)	3.69 (1.69)	3.62 (1.61)	3.67 (1.57)
Market success	4.32 (1.24)	4.10 (1.29)	3.98 (1.13)	3.95 (1.13)

Mean scores with SD in parenthesis

Fig. 8 Study 2: Treatments in the gentleness-dependent category; label designs from left to right: core attribute (natural ingredients), peripheral attribute (recycled packaging material), and regular product (no green attribute)

Fig. 9 Study 2: Treatments in the strength-dependent category; label designs from left to right: core attribute (natural ingredients), peripheral attribute (recycled packaging material), and regular product (no green attribute)

Fig. 10 Study 4: Treatments in the gentleness-dependent category; label designs from left to right (for both brand names): core attribute (natural ingredients), peripheral attribute (plant-based packaging material), and regular product (no green attribute)

Fig. 11 Study 4: Treatments in the strength-dependent category; label designs from left to right (for both brand names): core attribute (natural ingredients), peripheral attribute (plant-based packaging material), and regular product (no green attribute)

References

Auger, P., & Devinney, T. (2007). Do what consumers say matter? The misalignment of preferences with unconstrained ethical intentions. *Journal of Business Ethics, 76*(4), 361–383. https://doi.org/10.1007/s10551-006-9287-y.

Bonini, S., & Oppenheim, J. (2008). Cultivating the green consumer. *Stanford Social Innovation Review, 6*(4), 56–61.

Brough, A. R., Wilkie, J. E. B., Jingjing, M. A., Isaac, M. S., & Gal, D. (2016). Is eco-friendly unmanly? The green-feminine stereotype and its effect on sustainable consumption. *Journal of Consumer Research, 43*(4), 567–582. https://doi.org/10.1093/jcr/ucw044.

Chernev, A., & Carpenter, G. S. (2001). The role of market efficiency institutions in consumer choice: A case of compensatory inferences. *Journal of Marketing Research (JMR), 38*(3), 349–361.

Cho, Y.-N. (2015). Different shades of green consciousness: The interplay of sustainability labeling and environmental impact on product evaluations. *Journal of Business Ethics, 128*(1), 73–82. https://doi.org/10.1007/s10551-014-2080-4.

Cho, Y.-N., & Baskin, E. (2018). It's a match when green meets healthy in sustainability labeling. *Journal of Business Research, 86,* 119–129. https://doi.org/10.1016/j.jbusres.2018.01.050.

Dick, A., Chakravarti, D., & Biehal, G. (1990). Memory-based inferences during consumer choice. *Journal of Consumer Research, 17*(1), 82–93.

Du, S., Bhattacharya, C. B., & Sen, S. (2007). Reaping relational rewards from corporate social responsibility: The role of competitive positioning. *International Journal of Research in Marketing, 24*(3), 224–241. https://doi.org/10.1016/j.ijresmar.2007.01.001.

Falk, A., Meier, S., & Zehnder, C. (2013). DO lab experiments misrepresent social preferences? The case of self-selected student samples. *Journal of the European Economic Association, 11*(4), 839–852. https://doi.org/10.1111/jeea.12019.

Gershoff, A. D., & Frels, J. K. (2015). What makes it green? The role of centrality of green attributes in evaluations of the greenness of products. *Journal of Marketing, 79*(1), 97–110.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60*(4), 509–517. https://doi.org/10.1037/0022-3514.60.4.509.

Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology, 54*(5), 733–740. https://doi.org/10.1037/0022-3514.54.5.733.

Gosselt, J. F., van Rompay, T., & Haske, L. (2019). Won't get fooled again: The effects of internal and external CSR ECO-labelling. *Journal of Business Ethics, 155*(2), 413–424. https://doi.org/10.1007/s10551-017-3512-8.

Griskevicius, V., Tybur, J. M., & Van den Bergh, B. (2010). Going green to be seen: Status, reputation, and conspicuous conservation. *Journal of Personality and Social Psychology, 98*(3), 392–404. https://doi.org/10.1037/a0017346.

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature, 42*(4), 1009–1055. https://doi.org/10.1257/0022051043004577.

Keller, K. L. (1993). Conceptualizing, measuring, managing customer-based brand equity. *Journal of Marketing, 57*(1), 1–22.

Lin, Y.-C., & Chang, C.-C. A. (2012). Double standard: The role of environmental consciousness in green product usage. *Journal of Marketing, 76*(5), 125–134.

Luchs, M. G., Brower, J., & Chitturi, R. (2012). Product choice and the importance of aesthetic design given the emotion-laden trade-off between sustainability and functional performance product choice and the importance of aesthetic design given the emotion-laden trade-off between sustainability and functional performance. *Journal of Product Innovation Management, 29*(6), 903–916. https://doi.org/10.1111/j.1540-5885.2012.00970.x.

Luchs, M., & Kumar, M. (2017). 'Yes, but this Other one looks better/works better': How do consumers respond to trade-offs between sustainability and other valued attributes? *Journal of Business Ethics, 140*(3), 567–584. https://doi.org/10.1007/s10551-015-2695-0.

Luchs, M. G., Naylor, R. W., Irwin, J. R., & Raghunathan, R. (2010). The sustainability liability: Potential negative effects of ethicality on product preference. *Journal of Marketing, 74*(5), 18–31. https://doi.org/10.1509/jmkg.74.5.18.

Montoya, A. K., & Hayes, A. F. (2017). Two-condition within-participant statistical mediation analysis: A path-analytic framework. *Psychological Methods, 22*(1), 6–27. https://doi.org/10.1037/met0000086.

Newman, G. E., Gorlin, M., & Dhar, R. (2014). When going green backfires: How firm intentions shape the evaluation of socially beneficial product enhancements. *Journal of Consumer Research, 41*(3), 823–839.

Olsen, M. C., Slotegraaf, R. J., & Chandukala, S. R. (2014). Green claims and message frames: How green new products change brand attitude. *Journal of Marketing, 78*(5), 119–137.

Olson, E. (2013). It's not easy being green: The effects of attribute tradeoffs on green product preference and choice. *Journal of the Academy of Marketing Science, 41*(2), 171–184. https://doi.org/10.1007/s11747-012-0305-6.

Pancer, E., McShane, L., & Noseworthy, T. (2017). Isolated environmental cues and product efficacy penalties: The color green and eco-labels. *Journal of Business Ethics, 143*(1), 159–177. https://doi.org/10.1007/s10551-015-2764-4.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879–891.

Sachdeva, S., Jordan, J., & Mazar, N. (2015). Green consumerism: Moral motivations to a sustainable future. *Current Opinion in Psychology, in press.*

Seo, J., & Scammon, D. (2017). Do green packages lead to misperceptions? The influence of package colors on consumers' perceptions of brands with environmental claims. *Marketing Letters, 28*(3), 357–369. https://doi.org/10.1007/s11002-017-9420-y.

Unilever. (2017, January 1). *Report shows a third of consumers prefer sustainable brands.* https://www.unilever.com/news/Press-releases/2017/report-shows-a-third-of-consumers-prefer-sustainable-brands.html. Retrieved January 13, 2019.

Vanclay, J. K., Shortiss, J., Aulsebrook, S., Gillespie, A. M., Howell, B. C., Johanni, R., et al. (2011). Customer response to carbon labelling of groceries. *Journal of Consumer Policy, 34*(1), 153–160. https://doi.org/10.1007/s10603-010-9140-7.

Wood, S., Robinson, S., & Poor, M. (2018). The efficacy of green package cues for mainstream versus niche brands: How mainstream green brands can suffer at the shelf. *Journal of Advertising Research, 58*(2), 165–176. https://doi.org/10.2501/JAR-2018-025.