Journal of Cognitive Psychology

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/pecp21

Conceptual and methodological considerations to the negative footprint illusion: a reply to Gorissen et al. (2024)

Patrik Sörqvist & John E. Marsh

To cite this article: Patrik Sörqvist & John E. Marsh (10 Oct 2024): Conceptual and methodological considerations to the negative footprint illusion: a reply to Gorissen et al. (2024), Journal of Cognitive Psychology, DOI: 10.1080/20445911.2024.2412030

To link to this article: https://doi.org/10.1080/20445911.2024.2412030

Published online: 10 Oct 2024.

JOURNAL OF COGNITIVE PSYCHOLOGY
https://doi.org/10.1080/20445911.2024.2412030

Routledge
Taylor & Francis Group

BRIEF REPORT

Conceptual and methodological considerations to the negative footprint illusion: a reply to Gorissen et al. (2024)

Patrik Sörqvist[a,b] and John E. Marsh[b,c]

[a]Department of Building Engineering, Energy Systems, and Sustainability Science, University of Gävle, Gävle, Sweden; [b]Department of Health, Learning and Technology, Luleå University of Technology, Luleå, Sweden; [c]Human Factors Group, School of Psychology and Humanities, University of Central Lancashire, Preston, UK

ABSTRACT

When asked to estimate the *carbon footprint* of a bundle of low carbon footprint and high carbon footprint items, people typically report a lower value compared to estimating the high carbon footprint items alone. This finding is called the *negative footprint illusion*. Previous research suggests that people might be made less susceptible to this effect depending on whether they are asked to evaluate how *environmentally friendly* or how *environmentally damaging* the items are. In the current study, we used large instead of small stimulus sets (i.e. a more powerful experimental manipulation than that in previous research) and show under these circumstances it does not matter whether participants are required to make *friendliness* or *damaging* estimates. The role of attribute substitution along with other conceptual and methodological issues to the negative footprint illusion are discussed, particularly in relation to a recent paper by Gorissen et al. [2024. Green versus grey framing: Exploring the mechanism behind the negative footprint illusion in environmental sustainability assessments. *Sustainability*, *16*(4), 1411].

ARTICLE HISTORY
Received 29 May 2024
Accepted 29 September 2024

KEYWORDS
Negative footprint illusion; judgment; environmental impact; bias; attribute substitution

Introduction

When asked to estimate the *carbon footprint* of a bundle of relatively low carbon footprint and relatively high carbon footprint items, people typically report a lower value than when estimating the high carbon footprint items alone. This finding is called the *negative footprint illusion* (Gorissen & Weijters, 2016) and has been studied extensively during the last decade (see Sörqvist et al., 2020; and Andersson et al., 2024 for reviews).

The negative footprint illusion resembles the negative calorie illusion (Chernev, 2011) which is the observation that when asked to estimate the *calorific content* of a bundle of relatively low-calorie foods and relatively high-calorie foods, people typically report a lower calorie value than when estimating the high-calorie food items alone. For example, a hamburger (high-calorie food) together with some carrots (low-calorie food) are assigned a lower calorific value together in comparison with the hamburger alone. This similarity between the negative footprint illusion and the negative calorie illusion suggests that they are both part of a family of cognitive effects that emerge when people estimate the characteristics of bundles of vices and virtues (Chernev & Gal, 2010).

One key finding is that the magnitude of the negative footprint illusion (and thus the mechanism(s) that produce the effect) depends on the characteristics of the to-be-estimated stimuli. For example, when the to-be-estimated items are distributed irregularly over the visual field, the effect increases in magnitude, in comparison with when the high carbon footprint items are presented in one group, spatially separated from another group comprising the low carbon footprint items (Sörqvist et al., 2022). Moreover, the effect is larger for a set comprising many, as compared with few, low carbon footprint items, regardless of whether the number of high carbon footprint items in the set is constant: As the number of low carbon footprint items increases, the negative footprint illusion increases even if the ratio between the number of low and high carbon footprint items remains identical (Andersson et al., 2024). The type of item also seems to matter to some extent. When an organic apple (a low carbon footprint item) is added to a burger (a high carbon footprint item) and people are asked to estimate the carbon footprint of the full meal comprising both items, the negative footprint illusion arises (Gorissen & Weijters, 2016). The same thing happens when representations of

CONTACT Patrik Sörqvist patrik.sorqvist@hig.se University of Gävle, Kungsbäcksvägen 47, 801 76 Gävle, Sweden

green buildings (low carbon footprint items) are combined with conventional buildings (high carbon footprint items) into a common set (Andersson et al., 2024; Holmgren et al., 2018; Holmgren et al., 2018; Sörqvist et al., 2022; Sörqvist & Holmgren, 2022; Threadgold et al., 2021); when hybrid cars are combined with petrol cars (Holmgren et al., 2021; Threadgold et al., 2021); but not when organic apples are combined with regular apples (Threadgold et al., 2021).

Another general tendency emerging from the growing body of literature on the negative footprint illusion is that it appears to be resistant to variations in the dependent measure. The effect seems to behave the same way regardless of whether estimates are made of the items' carbon footprint or the items' carbon dioxide emissions (Holmgren et al., 2021). Further, the effect behaves similarly when participants are asked to make an indirect estimate of the carbon footprint of the items, by estimating how many trees would be needed to compensate for the emissions associated with the construction of buildings, through the process of carbon binding (Holmgren et al., 2018). The effect seems also to be insensitive to the colour of the response scale (Gorissen & Weijters, 2016). Moreover, most studies on the negative footprint illusion have used a type of response scale that may promote a qualitative rather than a quantitative mindset when making the estimates. Even so, when participants are asked to make the estimate on a quantitative/objective response scale (reporting a kilogram estimate to the question "how many kilograms CO_2 do the items generate?"), rather than a more qualitative/subjective response scale (reporting a small–large estimate to the question "how large is the item's carbon footprint?"), the negative footprint illusion still emerges (Sörqvist & Holmgren, 2022; see Biernat et al., 1991, for an influential paper on the importance of the difference between subjective and objective response scales in human judgment).

Against this background, the results from a recent study by Gorissen et al. (2024) are at first glance surprising. They found that estimates of the *environmental friendliness* of a set comprising a burger (high carbon footprint) and an organic apple (low carbon footprint) were higher than for a set comprising the burger alone, whereas no difference was found between the two sets when participants were asked to estimate how *environmentally damaging* they were. The authors argued that these results suggest an important difference between "green" and "grey" judgment scales. Green scales refer to estimates of how good something is for the environment, whereas grey scales refer to estimates of how bad something is for the environment. The authors argued that the difference arises because green scales trigger an evaluative (qualitative) mindset, whereas grey scales evoke a summative (quantitative) mindset (cf. Holmgren et al., 2018). Here, an evaluative mindset refers to a mindset wherein participants attend to the items' affective/qualitative value on a goodness/badness dimension. In turn, a summative mindset refers to a mindset wherein participants attend to the quantitative values of the items and how these add together. Thus, the idea is that a negative footprint illusion, here represented by the difference in estimates of the two sets (organic apple and burger vs. burger alone), appears with the green scale, but not with the grey scale, because the green but not grey scale supposedly triggers an evaluative mindset wherein the estimate results in an average of the goodness and the badness of the combined items rather than their sum.

In view of past research on this phenomenon, several issues arise from this line of thought. First, there is a conceptual problem. It could be argued that a dish comprising a burger and an organic apple is indeed more *environmentally friendly* than the burger alone, because the organic apple is perceived as an environmentally friendly object. Assigning a higher estimate for this set is thus normatively correct, even if the industrial production of an organic apple has a carbon footprint. This is different from the typical negative footprint illusion paradigm, wherein estimates are made of carbon footprint: A low carbon footprint house, in combination with a high carbon footprint house, does not have a lower carbon footprint in total than the high carbon footprint house alone. Assigning a lower *carbon footprint* for the combined set is therefore false. It can only become true through an attribute substitution process (Kahneman & Frederick, 2001), whereby participants (subconsciously) replace the question "how large is the carbon footprint of the items?" with the simpler one ("how environmentally friendly are the items?") of evaluating the environmental friendliness of the items instead of the items' carbon footprint. In other words, to say that a low carbon footprint house and a high carbon footprint house in combination are more *environmentally friendly* than the high carbon footprint house alone can indeed be correct, if the houses are evaluated on this friendly-unfriendly dimension, but to say that the combination has a lower *carbon footprint* would be normatively incorrect.

Second, as the authors accurately point out (Gorissen et al., 2024), past studies on the negative footprint illusion have all used a grey scale, whereby the participants have been asked to evaluate how environmentally damaging the items are (in terms of, e.g. the items' carbon footprint or how much CO_2 they generate).

However, a difference between the two sets (organic apple and burger vs. burger alone) was not found in Gorissen et al.'s study using the grey scale. Given the broad array of experiments that have incorporated grey scales and have indeed observed a negative footprint illusion with such scales, why did Gorissen et al. fail to observe an effect? We believe the answer to this question lies in the stimuli they adopt.

A small set size produces a small negative footprint illusion. Sometimes small set sizes result in no negative footprint illusion at all, but rather a "zero footprint illusion" (or a quantity insensitivity effect; Kim & Schuldt, 2018) whereby estimates for a small set comprising low and high carbon footprint items are no different from those of a small set comprising just the high carbon footprint items (Andersson et al., 2024). Estimates of similar magnitude for the two sets do not reflect a negative footprint illusion but parity of estimates is not normatively correct either. Such findings might reflect a weak effect from small experimental manipulations. When comparing estimates of a single-item set with a two-item set, as in Gorissen et al.'s (2024) study, a very small negative footprint illusion is thus expected. Consistently, the difference in the *environmental damage* estimates of the two sets was in the same direction as would be expected for a negative footprint illusion (i.e. a lower *environmental damage* estimate for the burger and apple combined as compared with the burger alone), but the difference was very small and not statistically significant. A larger set size might hence paint a different picture.

The purpose of this study was to critically test whether it matters to ask participants to estimate how *environmentally damaging* versus how *environmentally friendly* a set of items is, through using a comparably large stimulus set. Specifically, in this study participants viewed sets comprising a couple of dozen (sketched) houses, for which their degree of environmental friendliness varied. This variation was communicated by the colour of the houses whereby green indicated a low carbon footprint (low environmental impact), red illustrated a high carbon footprint (high environmental impact), and yellow depicted an intermediate carbon footprint (moderate environmental impact). It was hypothesised that participants would assign (1) a lower environmentally damaging estimate to a set comprising yellow and green houses in combination, in comparison with the estimate of the yellow houses alone; (2) a higher environmentally friendliness estimate to a set comprising yellow and green houses in combination, in comparison with the estimate of the yellow houses alone; and (3) that these two effects would be similar in magnitude. Furthermore, all possible combinations of yellow, green, and red houses were included to thoroughly explore the similarities in response patterns across the two judgment scales and across stimuli that varied in their environmental impact.

Methods

Participants

An a priori power analysis (using G*Power; Faul et al., 2007) based on the effect size of Cohen's $d_z = 0.53$ for the negative footprint illusion reported in Sörqvist and Holmgren (2022) revealed that 49 participants would be enough to detect the effect in a within-participants design with power (1 — β error probability) set to 0.95 and assuming a two-tailed hypothesis. A total of 59 individuals took part in this study (70% women, 29% men, and 1% who either did not want to reveal their gender or identified with another gender). Thus, the study was adequately powered. The participants' mean age was 26.04 years ($SD = 7.20$). They all received a small honorarium for their participation. The participants were recruited from the student pool at the University of Gävle and the data collector's social networks. The data collection adhered to the Declaration of Helsinki and the ethical guidelines given by APA. Written informed consent was obtained from all participants prior to participation and the study was approved by the Swedish Ethical Review Authority (Dnr 2023-01109-01).

Materials

Sketched houses were used as stimuli and were identical except that they were depicted in either green, yellow, or red colour. These houses were distributed across an invisible 9×8 matrix to create 7 pictures (see Figure 1 for examples): green-coloured houses only (24 green houses, pseudo-randomly distributed, such that each row had 3 houses), yellow-coloured only (24 yellow houses, distributed as the green houses in the matrix), red-coloured only (24 red houses, distributed as the green houses in the matrix), green and yellow houses (24 of each), red and yellow houses (24 of each), green and red houses (24 of each), and a picture with all three types of house (24 of each, thus all cells in the matrix were filled). These 7 pictures were also all inverted over the x-axis, to create another set of 7 pictures with the same contents but with the houses distributed differently.

Design and procedure

A within-participants experimental design was used. The participants sat alone, in a lab, in front of a desktop

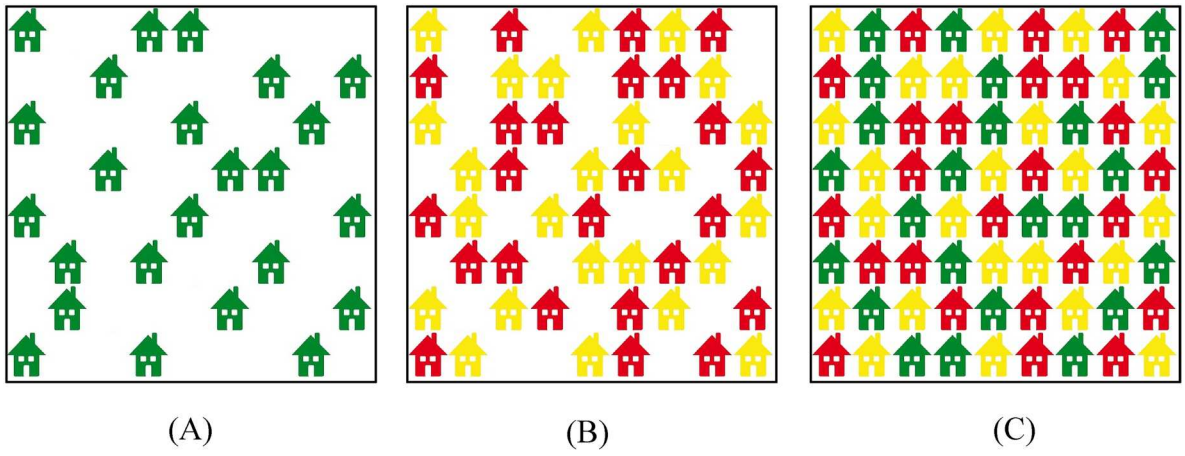

Figure 1. Examples of the stimulus material used in the experiment: only green houses (Panel A), yellow and red houses (Panel B), and a picture with all three types of houses (Panel C).

computer during the data collection. The computer controlled the presentation of instructions, stimulus material and response collection. At the onset, the participants read instructions about the environmental impact of buildings and how this environmental impact can be represented by a colour system ranging from green to red. They were also shown a picture illustrating a building environmental impact evaluation scale, ranging from dark green (lowest environmental impact), to red (highest environmental impact), with yellow (intermediate environmental impact) in the middle of the 7-point scale. Participants were also told that their task would be to view pictures of houses with varying environmental impact, corresponding to their colour as explained by the rating scale, and to make estimates of the houses' combined environmental impact in two ways: either by judging how environmentally friendly the houses in the picture are, or by judging how environmentally damaging the houses in the picture are. After reading the instructions, the participants pressed a button to proceed to the estimation task. The task was divided into two blocks. In one block, the computer presented each of the 7 pictures in random order and asked participants to estimate how environmentally friendly the items are, one picture/estimate at a time. Estimates were made on a seven-point scale (similar to Gorissen et al., 2024) ranging from 1 (not environmentally friendly) to 7 (environmentally friendly). The participants were allowed as much time as they needed to make an estimate and pressed a button to proceed to the next picture. The other block was identical, with the exception that here participants were asked to estimate how environmentally damaging the items are. Estimates were made on a seven-point scale ranging from 1 (environmentally damaging) to 7 (not environmentally damaging). Note that a low number for both scales represents "bad for the environment" and a high value represents "good for the environment". Seven estimates (out of 826 estimates in total) were missing, probably due to omissions. These missing data points were replaced by the average of the estimates of the corresponding item set. The order of the two blocks was counterbalanced between participants, so that half began with making damaging estimates in the first block they encountered, while the other half began with making friendliness estimates in the first block they encountered. At the switch between blocks, participants read a brief instruction of the change in the judgment scale. The regular pictures were presented in the first block, and the inverted versions of the pictures were presented in the second block. The whole experiment took 5–10 min to complete. The data can be accessed from doi.org/10.17605/OSF.IO/SXMQC.

Results

As can be seen in Figure 2, judgments made on the "green" scale were similar to judgments made on the "grey" scale. A 7(stimulus set) × 2(judgment scale) repeated measures analysis of variance was conducted first, as an overarching test of whether the judgment scale generally influenced participants' estimates and whether it interacted with stimulus sets. The analysis revealed a significant effect of stimulus set, $F(6, 348) = 305.84$, $p < .001$, $\eta_p^2 = .84$, $BF_{10} = \infty$. There was no effect of scale, $F(1, 58) = 0.47$, $p = .497$, $\eta_p^2 = .008$, $BF_{01} = 13.14$, and no interaction between the factors, $F(6, 348) = 1.57$, $p = .156$, $\eta_p^2 = .03$, $BF_{01} = 28.42$. Thus, we conclude that participants' estimates were consistent across all stimulus conditions, regardless of the judgment scale used (environmental friendliness or environmental damage).

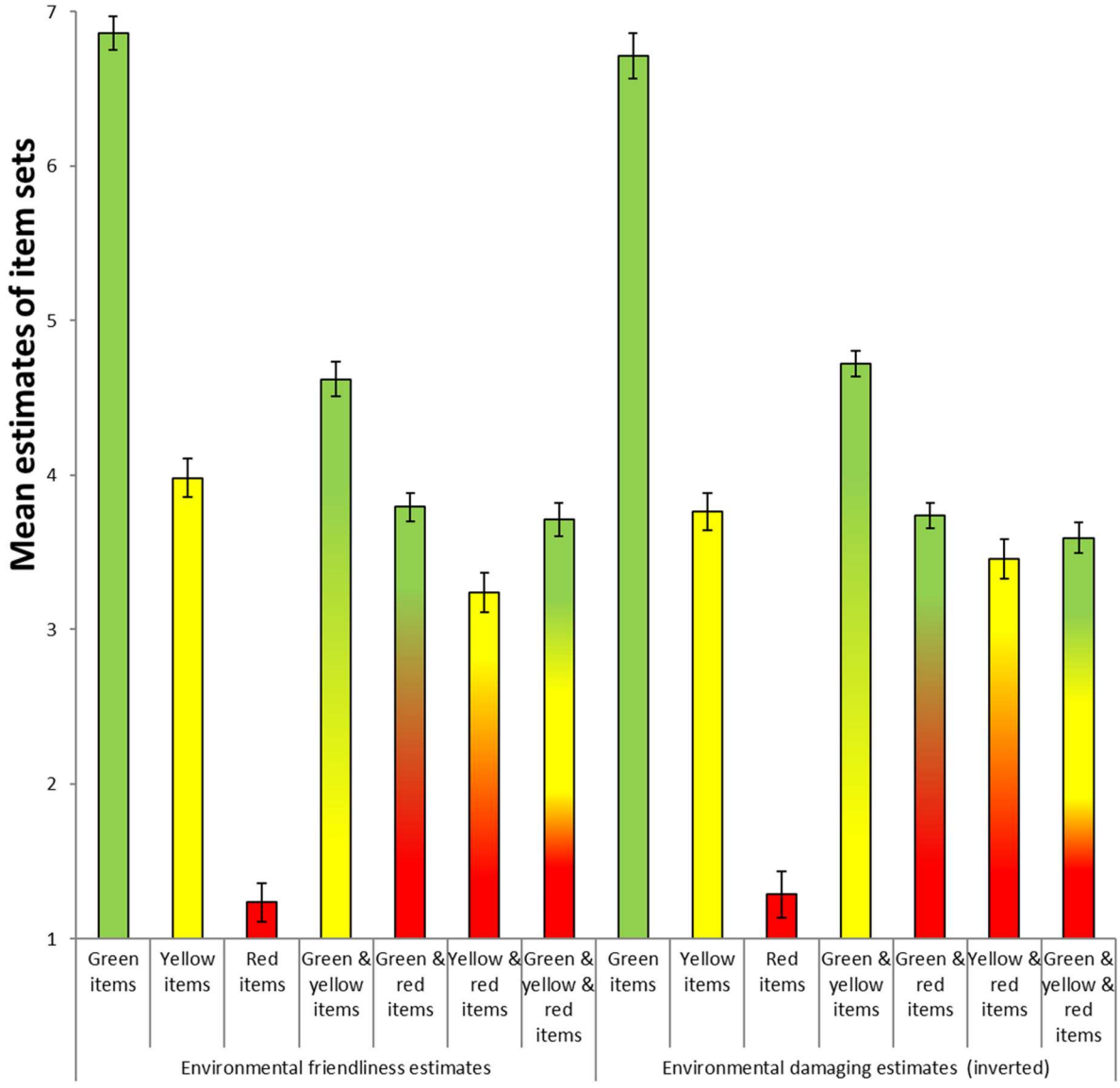

Figure 2. Mean estimates of different clusters of stimuli across two judgment scales: estimates of how environmentally friendly the items are and estimates of how environmentally damaging the items are. The colours of the bars represent the colours of the stimuli in the estimated set. Note that higher values for the environmental friendliness estimates (left) represent higher environmental friendliness, and higher values for the environmentally damaging estimates (right) represent lower environmental damage as the latter scale was inverted. Error bars represent standard error of conditional means.

Next, we turned to hypothesis testing. The first thing to note is that participants assigned a lower environmental damaging estimate to an item set comprising green and yellow houses in comparison with a set comprising the yellow houses alone, $t(58) = 8.30$, $p < .001$, Cohen's $d = 1.08$, $BF_{10} = 4.78 \times 10^8$. This confirms Hypothesis 1 and corresponds to the typical pattern found in the negative footprint illusion literature, but in that literature as opposed to here, participants have been requested to estimate the carbon footprint of the items.

The second thing to note is that participants also assigned a higher environmental friendliness estimate to an item set comprising green and yellow houses in comparison with a set comprising the yellow houses alone, $t(58) = 4.13$, $p < .001$, Cohen's $d = 0.54$, $BF_{10} = 190.70$. This confirms Hypothesis 2.

A third thing to note is that the difference between the two key conditions (i.e. stimulus comprising green houses and yellow houses versus yellow houses only) was similar in magnitude across the two scales. The mean difference scores for the estimates of a set with green + yellow houses and a set with yellow houses were 0.96 ($SD = 0.89$) for the grey scale and 0.64 ($SD = 1.18$) for the green scale, respectively. Although these means differed statistically, $t(58) = 2.05$, $p = .045$,

Cohen's d = 0.24, $BF_{01} = 1.32$, the Bayes Factor favours the null over the hypothesis. Hence, the empirical marker that corresponds to a negative footprint illusion seems similar across the green and grey judgment scales, confirming Hypothesis 3, but as the frequentist and the Bayesian analysis were inconsistent, this conclusion must be interpreted with caution. If anything, the difference was larger for the grey (environmentally damaging) than for the green (environmentally friendliness) ratings, which contradicts the findings of Gorissen et al. (2024).

It is also noteworthy that the typical "negative footprint illusion" effect pattern, observed when estimates of a stimulus set comprising low environmental impact (green) items and intermediate environmental impact (yellow) items are compared with estimates of the items with intermediate environmental impact alone, was also found with other stimulus combinations. For example, intermediate environmental impact (yellow) and high environmental impact (red) items, in combination, were rated as more environmentally friendly and less environmentally damaging, than items with high environmental impact (red) alone. A statistical analysis confirmed these conclusions. A 2(stimulus set: red and yellow versus red only) × 2(response scale: green versus grey) repeated measures analysis of variance revealed a significant effect of stimulus set, $F(1, 58) = 418.05$, $p < .001$, $BF_{10} = 1.50 \times 10^{15}$, but no effect of scale, $F(1, 58) = 1.86$, $p = .178$, $BF_{01} = 3.34$, and no interaction between the factors, $F(1, 58) = 0.73$, $p = .396$. $BF_{01} = 2.88$. This suggests that the typical pattern of judgment, with lower environmental impact estimates of an item set with low and intermediate carbon footprint items in comparison with estimates of an item set comprising only the intermediate carbon footprint items, is not restricted to estimates of items with a specific degree of environmental impact. It generalises to stimuli with other degrees of environmental impact as well. The findings also provide further evidence of the similarity between the two judgment scales.

A control analysis of potential effects of the order of judgment scale condition (i.e. making green ratings first versus making grey ratings first) was conducted, by running an analysis of variance with scale order as a between-participants independent variable, scale (green versus grey ratings) as a within-participants independent variable, and stimulus conditions (green and yellow houses versus yellow houses only) as a within-participants independent variable. The analysis found no three-way interaction, $F = 0.99$, $p = .323$, $BF_{01} = 10.87$, with a Bayes Factor in strong favour of the null-hypothesis, confirming that it did not matter whether participants made green or grey estimates first. Perhaps more importantly, an independent samples *t*-test revealed no difference in the magnitude of the difference between the key stimulus conditions (as calculated by taking estimates of green and yellow houses minus estimates of yellow houses only) among participants who made estimates on the green scale first ($M = 0.77$, $SD = 0.94$) and participants who made estimates on the grey scale first ($M = 0.68$, $SD = 1.10$), $t(57) = 0.34$, $p = .736$, Cohen's d = 0.09, $BF_{01} = 4.85$, confirming that the difference between the two key stimulus conditions was similar regardless of judgment scale.

Discussion

The experiment reported here revealed estimation patterns very similar for estimates of items' *environmental friendliness* and estimates of how *environmentally damaging* the items are. Specifically, the results revealed that a lower environmentally damaging estimate was assigned to a set comprising yellow and green houses in combination, in comparison with the estimate of the yellow houses alone (confirming Hypothesis 1 and at odds with the results reported by Gorissen et al., 2024); a higher environmental friendliness estimate was assigned to a set comprising yellow and green houses in combination, in comparison with the estimate of the yellow houses alone (confirming Hypothesis 2); and this effect was similar in magnitude across the two scales (confirming Hypothesis 3, and also at odds with the results reported by Gorissen et al., 2024).

There are several methodological differences between the current study and the study by Gorissen et al. (2024) that could contribute to this difference in results. For example, Gorissen et al. deployed a between-participants design whilst a within-participant design was adopted here. It is possible that the choice of study design was a driver for the difference in findings reported. We adopted a within-participants design to optimise the comparison of how people respond using the two judgment scales, making sure that differences between conditions/scales could not be attributed to individual differences between participants, which is often a source of confounding factors in between-participants designs. However, an interpretation of our results that would align with Gorissen et al. (2024) is that, in our within-participant design, the differentiation between qualitative and quantitative mindsets, arguably evoked by the green and grey scales (cf. Gorissen et al., 2024), may have been reduced with participants exposed to both types of scales somehow blending their responses. This might, for example, be due to participants becoming more familiar with the scales as they progress through the experiment and/or

through comparing their estimates on certain stimuli with those made to previous stimuli. If the two judgment scales produce different mindsets, the responses by participants who began with the green judgment scale should differ from the responses by participants who began with the grey judgment scale, especially in the first stimulus block they encounter, since this is when their mindsets became differently activated by the two judgment scales. This should be the case even if the effect of different mindsets is not seen in the overall analysis across all stimuli conditions. However, this idea was not supported by the results of our control analysis. In contrast, the control analysis suggests that the response pattern was similar regardless of which judgment scale was used first.

Another difference between the current study and the study by Gorissen et al. (2024) is the stimulus sets, both regarding the identity of the items (i.e. houses of different colours corresponding to the environmental impact of the items versus a burger and an apple with or without an eco-label) and the quantity of the items (i.e. several items versus just one or two items). As argued in the introduction, the negative footprint illusion is often small in magnitude when few to-be-estimated items are used and becomes larger in magnitude as a function of increasing number of items (Andersson et al., 2024). A possible explanation of the absence of an effect with the grey scale in the study by Gorissen et al. (2024) is, hence, the small stimulus set size.

In relation to the difference in identity of the stimulus items, past research suggests that the negative footprint illusion can be found with different stimulus identities, such as when estimates of a burger is compared with estimates of a burger and an organic apple (Gorissen & Weijters, 2016), when estimates of petrol and hybrid cars are compared with estimates of petrol cars (Holmgren et al., 2021), when estimates of regular and green buildings are compared with estimates of regular buildings alone (Holmgren et al., 2018), and so forth. Different stimulus identity between the current study and the study by Gorissen et al. (2024) seems, therefore, to be an unlikely explanation of the difference in results. However, one study failed to find a negative footprint illusion when estimates of regular apples were compared with estimates of regular apples and organic apples in combination (Threadgold et al., 2021). One possibility is that different stimuli trigger different mindsets, and this interacts with the judgment scale, which might explain the difference in results between the current study and that of Gorissen et al. (2024). Even though the current study provides evidence against the assumption that different judgement scales trigger different mindsets, mindset seems to play a role in the negative footprint illusion in other ways. For example, when a summative mindset is triggered by a priming technique, the negative footprint illusion is attenuated (Holmgren et al., 2021). Because of this, a discussion of the results in relation to this mindset framework might be informative.

Let us first consider the assumption that the stimulus sets used here trigger a quantitative mindset. Whilst our results revealed evidence for a difference in estimates of green and yellow houses versus yellow houses only (i.e. the empirical signature of the negative footprint illusion, if the estimates had concerned the items' carbon footprint) for the green and grey scales, the effect was somewhat more pronounced for the grey (environmentally damaging scale). This difference between the two scales was statistically weak and can possibly be attributed to chance but may still deserve an in-depth discussion within the mindset framework proposed by Gorissen et al. (2024). One possibility for the more pronounced effect of the grey scale is that the sketched houses in the study are inherently more quantifiable with characteristics such as colour, number and their assigned environmental impact being more easily compared and calculated. For example, participants might mentally subtract the number of green houses from yellow ones to estimate a lower overall impact in line with "zero-sum game" reasoning (cf. Rozycka-Tran et al., 2015) whereby participants believe the total environmental damage to be fixed such that if one aspect increases (e.g. adding eco-friendly houses) so too must another aspect (e.g. the carbon footprint of the conventional houses) decrease by an equivalent amount. On this line of reasoning the stimuli set (apple and burger) deployed by Gorissen et al. (2024) does not lend itself so easily to such quantitative reasoning because it involves just one apple and one burger without explicit quantities, apart from the apple's bio-label. The difficulty in comparing these dissimilar items might thereby promote adoption of a more evaluative and qualitative mindset than a quantitative one. On this view, the greater as opposed to lesser effect with the grey scale found here might be attributed to a mindset explanation similar to the one proposed by Gorissen et al. (2024). Yet, this line of thought—suggesting a larger negative footprint illusion when participants approach the task with a quantitative mindset—is inconsistent with Gorissen et al.'s (2024) suggestion that the negative footprint illusion should rather be smaller with a quantitative mindset. Furthermore, the idea that a mindset focused on numerical processing and calculation should produce a *larger* rather than *smaller* negative footprint illusion is also at odds with previous

research suggesting the opposite (Holmgren et al., 2021; Sörqvist & Holmgren, 2022). In relation to comparing the magnitude of effect across the two scales, it should also be noted that the Bayes Factor favoured the null hypothesis at odds with the frequentist statistic and as such a theoretical interpretation of this result is only treated tentatively.

Let us instead consider the possibility that the stimuli used here trigger a more qualitative mindset rather than a quantitative mindset. The variety engendered through using multiple objects with varying carbon footprints (arguably a complex stimuli) in our study could favour the adoption of a qualitative mindset, irrespective of the judgment scale type. In contrast, the study by Gorissen et al. (2024), which utilised only two items without explicit visual cues communicating their environmental impact, may have fostered a clear-cut qualitative mindset only when a green scale was employed. If the stimulus sets used here triggers a qualitative mindset regardless of scale type, but the stimulus sets used by Gorissen et al. (2024) only triggers a qualitative mindset with a green scale, this could explain the different pattern of results across the two studies.

A key consideration in future research will thus be the nature of stimuli—whether they are easily quantifiable and comparable—which can modulate the cognitive processes responsible for computing environmental impact, and possibly interact with the type of judgment scale. The distinction between the types of stimuli (similar and quantifiable versus dissimilar and qualitative) could be an important factor in understanding how the negative footprint illusion manifests in different contexts. However, the emergence of a negative footprint illusion with the deployment of a quantitative/objective response scale (e.g. kilograms of CO_2; Sörqvist & Holmgren, 2022) as well as a qualitative/subjective response scale (e.g. carbon footprint estimates; Gorissen & Weijters, 2016) suggests the effect may not be strictly dependent on the adoption of qualitative reasoning or a qualitative mindset.

While judgment scale could possibly interact with stimulus characteristics to produce different magnitudes of the negative footprint illusion, we propose in contrast to Gorissen et al. (2024) that the magnitude of the negative footprint illusion is not highly dependent on the judgment scale, and the reason for this is attribute substitution (Kahneman & Frederick, 2001). There is a clear resemblance between the pattern of results obtained with the two judgment scales used here and the judgments of carbon footprint used in the typical negative footprint illusion paradigm (e.g. Gorissen & Weijters, 2016; Holmgren et al., 2018). This suggests that a driving force behind the negative footprint illusion is indeed attribute substitution (Kahneman & Frederick, 2001), whereby participants replace the more difficult question ("how large is the carbon footprint of these items?") that requires them to assign a carbon footprint value to the objects with the easier question ("how environmentally bad are these items?") that requires them to evaluate the object on a good vs. bad continuum and assign a value based on that instead (cf. Holmgren et al., 2018). This assumption also coheres with the idea that many judgments under uncertainty are guided by an affect heuristic (Slovic et al., 2007)—the tendency to base judgments on the perceived positive ("goodness") or negative ("badness") qualities of a stimulus instead of their quantitative properties, often at the expense of more analytical information processing (cf. Evans & Stanovich, 2013a, 2013b).

Attribute substitution could explain why the negative footprint illusion seems to be driven by an averaging bias (Holmgren et al., 2018)—that is, the tendency to assign the average carbon footprint rather than the sum to combination of items. For example, if the task is to estimate the carbon footprint of a set comprising high and low carbon footprint houses, but the participants replace the task with estimating how environmentally bad the items are instead, then assigning a lower value to the combined set of low and high carbon footprint houses in comparison with the high carbon footprint houses alone becomes a true (normatively accurate) response. As the average of "environmentally good" and "environmentally bad" items would correspond to something like "intermediate environmental impact", this response can in this circumstance be seen as a true response. When participants adopt this interpretation, they can be thought of as providing a normative response to a problem that is different from what was intended in the study (Toplak, 2021). Indeed, the averaging bias can also explain related phenomena seen in studies on environmental impact estimates such as the quantity insensitivity effect (Kim & Schuldt, 2018; Kusch & Fiebelkorn, 2019). For example, environmental impact estimates of a set comprising two low-carbon footprint items (two hybrid cars) are no different from estimates of one low-carbon footprint item (one hybrid car) (Kim & Schuldt, 2018). This quantity insensitivity effect can be perfectly explained by the averaging bias account, as the average of two identical items is the same as the average of one item alone.

Yet, attribute substitution (and the averaging bias) seems not to offer a full explanation of the mechanisms driving the negative footprint illusion. For instance, attribute substitution by itself can hardly explain why stimulus distribution across the visual field influences the

magnitude of the effect (Sörqvist et al., 2022). An irregular stimulus distribution of the low carbon footprint items might increase the perceived environmental friendliness of the items, but it is unclear why the irregular distribution of the high carbon footprint items—following the same line of thought—does not compensate for this by increasing the perceived environmental unfriendliness of these items. Likewise, attribute substitution cannot explain why larger sets receive smaller carbon footprint estimates than smaller sets, even when the ratio between high carbon footprint and low carbon footprint items in the sets are held constant (Andersson et al., 2024). Larger sets of low carbon footprint items could increase the perceived environmental friendliness of the set, thereby leading to a lower carbon footprint estimate for these sets, but it is unclear why this would not be contradicted by the corresponding increase in the number of high carbon footprint items. One possibility is that there is a positivity bias involved, wherein more weight is assigned to the environmentally friendly than to the environmentally unfriendly items during the judgment formation process. When an irregular distribution makes the stimulus sets become perceived as larger, or when objects in the stimulus sets increase in number, this could interact with a positivity bias to produce a larger negative footprint illusion. This might be worthwhile exploring in future studies.

In conclusion, attribute substitution and the conceptual problems raised in the introduction section above stress the importance of a careful consideration of whether the selected response scale measures the negative footprint illusion, or whether it measures something else. In other words, if participants are not asked to estimate the carbon footprint of the items (or at least something related to the items' carbon footprint, e.g. CO_2 emissions or the number of trees needed to compensate for emissions), but instead how environmentally friendly the items are, is the negative footprint illusion then actually studied? In any case, participants' response pattern is highly similar when requested to evaluate the items' environmental friendliness or how environmentally damaging the items are. The current study shows that using a grey judgment scale does not attenuate the effect, which is consistent with a large body of past research (Sörqvist & Holmgren, 2022), and speaks against the policy implications of Gorissen et al.'s (2024) study. The negative footprint illusion appears to be relatively insensitive to the type of response scale used, but it is influenced by stimulus characteristics that may interact with what participants are asked to evaluate. Future studies on the negative footprint illusion could explore those interactions in more detail.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Andersson, H., Holmgren, M., Sörqvist, P., Threadgold, E., Beaman, C. P., Ball, L. J., & Marsh, J. E. (2024). The negative footprint illusion is exacerbated by the numerosity of environment-friendly additions: Unveiling the underpinning mechanisms. *Journal of Cognitive Psychology*, *36*(2), 295–307. https://doi.org/10.1080/20445911.2024.2313568

Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology*, *60*(4), 485–499. https://doi.org/10.1037/0022-3514.60.4.485

Chernev, A. (2011). The dieter's paradox. *Journal of Consumer Psychology*, *21*(2), 178–183. https://doi.org/10.1016/j.jcps.2010.08.002

Chernev, A., & Gal, D. (2010). Categorization effects in value judgments: Averaging bias in evaluating combinations of vices and virtues. *Journal of Marketing Research*, *47*(4), 738–747. https://doi.org/10.1509/jmkr.47.4.738

Evans, J. S. B. T., & Stanovich, K. E. (2013a). Dual-process theories of higher cognition: advancing the debate. *Perspectives in Psychological Sciences*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Evans, J. S. B. T., & Stanovich, K. E. (2013b). Theory and metatheory in the study of dual processing: reply to comments. *Perspectives in Psychological Science*, *8*(3), 263–271. https://doi.org/10.1177/1745691613483774

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Gorissen, K., & Weijters, B. (2016). The negative footprint illusion: Perceptual bias in sustainable food consumption. *Journal of Environmental Psychology*, *45*, 50–65. https://doi.org/10.1016/j.jenvp.2015.11.009

Gorissen, K., Weijters, B., & Deltomme, B. (2024). Green versus grey framing: Exploring the mechanism behind the negative footprint illusion in environmental sustainability assessments. *Sustainability*, *16*(4), 1411. https://doi.org/10.3390/su16041411

Holmgren, M., Andersson, H., Ball, L. J., & Marsh, J. E. (2021). Can the negative footprint illusion be eliminated by summative priming? *Journal of Cognitive Psychology*, *33*(3), 337–356. https://doi.org/10.1080/20445911.2021.1903012

Holmgren, M., Andersson, H., & Sörqvist, P. (2018a). Averaging bias in environmental impact estimates: Evidence from the negative footprint illusion. *Journal of Environmental Psychology*, *55*, 48–52. https://doi.org/10.1016/j.jenvp.2017.12.005

Holmgren, M., Kabanshi, A., Marsh, J. E., & Sörqvist, P. (2018b). When A + B A: Cognitive bias in experts' judgment of environmental impact. *Frontiers in Psychology*, *9*, Article 823, 1–6. https://doi.org/10.3389/fpsyg.2018.00823

Kahneman, D., & Frederick, S. (2001). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive judgment* (pp. 49–81). Cambridge University Press.

Kim, B., & Schuldt, J. P. (2018). Judging the environmental impact of green consumption: Evidence of quantity insensitivity. *Journal of Environmental Psychology*, *60*, 122–127. https://doi.org/10.1016/j.jenvp.2018.10.005.

Kusch, S., & Fiebelkorn, F. (2019). Environmental impact judgments of meat, vegetarian, and insect burgers: Unifying the negative footprint illusion and quantity insensitivity. *Food Quality and Preference*, *103731*, 1–10. https://doi.org/10.1016/j.foodqual.2019.04.003

Różycka-Tran, J., Boski, P., & Wojciszke, B. (2015). Belief in a zero-sum game as a social axiom: A 37-nation study. *Journal of Cross-Cultural Psychology*, *46*(4), 525–548. https://doi.org/10.1177/00220221155722

Slovic, P., Finucane, M. L., Peter, E., & MacGregor, D. G. (2007). The affect heuristic. *European Journal of Operational Research*, *177* (3), 1333–1352. https://doi.org/10.1016/j.ejor.2005.04.006

Sörqvist, P., Colding, J., & Marsh, J. E. (2020). Psychological obstacles to the efficacy of environmental footprint tools. *Environmental Research Letters*, *15*(9), 091001. https://doi.org/10.1088/1748-9326/ab9968

Sörqvist, P., & Holmgren, M. (2022). The negative footprint illusion in environmental impact estimates: Methodological considerations. *Frontiers in Psychology*, *13*, 990056. https://doi.org/10.3389/fpsyg.2022.990056

Sörqvist, P., Volna, I., Zhao, J., & Marsh, J. E. (2022). Irregular stimulus distribution increases the negative footprint illusion. *Scandinavian Journal of Psychology*, *63*(5), 530–535. https://doi.org/10.1111/sjop.12829

Threadgold, E., Marsh, J. E., Holmgren, M., Andersson, H., Nelson, M., & Ball, L. J. (2021). Biased estimates of environmental impact in the negative footprint illusion: The nature of individual variation. *Frontiers in Psychology*, *12*, Article 648328, 1–16.

Toplak, M. E. (2021). *Cognitive sophistication and the development of judgment and decision-making*. Academic Press.