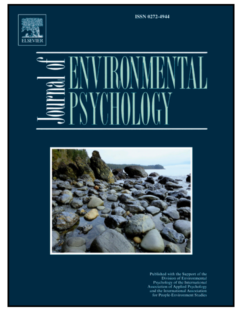
Journal Pre-proof



When More is Less: Cognitive Bias from Adding Recyclable Products in Green Consumption

Sarah Ormesher, Kathleen C. McCulloch, Beth H. Richardson, Patrik Sörqvist, John E. Marsh

PII: S0272-4944(26)00184-2

DOI: https://doi.org/10.1016/j.jenvp.2026.103083

Reference: YJEVP 103083

To appear in: *Journal of Environmental Psychology*

Received Date: 16 January 2026

Revised Date: 27 April 2026

Accepted Date: 19 May 2026

Please cite this article as: Ormesher, S., McCulloch, K.C., Richardson, B.H., Sörqvist, P., Marsh, J.E., When More is Less: Cognitive Bias from Adding Recyclable Products in Green Consumption, *Journal of Environmental Psychology*, https://doi.org/10.1016/j.jenvp.2026.103083.

When More is Less: Cognitive Bias from Adding Recyclable Products in Green Consumption

Sarah Ormesher[1], Kathleen C. McCulloch[1], Beth H. Richardson[1], Patrik Sörqvist[2,3], & John E. Marsh[1,2,4]

[1]Human Factors Laboratory, School of Psychology and Humanities, University of Lancashire, Preston, U.K

[2]Department of Health, Learning and Technology, Luleå University of Technology, Luleå, Sweden

[3]Department of Building Engineering, Energy Systems, and Sustainability Science, University of Gävle, Gävle, Sweden

[4]Faculty of Society and Design, Bond University, Gold Coast, Australia

Author note

Sörqvist was supported by a grant from Stiftelsen Riksbankens Jubileumsfond (P23-0067)

Correspondence:

John E. Marsh

Room 1.08 Darwin Building,

School of Psychology and Humanities

University of Lancashire

Marsh Lane

Preston, Lancashire

PR1-2HE

E-mail: jemarsh@lancashire.ac.uk

Abstract

The negative footprint illusion (NFI) refers to the counterintuitive tendency to judge a mixture of environmentally friendly and conventional items as producing less total impact than the conventional items alone. Although typically attributed to intuitive averaging processes, little is known about the conditions under which this bias emerges. Across two experiments using housing and packaging stimuli in recycling-relevant judgment contexts, the NFI appeared reliably under conditions that reduced support for explicit summation, most clearly at larger set sizes, but was absent when additive evaluation remained tractable. The effect generalised across item categories and domains and persisted even when recyclable and non-recyclable materials belonged to different categories, such as bottles and cans. Affective evaluations (valence and arousal) and behavioural intentions did not predict individual differences in illusion magnitude, providing little support for affective-halo or moral-compensation interpretations. Instead, the findings support a cue-integration account in which “green” and “brown” signals are combined into a single composite impression rather than summed. These results extend the NFI to recycling judgments and demonstrate that intuitive averaging processes can distort perceived environmental impact in everyday decision-making.

Keywords: negative footprint illusion, averaging bias, environmental judgment, cue integration, recycling, affective evaluation.

Significance Statement:

By showing that the negative footprint illusion emerges in recycling contexts and under conditions that trigger heuristic averaging, this study points to cognitive averaging, rather than moral or affective evaluation, as the key mechanism driving biased environmental judgments.

When More is Less: Cognitive Bias from Adding Recyclable Products in Green Consumption

Human judgments of environmental impact are often disconnected from physical reality. A particularly compelling example of this is the negative footprint illusion (NFI). This refers to the tendency to judge a combination of high carbon footprint and low carbon footprint items as producing less total carbon footprint than the high carbon items alone. First reported in the domain of food (Gorissen & Weijters, 2016) and housing (Holmgren et al., 2018a), the NFI exemplifies how the heuristic integration of positive ("green") and negative ("brown") environmental cues can distort quantitative reasoning about sustainability (Sörqvist et al., 2020). Here, "green" refers to items characterised as relatively low in environmental impact (e.g., recyclable or environmentally certified), whereas "brown" refers to conventional items associated with higher environmental impact. Because environmental judgments underlie our everyday choices, such as what to buy (Sörqvist et al., 2015), recycle (van Doorn & Kurz, 2021), or discard (Nadricka et al., 2024), the NFI provides a window into the cognitive mechanisms that influence and guide our sustainable behaviours. The present research examines whether the NFI extends to recycling judgments, specifically whether combinations of recyclable and non-recyclable items are judged as less environmentally damaging than non-recyclable items alone, and tests competing explanations for the NFI effect.

Several theoretical accounts have been advanced to explain the NFI. The leading proposal is an averaging-bias account. This holds that people intuitively average across positive and negative environmental cues rather than summing them (Andersson et al., 2024; Holmgren et al., 2018a). When a "green" product is added to a "brown" one, the addition of a green product with a lower environmental impact reduces the perceived mean environmental impact of the set, thereby yielding the paradoxical impression that total emissions decrease. This interpretation is consistent with evidence that people often replace difficult summation problems with simpler, impression-based evaluations when integrating mixed information. In multi-attribute judgment, this frequently yields mean-like integration rules rather than additive ones, producing systematic underestimation whenever positive and negative cues co-occur.

This interpretation aligns with the broader heuristic tradition showing that people often replace difficult quantitative evaluations with simpler, impression-based judgments grounded in representativeness (Tversky & Kahneman, 1974). In multi-attribute contexts, the resulting integration rule is frequently *mean-like* rather than additive (Anderson, 1981; Birnbaum, 1974). In practical terms, this means that instead of mentally adding together the impacts of each element in a set (summation), people tend to form a single overall impression that reflects the average "quality" of the items. When positive and negative cues coexist, averaging lowers the perceived overall impact relative to true addition, because low-impact ("green") elements dilute the perceived severity of high-impact ("brown") ones. Such averaging tendencies are consistent with broader models of judgment integration. For example, range-frequency theory (Parducci, 1965) proposes that people evaluate stimuli relative to the distribution of values encountered in the immediate context rather than in absolute terms. Judgments therefore reflect how an item compares to others in the set, not simply its objective magnitude. Similarly, configural-weight models (Birnbaum, 1973) suggest that when combining multiple pieces of information, people assign different subjective weights to components depending on how they are configured and often compress extremes toward a central tendency. In both approaches, complex information is reduced to a simplified representation that prioritises relative standing or overall impression rather than strict numerical addition. Collectively, these perspectives centre on the idea that environmental judgments may rely on impression-based integration rules that privilege averages or relative standing over exact totals.

A second suite of explanations for the NFI emphasises affective and moral processes. For example, according to an affective-halo view, the presence of an eco-label or green colouration evokes positive affect that generalises to the entire set (Gorissen et al., 2024; Kahneman & Frederick, 2002; Threadgold et al., 2022). Relatedly, the compensatory-belief or moral-licensing account proposes that people believe environmentally virtuous actions can offset environmentally harmful ones (MacCutcheon et al., 2020). In this interpretation, the NFI reflects motivated reasoning or self-justification rather than purely an averaging process.

All three explanations (the cognitive-averaging account, the affective-halo view and the moral-licensing account) also echo more general accounts of heuristic substitution and bounded rationality in judgment (Gigerenzer & Todd, 1999; Kahneman & Frederick, 2002), where difficult quantitative evaluations are replaced by qualitatively meaningful impressions. For example, when asked to estimate the carbon footprint (CF) of a set of high-CF items and low-CF items, the person making the judgment can (subconsciously) replace the task with a simpler task of estimating the items' environmental friendliness instead. A mixed set containing both high-CF and low-CF items is, in fact, more environmentally friendly on average than a set containing only high-CF items. If participants substitute the instructed question ("What is the total carbon footprint?") with an easier one ("How environmentally friendly does this set seem overall?"), then reporting the average becomes a normatively correct response to the substituted task. However, it remains normatively incorrect for the original summation task, which requires estimating total rather than average impact (Sörqvist & Marsh, 2024). This illustrates how attribute substitution can underlie the cognitive-averaging account.

Although the three accounts outlined above can explain why adding a "green" item makes a set appear less environmentally harmful, they make different and therefore competing predictions. If the NFI arises from averaging within a single cognitive representation of environmental impact, then it should appear consistently across domains and remain unrelated to affective valence or arousal. By contrast, if it stems from affective or compensatory processes, the magnitude of the NFI should covary with emotional evaluations or pro-environmental attitudes.

Some evidence supports the cognitive (e.g., averaging) interpretation. The illusion has been replicated across diverse items including buildings (Holmgren et al., 2018a), food (Gorissen & Weijters, 2016), vehicles (Holmgren et al., 2021), and temporal consumption sequences (Holmgren et al., 2021). This suggests that it is not bound to any one domain. Recent methodological refinements have further shown that the effect is robust to response scale type and experimental design. Using ratio-scale CO_2 estimates, Sörqvist and Holmgren (2022) found reliable NFI effects even when participants estimated emissions in kilograms and within-participant comparisons were possible. This indicates that the illusion is not an artifact of ordinal rating scales or between-group designs (see also Holmgren et al., 2018b). Moreover, the effect is slightly larger with brief than with extended response deadlines, which implies that a more intuitive judgment produces a larger effect (Sörqvist & Holmgren, 2022). Combined, these findings support the view that NFI reflects a basic, automatic cognitive process.

Spatial and numerical manipulations have also revealed systematic patterns consistent with cognitive averaging. Sörqvist et al. (2022) observed that when environmentally certified buildings were intermixed with conventional ones rather than grouped separately, perceived impact decreased further. This effect was attributed to greater salience of the green subset and seems difficult to reconcile with affective accounts. Similarly, increasing the numerosity of green items among a constant number of brown items exaggerates the illusion (Andersson et al., 2024). These regularities indicate that the cognitive system integrates category cues roughly proportionally to their frequency.

Such numerosity-driven integration echoes the statistical averaging found in perceptual and judgmental tasks (Ariely, 2001; Haberman et al., 2009) akin to the representational and integrative difficulty imposed by large numerosity sets. Large sets do not merely make judgments harder in a general sense, rather they reduce the availability of explicit summation as a viable strategy. Under such conditions, judgments are more likely to rely on ensemble-like representations of the set's overall environmental "tone," consistent with evidence that people extract summary statistics from groups of items rather than computing totals (Ariely, 2001; Whitney & Yamanashi Leib, 2018). In this sense, integrative difficulty reflects not general processing demand, but the extent to which the task supports or discourages item-wise addition.

Building on these principles, an important question concerns when such averaging processes are most likely to be engaged. Although prior research shows that numerosity can accentuate the illusion (Andersson et al., 2024), large set size should not be interpreted as a necessary or sufficient cause of the effect. Rather, numerosity appears to function as one reliable trigger of integrative difficulty within a broader class of task conditions that promote heuristic processing. When the task structure provides fewer cues that support item-wise summation, for example because set sizes are large or because quantity variation is not made explicit, individuals appear to substitute the difficult summation task with a simpler impression-based judgment of overall environmental "greenness." In this sense, the emergence of the NFI is best understood as reflecting a threshold process whereby averaging becomes more likely when conditions favour set-level evaluation over item-wise addition, but similar reliance on averaging may also occur under other conditions that increase representational difficulty, such as coarse response scales, ambiguous anchors, or evaluative framing that encourages holistic rather than additive judgments. Consequently, numerosity should be treated as one important, but not exclusive, pathway through which averaging is triggered.

The observed convergence across paradigms has positioned the NFI as a benchmark phenomenon for studying cognitive integration of qualitative and quantitative information. Nevertheless, several questions remain about the extent to which affect and moral evaluation contribute to the NFI. Threadgold et al. (2022) examined individual-difference correlates of the NFI, including moral and cognitive reasoning styles, but found no evidence that affective or belief-based tendencies reliably modulate the illusion. In contrast, MacCutcheon et al. (2020) demonstrated that individuals who endorsed compensatory green beliefs (e.g., the idea that "eco-friendly" actions offset harmful ones) showed a stronger NFI. Furthermore, Gorissen et al. (2024) found stronger effects in participants with strong environmental values. The evidence is thus mixed on the topic of whether there is some role for motivational factors in the NFI.

Recycling decisions provide a particularly compelling test case. Recycling is widely regarded as the quintessential "green" behaviour (Whitmarsh & O'Neill, 2010). Recent survey research shows that many people overestimate its environmental benefits relative to waste reduction or reuse (Barnett et al., 2023). This public "recycling bias" parallels the logic of the NFI. Individuals appear to treat the act of recycling as offsetting the creation of waste, rather than as one stage in a larger material flow. However, despite its conceptual closeness, the NFI has never been examined in recycling contexts where recyclable and non-recyclable items coexist. Investigating this domain not only presents a way to test whether the illusion generalises to everyday waste judgments; it also allows the disentangling of cognitive from affective explanations.

The present research had three aims. First, we tested whether the NFI generalises to recycling judgments, specifically, whether combinations of recyclable and non-recyclable items are judged to have lower total environmental impact than non-recyclable items alone. Second, we examined whether the illusion is moderated by item category, comparing within-category bundles (e.g., bottles + bottles) to cross-category bundles (e.g., bottles + cans). Third, we evaluated competing explanations by incorporating affective evaluations (valence and arousal) and a behavioural-intention proxy (likelihood of buying) for a subset of participants in Experiment 2. These measures were included because affective-halo and moral-compensation accounts propose that impact estimates are partly influenced by evaluative responses to “green” versus “brown” items: participants who show stronger positive evaluation and/or stronger behavioural preference for recyclable items should also show a larger downshifting of perceived total impact when recyclable items are added (i.e., a larger NFI). In contrast, a cognitive averaging account predicts that the magnitude of the NFI depends primarily on integrative demands and cue integration, and therefore should be weakly related (or unrelated) to individual differences in affective tone or buying likelihood.

To preview, across two experiments, participants estimated the total environmental impact of sets of bottles and cans varying in recyclability. The NFI emerged reliably in Experiment 1 and selectively under high-quantity conditions in Experiment 2, revealing boundary conditions under which averaging is engaged. Further, affective ratings failed to predict the size of the NFI. These findings support the view that the NFI arises from a domain-general cognitive averaging process rather than affective or moral influences. Through extending the paradigm to recycling, the present study demonstrates that intuitive cue-integration mechanisms, originally documented for energy use and carbon estimates (Holmgren et al., 2021; Sörqvist & Holmgren, 2022), also influence everyday waste judgments. This reveals how simple cognitive operations can produce systematically biased perceptions of environmental impact. By testing the NFI in a recycling context, the present research provides a behavioural testbed for understanding how intuitive averaging processes, akin to those observed in ensemble perception (Ariely, 2001; Haberman et al., 2009) and moral evaluation (Hsee, 1998), influence environmental judgments. Critically, the two experiments differed not only in numerosity, but in how strongly the task structure made quantitative relations between baseline and addition sets explicit, allowing us to test whether the NFI depends on the availability of additive reasoning rather than set size alone.

Experiment 1

The first experiment provided an initial test of whether the NFI generalises to recycling judgments across two distinct item categories (houses and bottles). Participants estimated the total environmental impact of sets of items that varied in recyclability. Following the averaging-bias account, we predicted that combinations of recyclable and non-recyclable containers would be judged as producing less total impact than non-recyclable containers alone. If the NFI primarily reflects a cognitive averaging process, it should emerge consistently regardless of the category of item. Experiment 1 offers a foundational demonstration of the NFI in a recycling context, establishing the baseline pattern that Experiment 2 builds upon to assess whether affective evaluations contribute to the illusion.

Method

Design

The experiment had a 2 (Item Type: houses, bottles) × 2 (Set Type: no addition, addition) × 2 (Order: houses then bottles, bottles then houses) mixed factorial design. Item Type and Set Type were manipulated within participants, while Order was a between-participant variable. The dependent variable (DV) was participants' estimates of carbon footprint / environmental impact on a 1-100-point scale. Counterbalancing was used to present participants with either houses or bottles first.

Participants

Participants were recruited via Prolific (www.prolific.com). A power analysis using G*Power (Faul et al., 2007) was conducted for a paired-sample t-test comparing estimates in the "no addition" and "addition" conditions. To detect a medium effect size (*dz* = 0.50) with 95% power (α = .05), a minimum of 54 participants was required. To allow for potential exclusions, 64 participants were recruited (*M*_age = 34.5 years, *SD* = 7.7, 55% female, all native English speakers). This final sample size is adequate to detect a medium effect size (*dz* = 0.5) with 97.5% power, or a modest effect size of *dz* = 0.356 with 80% power. All participants self-reported English as their first language. The study took approximately 15 minutes, and participants were compensated £2.25. The study was approved by the University of Lancashire's Science Ethics Panel (SCIENCE 0202). The study was not preregistered, but all data are available on the Open Science Framework (OSF; https://doi.org/10.17605/OSF.IO/2JN7U).

Materials and Procedure

The experiment was conducted online via Prolific, with participants redirected to Qualtrics (Provo, UT). Participants first read an information sheet outlining the study before providing informed consent. They then completed a demographics form (age, gender) before proceeding to the main task, which consisted of environmental impact estimations for houses and plastic bottles.

In the houses condition (Figures 1 and 2), participants were instructed to estimate the carbon footprint / environmental impact of houses using a 1-100 scale (higher scores indicate higher carbon emissions). They were given the following definition:

Carbon footprint refers to emissions arising from electricity consumption, heating, and other household activities. A higher score indicates greater environmental impact, while a lower score indicates reduced impact.

To standardise responses, participants were provided with a reference point:

A building of 18 apartments would score 50 (midway) on the scale.

They were then presented with the first condition (15 conventional houses) and estimated the carbon footprint.

[Insert Figure 1**]**

Non-Addition Condition (Houses):

Illustration of the stimulus used in the non-addition condition for houses. Participants viewed an image of 15 conventional houses, representing standard housing with no environmental certification, and estimated their carbon footprint / environmental impact on a 1-100 scale.

Next, they saw the "addition" condition, where environmentally certified houses were introduced: 15 conventional houses + 15 environmentally certified houses. Participants were informed that environmentally certified houses are built using sustainable materials and energy-efficient technology, reducing their carbon footprint.

[Insert Figure 2]

Addition Condition (Houses):

Illustration of the stimulus used in the addition condition for houses. Participants viewed an image of 15 conventional houses plus 15 environmentally certified houses, where the latter were designed with sustainable materials and energy-efficient technology. Participants estimated the overall carbon footprint / environmental impact.

Participants then made a second carbon footprint / environmental impact estimate using the same 1-100 scale.

For the bottle condition (Figures 3 and 4), a similar procedure was followed for plastic bottles. Participants were instructed that carbon footprint estimations referred to the environmental impact of bottle production and disposal. The reference point was:

> *A shelf containing 18 single-use, non-recyclable plastic yoghurt tubs would score 50 (midway) on the scale.*

They were then presented with the first condition: 15 single-use, non-recyclable plastic bottles, and participants estimated the carbon footprint / environmental impact.

[Insert Figure 3]

Non-Addition Condition (Plastic Bottles):

Illustration of the stimulus used in the non-addition condition for plastic bottles. Participants viewed an image of 15 yellow bottles, representing single-use, non-recyclable plastic bottles, and were asked to estimate their carbon footprint / environmental impact on a 1-100 scale.

In the "addition" condition, they were informed that recyclable bottles were added: 15 single-use, non-recyclable plastic bottles + 15 environmentally certified plastic bottles (*Environmentally certified bottles are recyclable and made from recycled materials, reducing their carbon footprint and environmental impact*).

[Insert Figure 4]

Addition Condition (Plastic Bottles):

Illustration of the stimulus used in the addition condition for plastic bottles. Participants viewed an image of 15 yellow bottles (single-use, non-recyclable plastic bottles) plus 15 green bottles (recyclable plastic bottles) and estimated the overall carbon footprint / environmental impact.

Participants then provided a second estimate on the 1-100 scale.

Results

The analysis begins with focusing on the negative footprint illusion (NFI) for the two item types; houses and bottles. A 2 (Item Type: house, bottles) × 2 (Set Type: no addition, addition) × 2 (Order: houses à bottles, bottles à houses) mixed analysis of variance (ANOVA) with estimated carbon footprint / environmental impact score as dependent variable was conducted, to test whether environmental impact ratings for plastic bottles would exhibit the NFI similarly to houses. The main effect of set type (no addition vs addition) was significant, $F(1, 61) = 10.20$, $MSE = 469.71$, $p = .002$, $\eta^2p = .143$, with lower estimates in the addition condition ($M = 55.64$, $SD = 21.73$) compared to the no addition condition ($M = 64.45$, $SD = 17.99$), replicating the typical NFI. Importantly, there was no significant interaction between item type and set type, $F(1, 61) = 0.10$, $MSE = 46.86$, $p = .754$, $\eta^2p = .002$, suggesting that the magnitude of the NFI was similar in size for the two item types.

The main effect of item type (houses vs. bottles) was not significant, $F(1, 61) = 0.023$, $MSE = 278.16$, $p = .879$, $\eta^2p < .001$, indicating that environmental impact estimates did not differ between houses ($M = 60.21$, $SD = 19.69$) and bottles ($M = 59.89$, $SD = 20.02$). Across item types, the NFI was of moderate magnitude, $t(61) = 3.19$, $p = .002$, $dz = 0.41$, with Bayesian analysis providing strong evidence for the effect ($BF_{10} = 10.42$). Follow-up analyses showed that the size of the NFI was comparable for houses (no-addition: *M* = 64.47, *SD* = 16.60; addition: *M* = 55.90, *SD* = 22.78), *t*(61) = 2.97, *p* = .004, *dz* = 0.38, BF_{10} = 5.85, and bottles (no-addition: *M* = 64.42, *SD* = 19.37; addition: *M* = 55.35, *SD* = 20.68), *t*(61) = 3.12, *p* = .003, *dz* = 0.39, BF_{10} = 8.47.

For control purposes, we also analysed effects of Order (estimates of houses first vs estimates of bottles first). The main effect of order was non-significant $F(1, 60) = 0.28$, *MSE* = 812.25, *p* = .596, η^2p = .005, indicating that whether participants completed the houses or bottles condition first had no effect on environmental impact estimations. Similarly, none of the interactions involving order reached significance: Order × Item type, $F(1, 60) = 0.09$, *MSE* = 282.35, *p* = .760, η^2p = .002; Order × Set Type, $F(1, 60) = 0.26$, *MSE* = 475.49, *p* = .613, η^2p = .004; Order × Item type × Set Type, $F(1, 60) = 3.43$, *MSE* = 45.06, *p* = .069, η^2p = .054.

Discussion

The results of Experiment 1 provide clear evidence that the NFI extends to a recycling context. Participants consistently judged sets containing a mix of recyclable and non-recyclable containers as producing less total environmental impact than sets composed entirely of non-recyclable containers, despite both sets containing the same number of high-impact items. This pattern replicates the core NFI observed in previous domains such as buildings and food (e.g., Gorissen & Weijters, 2016; Holmgren et al., 2018a) and demonstrates that the illusion generalises to consumer waste materials that people encounter in everyday life. Although houses and bottles were not combined within the same stimulus set, the replication of the NFI across these categorically unrelated domains indicates that the averaging process is not tied to a particular object class but emerges independently of item type.

Averaging accounts provide a straightforward explanation of this finding. When estimating total environmental impact, participants appear to form a single composite impression of the set and judge this average rather than summing the environmental burdens of its components. The presence of environmentally “friendly” items (recyclable containers) thus lowers the perceived overall impact in a way that is inconsistent with the true additive structure of emissions. This interpretation aligns with longstanding evidence that people frequently apply intuitive averaging when combining mixed positive and negative attributes (e.g., Anderson, 1981; Birnbaum, 1974) and with recent work showing that environmental judgments are similarly susceptible to mean-based integration processes (Holmgren et al., 2018a, 2018b; Sörqvist & Marsh, 2024).

The recycling domain provides a meaningful context in which to consider how cognitive heuristics may guide environmental judgment. Recycling is widely viewed as a pro-environmental behaviour, and recyclable materials often carry normative or conceptual cues of “greenness.” Such cues may feed into the representational space on which averaging processes operate, even when no explicit affective reaction is measured. That the illusion emerges robustly in this domain suggests that cognitive integration of “green” and “brown” cues can distort perceived environmental impact in everyday decision-making.

At an applied level, these findings highlight a potential cognitive barrier to accurate environmental reasoning. When people encounter mixed sets of waste or products, the inclusion of recyclable or eco-friendly elements may unintentionally downweight the perceived harm of the overall set. This may help explain why consumers often overestimate the benefits of recycling relative to waste reduction and why organisations sometimes overemphasise marginal sustainability improvements alongside more impactful conventional practices. Interventions that encourage explicit summation, such as prompting individuals to consider total emissions rather than overall “greenness”, may therefore reduce susceptibility to the NFI (Holmgren et al., 2021) and support more accurate environmental decision-making.

Because Experiment 1 did not include affective or behavioural measures, it cannot determine whether emotional evaluations or purchase tendencies contribute to the NFI. The recyclable items used here may have carried implicit positive associations, but no direct assessment of affect or buying likelihood was obtained. Experiment 1 therefore establishes the existence of the illusion in a recycling context but does not test competing explanations. Experiment 2 addresses these issues by incorporating a broader factorial manipulation of quantity and category, and by including, for a subset of participants, direct assessments of valence, arousal, and buying likelihood to examine whether these evaluations predict susceptibility to the illusion.

Experiment 2

Although Experiment 1 demonstrated that the NFI occurs in recycling judgments, it does not reveal whether the cognitive averaging process depends on items belonging to a single coherent category. Averaging mechanisms rely on the integration of multiple elements within a unified representational set (Anderson, 1981; Ariely, 2001). If people encode recyclable and non-recyclable items of the same category as forming a single set, a composite “environmental impact” value can be computed, producing the illusion. However, if category boundaries, such as bottles versus cans, segment the set into distinct mental groups, separate averages may be computed, reducing or eliminating the NFI. To date, the NFI has been documented primarily for within-category combinations (e.g., organic vs. non-organic food: Gorissen & Weijters, 2016; “green” vs. conventional buildings: Andersson et al., 2024; hybrid vs. petrol cars: Holmgren et al., 2021). Testing whether the NFI persists across categorical boundaries therefore provides a critical test of whether averaging is a domain-general integrative process or one constrained by perceptual or semantic grouping. Experiment 2 was designed to address this question.

Method

Design

The experiment employed a 2 (Item Type: cans, bottles) × 2 (Set Type: no addition, addition) × 2 (Category: no cross-category, cross-category) × 3 (Quantity of Base Set: 10 items, 20 items, 40 items) factorial design. The number of recyclable additions was held constant (20) to isolate the effect of baseline set size on integration strategy while keeping the "green" signal strength comparable across conditions. All variables were manipulated within-participants.

Given the complexity of the design, full counterbalancing of all experimental factors was neither feasible nor desirable. Instead, a constrained randomisation procedure was implemented. Participants were assigned to one of four predefined block-order rotations that varied the order in which item types (bottles vs. cans) were presented. In each rotation, same-category (pure) blocks were presented together, and cross-category (mixed) blocks were presented together, but the item type alternated within these higher-order blocks. This ensured that participants did not evaluate the same material type in immediate succession while maintaining a stable category structure within each block.

Within each block, the three quantity conditions (10, 20, and 40 baseline items) were presented in random order, with Qualtrics' "Evenly Present Elements" option enabled to ensure that quantities were approximately balanced across serial positions at the sample level. All participants completed all conditions, but order effects were minimised through block-order rotation and within-block randomisation. The dependent variable (DV) was the estimate of environmental impact given on a 1-100-point scale.

A subset of participants (51) additionally completed ratings of valence, arousal, and likelihood of buying for each item type (recyclable vs. non-recyclable bottles and cans). These measures were intended to determine whether affective responses or behavioural preferences predict susceptibility to the NFI.

Including these measures for only half of the sample served three purposes. First, it avoided contamination of the main judgment task. Completing detailed affective and evaluative ratings before the impact-estimation trials could prime participants to attend to particular stimulus dimensions (e.g., greenness, desirability), potentially exaggerating or suppressing the NFI. Similarly, the affective and evaluative ratings could potentially be influenced by the environmental-impact ratings if done after. To prevent systematic bias, only a randomly assigned subset completed these ratings after the environmental-impact tasks.

Second, it minimised participant burden and fatigue. The full factorial design of Experiment 2 was already demanding. Collecting affective and behavioural measures from all participants would have substantially lengthened the task, risking disengagement and increasing noise in impact estimates.

Third, using a subset of participants maximises validity while preserving statistical power. Because the affective/behavioural measures were used to test individual-difference predictions (i.e., whether affect predicts NFI magnitude), full-sample measurement was not required. The subset size was sufficient for correlational tests while keeping the primary NFI manipulation uncontaminated.

Participants

One hundred adults (53 female, 47 male) were recruited via Prolific (www.prolific.com), restricted to UK residents who were native English speakers and at least 18 years old, consistent with the criteria used in Experiment 1. A power analysis using PANGEA (Westfall, 2015) indicated that a total sample of 100 participants, producing 2,400 observations across the factorial design, would provide at least 95 percent power to detect medium within-person effects in the environmental-impact estimation task.

Because the design included multiple within-participant factors, the study was powered primarily to detect medium-sized within-person effects of set type and its interaction with quantity, which constituted the theoretically central tests of the integrative-difficulty account. Sensitivity analyses indicated that the achieved sample size provided high power (> .95) for detecting medium main effects and medium two-way interactions, whereas smaller higher-order interactions would require substantially larger samples. Consequently, null findings for higher-order interactions should be interpreted cautiously, and the conclusions of the present research focus primarily on the robust and theoretically predicted effects of set type and integrative difficulty. Whereas non-significant *p*-values indicate insufficient evidence to reject the null, Bayesian factors allow direct evaluation of whether the data provide support for the absence of predicted effects, which was particularly relevant for testing affective and higher-order moderation predictions.

Because the experiment additionally incorporated affective evaluations and likelihood-of-purchase ratings for approximately half of the participants, a second PANGEA analysis was conducted to ensure adequate sensitivity for this component. This analysis indicated that a subgroup of 51 participants would yield approximately 80 percent power to detect medium-sized within-person differences in valence and arousal ratings ($dz \approx .52$) at $\alpha = .05$.

Participants were therefore randomly assigned to one of two groups: a judgment-only group (25 female, 24 male), who completed the environmental-impact estimation task, and an extended-evaluation group (28 female, 23 male), who completed the same task followed by valence, arousal, and buying-likelihood ratings. This structure ensured sufficient power for both the primary NFI analyses and the affective subsidiary analyses while avoiding unnecessary task load for all participants. One participant who provided invariant default responses across all second estimates was excluded and replaced, preserving the intended sample size.

Participants varied in educational background. Twenty-seven percent held secondary-level qualifications (e.g., GCSEs, A Levels, NVQ Level 1–3), seven percent reported further-education or vocational certifications (e.g., diplomas, technical college, NVQ4), and two percent had attended university without completing a degree. Forty-five percent held an undergraduate degree (e.g., BA, BSc), fourteen percent held a postgraduate degree (e.g., MA, MSc), and five percent held a doctoral qualification. The distribution closely matches typical UK Prolific samples. The final sample of 100 participants had a mean age of 41.09 years (*SD* = 13.71; *M*_age = 42, *SD* = 13.71 judgement only; *M*_age = 42, *SD* = 14 judgement and rating). The study was approved by the University of Lancashire Science Ethics Panel (reference SCIENCE 0202). The study was not preregistered, but all data are available on the Open Science Framework (OSF; https://doi.org/10.17605/OSF.IO/2JN7U).

Materials and Procedure

The materials for the environmental-impact task followed the structure used in Experiment 1, with two key extensions: (a) the inclusion of cross-category combinations of items, and (b) the addition of affective and behavioural-evaluation measures for the subset of participants assigned to the extended-evaluation group. All tasks were presented online using Qualtrics.

Environmental-Impact Estimation Task

Participants first completed the main NFI task. On each trial, they viewed an array of items representing either cans or plastic bottles. In the first estimate ("no-addition" condition), participants judged the total environmental impact of a set of non-recyclable items that varied in quantity (10, 20, or 40 items). In the second estimate ("addition" condition), they judged the impact of a mixed set containing the same non-recyclable items plus 20 recyclable items. The recyclable additions were either from the same category (e.g., non-recyclable bottles + recyclable bottles) or a different category (e.g., non-recyclable bottles + recyclable cans), depending on the condition (see Table 1 for all of the conditions).

Stimuli were presented as arrays of items arranged in a grid, analogous to those used in Experiment 1. An example of a cross-category stimulus (non-recyclable cans with recyclable bottles added) is shown in Figure 5. Environmental impact was rated on a 1–100 scale, where higher scores indicated greater environmental impact. The order of quantities, categories, and item types was counterbalanced across participants.

[Insert Figure 5]

Illustration of the stimulus used in the cross-category cans 10 set. Participants viewed 10 non-recyclable cans, followed by 10 non-recyclable cans plus 20 recyclable plastic bottles and estimated their environmental impact on a 1-100 scale.

Rationale for behavioural-intention measures. Likelihood-of-buying ratings were included as an evaluative index that captures the motivational/behavioural attractiveness of recyclable versus non-recyclable items. Under affective-halo and moral-compensation interpretations, such evaluative preferences should be reflected in impact judgments: stronger preference for recyclable items (relative to non-recyclable) should predict a larger reduction in perceived total impact following the addition of recyclable items. Under a cognitive averaging account, buying likelihood may differ strongly between recyclable and non-recyclable items, but these preferences should not systematically predict the integration rule used to estimate total impact, and thus should show little relation to NFI magnitude.

Affective and Buying-Likelihood Measures (Extended-Evaluation Group Only)

Approximately half of the participants (n = 51) completed additional evaluative measures after finishing all environmental-impact estimates. These measures assessed valence (1 = happy, 9 = sad), arousal (1 = high arousal 9 = low arousal), and likelihood of buying (1 = very unlikely, 9 = very likely).

Each participant in the extended-evaluation group viewed four item types individually: recyclable cans, non-recyclable cans, recyclable bottles, and non-recyclable bottles. For each stimulus, participants rated how the item made them feel using the Self-Assessment Manikin scales for valence and arousal (Bradley & Lang, 1994), presented as pictorial figures on screen (Figure 6). They were informed that the valence scale ranged from 1 (happy) to 9 (sad), and the arousal scale ranged from 1 (high arousal) to 9 (low arousal), and that they could also use the boxes between the numerical anchors to provide more fine-grained judgments.

The order of presentation of the four item types was counterbalanced. These affective and buying-likelihood ratings were always completed after the environmental-impact task to avoid priming participants to focus on particular evaluative dimensions (e.g., “greenness” or desirability) during the main NFI judgments, and they were collected only from the extended-evaluation subgroup to minimise task burden and fatigue.

[Insert Figure 6]

Illustration of the valence and arousal judgment scales.

General Procedure

Participants accessed the study via a Prolific link and were redirected to Qualtrics. After reading the information sheet and providing consent, they completed demographic questions before beginning the environmental-impact estimation trials. Participants in the judgment-only group completed the task and were then debriefed. Participants in the extended-evaluation group proceeded to the affective and buying-likelihood ratings before debriefing. Total participation time was approximately 15–20 minutes for the judgment-only group and 20–25 minutes for the extended-evaluation group.

Results

Environmental-impact estimation task

As can be seen in Tables 2 and 3, a negative footprint illusion was found in estimates of within-category bundles as well as cross-category bundles, but the effect only emerged when set size was large enough. A 2 (Item Type: bottles, cans) × 2 (Set Type: no addition, addition) × 2 (Category Type: same-category, cross-category) × 3 (Quantity of Base Set: 10, 20, 40 items) repeated measures ANOVA was conducted. The dependent variable was participant's environmental impact estimation score on a 1-100 scale. The results from this ANOVA are summarized in Table 4.

In this complex analysis, we focus on the presence of the NFI and factors modulating this effect. The interaction between category type and set type was not significant, suggesting that the magnitude of the NFI was similar in size across the two types of bundles (same-category bundles vs cross-category bundles). In turn, a significant interaction emerged between set type and quantity, demonstrating that the magnitude of the NFI was modulated by the quantity of items. Also, the four-way interaction was weakly significant and arguably of less importance than the interaction between set type and quantity of base set. All other interactions were non-significant.

To further explore the effects of quantity and set type on environmental impact estimations, paired-sample *t*-tests were conducted. Given that category type did not have a significant main effect and did not interact with set type (Table 4), data were collapsed across categories to focus on the significant effects of item type and quantity. When 20 recyclable items were added to 40 non-recyclable items, participants' environmental impact estimations significantly decreased, $t(99) = 3.62$, $p < .001$, Cohen's $dz = 0.362$, $BF_{10} = 34.48$. This suggests that adding recyclable items led participants to downshift the total environmental impact, consistent with the NFI. When 20 recyclable items were added to 20 non-recyclable items, there was no significant difference in estimates, $t(99) = 0.30$, $p = .763$, Cohen's $dz = 0.03$, $BF_{01} = 12.097$. Similarly, when 20 recyclable items were added to 10 non-recyclable items, no significant difference was found, $t(99) = -1.453$, $p = .149$, Cohen's $dz = -0.145$, $BF_{01} = 4.50$.

Valence, arousal, and behavioural intention ratings

Valence ratings

There was a main effect of recyclability, $F(1, 50) = 41.21$, $MSE = 7.55$, $p = .001$, $\eta^2 p = .452$. Recyclable items ($M = 3.71$, $SE = 0.24$) were rated more positively (1 = happy, 9 = sad) than non-recyclable items ($M = 6.18$, $SE = 0.26$) in perceived valence. There was no main effect of item type, $F(1, 50) = 1.75$, $MSE = 1.35$, $p = .191$, $\eta^2 p = .034$. Cans and bottles were similarly rated (cans: $M = 4.83$, $SE = 0.17$; bottles: $M = 5.05$, $SE = 0.19$). There was no interaction effect for Recyclability

× Item type, $F(1, 50) = 0.49$, $MSE = 1.55$, $p = .576$, $\eta^2 p = .006$. Thus, perceived valence was primarily driven by recyclability rather than item type.

Arousal ratings

There was no main effect for recyclability, $F(1, 50) = 0.43$, $MSE = 7.13$, $p = .515$, $\eta^2 p = .009$. Recyclable and non-recyclable items were similarly rated (non-recyclable: $M = 5.52$, $SE = 0.33$; recyclable: $M = 5.77$, $SE = 0.28$). There was a main effect of item type, $F(1, 50) = 5.30$, $MSE = 1.01$, $p = .026$, $\eta^2 p = .096$. Bottles ($M = 5.49$, $SE = 0.27$) received lower ratings (higher arousal; 1 = high arousal 9 = low arousal) than cans ($M = 5.81$, $SE = 0.24$). There was no interaction effect for Recyclability × Item type, $F(1, 50) = 1.41$, $MSE = 1.01$, $p = .241$, $\eta^2 p = .027$. Participants showed stronger emotional engagement with item type but not with recyclability.

Behavioural intentions

There was a main effect of recyclability, $F(1, 50) = 160.42$, $MSE = 1.52$, $p < .001$, $\eta^2 p = .762$. Recyclable items received lower ratings (i.e., greater likelihood of purchase; 1 = very likely, 9 = very unlikely) than non-recyclable items (recyclable: $M = 2.00$, $SE = 0.11$; non-recyclable: $M = 4.20$, $SE = 0.33$). There was no main effect of item type, $F(1, 50) = 0.64$, $MSE = 0.19$, $p = .429$, $\eta^2 p = .013$, indicating similar buying-likelihood ratings for cans and bottles within each recyclability level (recyclable cans: $M = 2.04$, $SE = 0.13$; recyclable bottles: $M = 2.01$, $SE = 0.11$; non-recyclable cans: $M = 4.28$, $SE = 0.10$; non-recyclable bottles: $M = 4.11$, $SE = 0.09$). The Recyclability × Item type interaction was not significant, $F(1, 50) = 0.59$, $MSE = 0.30$, $p = .168$, $\eta^2 p = .038$.

Valence, arousal, and behavioural intention ratings as predictors of NFI magnitude

To assess whether affective evaluations predicted susceptibility to the NFI, valence and arousal difference scores were computed separately for bottles and cans. For each participant, affective contrast scores were created by subtracting the rating of the recyclable item from the rating of the non-recyclable item (i.e., difference score for valence ratings of bottles = valence rating of nonrecyclable bottles – valence rating of recyclable bottles). The same computation was applied to arousal ratings. Positive values reflect that non-recyclable items were evaluated more negatively or as more arousing than recyclable items, whereas values near zero indicate minimal affective differentiation between the recyclable and non-recyclable items.

Because the NFI emerged only in the high-quantity condition (40-item baseline with 20 recyclable items added), correlations between affective contrast scores and NFI magnitude were conducted exclusively for this condition. To ensure conceptual alignment between predictors and outcomes, affective difference scores for bottles were correlated with same-category bottle NFI scores, and affective difference scores for cans were correlated with same-category can NFI scores. Cross-category NFI values were not included because they combine material types and therefore cannot be meaningfully related to item-specific affective predictors.

NFI scores for bottles and cans were positively correlated ($r(49) = .602$, $p < .001$), indicating that individuals who showed a stronger illusion for one material tended to show it for the other as well.

In line with the competing-account logic outlined above, we tested whether individual differences in evaluative responses (valence, arousal, and buying likelihood contrasts between recyclable and non-recyclable items) predicted the magnitude of the NFI in the condition where the illusion was reliably observed. No significant correlations between valence/arousal and the magnitude of the NFI were found (Valence bottles: $r(49) = -.151$, $p = .291$; Valence cans: $r(49) = .063$, $p = .661$. Arousal bottles: $r(49) = -.191$, $p = .180$; Arousal cans: $r(49) = -.195$, $p = .170$). This indicates that affective responses did not predict susceptibility to the illusion. Positive correlations were found between valence and arousal scores for the same item types ($r(49) = .373$, $p = .007$ [bottles]; $r(49) = .226$, $p = .110$ [cans]) suggesting that participants who felt more positive about an item also tended to rate it as more arousing. There were also strong correlations between the valence scores across items ($r(49) = .660$, $p < .001$) and the arousal scores across items ($r(49) = .757$, $p < .001$).

A similar pattern emerged for the relation between behavioural intention and the NFI magnitude. Behavioral intention did not predict the magnitude of the NFI (cans: $r(49) = .097$, $p = .500$; bottles: $r(49) = .027$, $p = .851$). Thus, differences in behavioural intentions toward recyclable versus non-recyclable products did not predict individual susceptibility to the NFI.

Discussion

Experiment 2 examined two central questions: whether the negative footprint illusion generalises to estimates of item bundles comprising items from different categories (bottles versus cans) and whether the size of the item set influences the emergence of the effect. A further aim, tested in a subset of participants, was to determine whether affective evaluations or behavioural preferences predict susceptibility to the illusion.

A clear pattern emerged. Unlike Experiment 1, Experiment 2 did not produce a significant overall NFI across all conditions. Instead, the illusion appeared only when participants evaluated large sets containing 40 non-recyclable items at baseline. The addition of 20 recyclable items to smaller sets (10 or 20 non-recyclable items) did not reduce perceived environmental impact.

This quantity-dependent pattern is consistent with earlier work indicating that the NFI arises primarily when the task structure provides fewer cues for item-wise summation, as in displays with larger set sizes or in contexts that do not foreground exact quantity relations (Andersson et al., 2024; Sörqvist & Marsh, 2024). In such contexts, observers rely on approximate representations of set properties (e.g., mean or proportion) rather than computing exact totals (Ariely, 2001; Pelham et al., 1994; Whitney & Yamanashi Leib, 2018). When set size is small and the summation problem remains tractable, participants appear able to engage in more explicit or quasi-additive reasoning. By contrast, larger sets provide fewer practical cues for tracking item-wise totals and therefore make explicit summation less tractable or less likely to be spontaneously adopted (or at least trigger reliance on more intuitive thought processes), prompting heuristic averaging strategies in which the presence of recyclable items lowers the perceived mean "environmental impact" of the set. Whereas Experiment 1 provided a relatively holistic evaluative context without explicit variation in baseline numerosity that likely encouraged impression-based integration even at moderate set sizes, Experiment 2 made quantity structure explicit and therefore permitted additive reasoning when the task structure made additive reasoning readily available. Under this stronger quantitative framing, averaging was predicted to emerge selectively at higher quantities, which is precisely the pattern observed.

The apparent difference between Experiments 1 and 2 is better understood in terms of task structure than in terms of raw set size alone. In Experiment 1, participants made set-level judgments using a single reference point without any explicit manipulation of baseline numerosity. Under these conditions, the display could readily be encoded as a whole and evaluated in terms of its overall environmental meaning, making impression-based averaging a plausible basis for judgment. In contrast, Experiment 2 systematically varied the number of baseline items (10, 20, 40) while holding the recyclable addition constant. This repeated manipulation made quantitative structure more transparent and likely increased the salience of exact quantity, thereby supporting additive or quasi-additive reasoning when set sizes remained tractable. On this interpretation, the experiments do not conflict. Rather, Experiment 2 provided stronger support for summation, so averaging emerged only when numerosity increased to a level at which explicit addition became less feasible or less attractive (Ariely, 2001; Holmgren et al., 2021; Pelham et al., 1994). As an analogy, Experiment 1 encouraged judgments akin to assessing the overall "shade" of a display, whereas Experiment 2 made it easier to track how many items contributed to the total. Adding some "green" elements changes the overall impression immediately, even though it does not reduce the cumulative amount.

The quantity findings have important implications for interpreting the averaging mechanism that underpins the NFI. They suggest that the illusion is not an automatic response to mixed recyclable–non-recyclable sets but instead emerges selectively when the task demands exceed participants' willingness or ability to compute additive totals. This interpretation aligns with research on numerosity and summary-statistic processing (Ariely, 2001; Pelham et al., 1994), which demonstrates that judgments of large item sets are disproportionately influenced by perceived averages rather than sums. In the present context, as set size increases, individuals appear to substitute the challenging question "What is the total environmental impact?" with the more intuitive question "How green does this set feel overall?", generating the characteristic underestimation of mixed sets.

The categorical manipulation sheds further light on the structure of this averaging process. When analysing only the high-quantity condition, where the NFI reliably occurred, the effect was comparable for same-category (e.g., bottles + bottles) and cross-category (e.g., bottles + cans) configurations. This indicates that once heuristic averaging is engaged, the cognitive system integrates environmental cues across perceptually distinct object classes. However, because no NFI appeared at lower quantities, cross-category generalisation can only be meaningfully inferred when the averaging mechanism is active. In other words, category boundaries do not constrain the illusion, but they do not themselves trigger it; only under conditions of high integrative demand do cross-category cues combine into a unified environmental impression.

The extended-evaluation subset provided additional theoretically important evidence. Although recyclable items were rated as more positive and substantially more likely to be purchased than non-recyclable items, these affective or behavioural evaluations did not predict the magnitude of the NFI. Participants who showed strong preferences for recyclable products were no more or less susceptible to downshifting their estimates of the total environmental impact of mixed sets. The absence of associations between buying-likelihood contrasts and NFI magnitude is difficult to reconcile with affective-halo and moral-licensing accounts for which impact estimates are directly influenced by evaluative/motivational responses to recyclable items, but is consistent with the view that preferences coexist with, rather than determine, the averaging-based integration rule that produces the illusion. These findings converge with previous work showing that compensatory green beliefs and moral-reasoning tendencies are unreliable predictors of the NFI (Threadgold et al., 2022; cf. MacCutcheon et al., 2020).

Instead, the results serve to reinforce the interpretation of the NFI as a cognitive averaging phenomenon. When estimating total impact, participants appear to combine “recyclable” and “non-recyclable” cues into a single composite representation, judging its mean rather than its sum. Affect may be encoded, indeed, recyclable items were rated as slightly more arousing than non-recyclable ones, but these affective properties do not appear to be weighted in the computation of environmental impact. This dissociation suggests that affective impressions accompany the stimuli but do not influence the integration rule. Such selective cue weighting aligns with models of bounded rationality and heuristic substitution (Gigerenzer & Todd, 1999; Kahneman & Frederick, 2002), in which multiple attributes may be represented but only a subset guides judgment. Here, the dominant attribute is recyclability, treated as a conceptual cue to environmental friendliness and incorporated through averaging rather than addition.

In summary, Experiment 2 shows that the negative footprint illusion is not universal but instead emerges under conditions that promote summary-statistic processing. Crucially, large numerosity should be understood as one reliable trigger of such processing rather than a necessary condition, as averaging may also be induced at lower quantities under conditions that increase integrative difficulty or encourage qualitative evaluation (e.g., under time pressure, coarse scaling, or evaluative framing). When item sets are numerically large, participants appear to rely on heuristic averaging, treating recyclable and non-recyclable items as points along a shared conceptual dimension of environmental friendliness. Under these conditions, category boundaries cease to constrain judgment, and the illusion generalises to estimates of cross-category bundles. Crucially, affective impressions and behavioural preferences do not predict variability in the effect, indicating that the NFI reflects a context-sensitive cognitive averaging mechanism rather than emotional or motivational processes. This pattern helps explain why the illusion is robust across domains yet does not appear reliably under all conditions.

General Discussion

Across two experiments, the present research demonstrates that the NFI emerges reliably in recycling judgments, although not uniformly. Participants downshifted their estimates of total impact when “green” items were added to “brown” in Experiment 1 and in the high-quantity conditions of Experiment 2. This pattern indicates that the illusion is robust when the task structure encourages reliance on heuristic processing, particularly under higher-quantity conditions. In these contexts, participants downshifted their estimations following a green addition to the target set, regardless of category composition or stimulus variability.

The quantity-dependent pattern observed in Experiment 2 aligns with the mechanism proposed by Andersson et al. (2024), even though their design held the number of brown items constant and varied the size of the green addition, whereas ours held the size of the green addition constant and varied the number of brown items. Andersson et al. showed that when item sets become perceptually or numerically large, people shift from item-wise evaluation to proportion-based summary processing, relying on ensemble-like averaging rather than additive integration. Crucially, this mechanism does not depend on fixed ratios but on the cognitive demands imposed by set size. The present findings replicate this load-sensitive shift: when sets were small (10 or 20 items), summative reasoning remained tractable and no NFI emerged; when sets were large (40 + 20 items), participants appeared to encode the set's overall green–brown composition and apply averaging. Thus, the convergence with Andersson et al. lies in demonstrating the same underlying cognitive transition from additive processing to summary-statistic integration once numerosity exceeds a cognitive threshold. This pattern raises an important question for future research: would interventions that explicitly support summation, such as summation prompts or cumulative-impact displays, eliminate the NFI at high quantities, where averaging appears to dominate? At the same time, large numerosity should be treated as a reliable trigger of averaging rather than a necessary condition. The NFI can also emerge at lower quantities (see for example Gorissen & Weijters, 2016) when other features increase integrative difficulty or promote impression-based responding (e.g., time pressure, coarse or comparative response formats, weak or ambiguous anchors, or framing that encourages "overall greenness" rather than totals).

Importantly, the present findings do not imply that numerosity alone determines whether the NFI appears. Rather, numerosity interacted with task structure. Experiment 1 encouraged participants to treat each array as a whole and place it on an evaluative environmental-impact continuum, whereas Experiment 2 repeatedly foregrounded quantity by varying baseline numerosity while holding the recyclable addition constant. In this sense, Experiment 1 was more compatible with a qualitative, set-level judgment, while Experiment 2 made additive structure more visible. The critical distinction, therefore, is not that 15 + 15 is inherently more demanding than 20 + 20, but that these displays were embedded in different judgment contexts that either obscured or highlighted the possibility of summation. Instead, Experiment 1 did not provide participants with repeated exposure to varying quantities, and therefore did not foreground the additive structure of the task, whereas Experiment 2 explicitly manipulated quantity across trials, making summation a more available strategy.

The findings reported here extend the NFI beyond earlier domains such as buildings, food, and vehicles (Gorissen & Weijters, 2016; Holmgren et al., 2018a, 2021) and show that the bias also arises in an everyday sustainability context. A consistent interpretation is that the bias reflects a cognitive simplification strategy whereby mixed environmental information is collapsed into an averaged mental representation of “greenness.” This process explains why people often underestimate the environmental consequences of mixed or partly green product sets (Kim & Schuldt, 2018) and why small eco-friendly features can unduly reduce perceived harm (Gorissen et al., 2024). From an applied perspective, the findings highlight the value of communication strategies that emphasise additive totals rather than qualitative impressions, for example, by presenting cumulative carbon or waste metrics instead of eco-labels.

The results also support a cognitive-averaging account rather than an affective or moral-licensing account. Experiment 2 showed that averaging is not automatically applied but emerges when the task structure provides fewer cues for item-wise summation, for example when baseline quantity is large and the display encourages set-level rather than element-by-element evaluation, that is, when large item sets impose numerosity-driven integration difficulty that makes explicit summation impractical. Under these conditions, participants appear to integrate green and brown cues into a single mental representation of environmental impact and evaluate this representation by its mean rather than its sum. This echoes long-established findings in the judgment literature showing that individuals often rely on configural and averaging rules when integrating multi-attribute information (Anderson, 1981; Birnbaum, 1974). Related models of range–frequency scaling (Parducci, 1965) and configural-weight integration (Birnbaum, 1973) similarly propose that people compress multidimensional information into a representative average that offers computational efficiency at the expense of accuracy. By reducing a complex summation task to a simpler heuristic, cognitive demands are lowered but systematic underestimation is introduced whenever positive and negative elements co-occur.

The findings further contribute to debates on bounded rationality and heuristic substitution (Gigerenzer & Todd, 1999; Kahneman & Frederick, 2002). When faced with a complex integrative task such as estimating the combined footprint of mixed materials, people substitute an accessible qualitative question, for example, how green the set seems, for a difficult quantitative one, such as the actual total carbon output. Dual-process theories (Evans & Stanovich, 2013) suggest that such substitution reflects rapid, intuitive Type 1 reasoning that may resist reflective correction. The recycling NFI exemplifies this form of substitution and illustrates how perceptual and conceptual averaging mechanisms influence higher-order environmental judgments. Formalising this process within Bayesian cue-integration frameworks (Körding et al., 2007), in which cues are weighted by subjective reliability, may help connect cognitive theories of judgment aggregation with environmental-psychology accounts of moral spillover and compensatory behaviour.

The category manipulation in Experiment 2 provides further evidence that averaging operates at an abstract cognitive level. The illusion appeared when recyclable and non-recyclable items belonged to different object classes, such as bottles and cans, although these share a superordinate category (containers). A related, yet not peer reviewed, undergraduate thesis found further evidence of a cross-category NFI in estimates of bundles of clearly distinct items (buildings and electric bicycles; Ternerot & Piccardo, 2023). When considered alongside Experiment 1, where the NFI generalised across completely distinct domains (e.g., houses vs. bottles), a clear conclusion emerges: the averaging mechanism is not tied to a particular object class. Instead, once averaging is triggered by task demands, the cognitive system integrates cues across conceptually unrelated categories. Category boundaries do not constrain averaging when it is engaged, although they do not independently trigger the illusion. This interpretation aligns with research on ensemble coding and summary-statistic representations (Ariely, 2001; Whitney & Yamanashi Leib, 2018), which shows that global averages are extracted across heterogeneous inputs that share a meaningful conceptual dimension, and with norm-based accounts (Leopold et al., 2001) in which judgments are encoded relative to a conceptual environmental norm.

A key contribution of Experiment 2 is clarifying the limited role of affect and moral reasoning in producing the NFI. Neither valence nor arousal predicted the magnitude of the illusion, even when stimulus variability was maximised. This null finding complements earlier evidence showing weak or inconsistent relationships between the NFI and compensatory green beliefs or moral-licensing tendencies (Threadgold et al., 2022). The combined pattern suggests that affective and motivational factors may shape general attitudes toward sustainability (Gorissen et al., 2024) but do not underlie the cognitive mechanism responsible for the illusion itself. Comparable dissociations between affect and quantitative estimation appear in scope-insensitivity effects (Fetherstonhaugh et al., 1997) and in averaging biases in moral evaluation (Baron & Ritov, 2004; Hsee, 1998), supporting the view that the tendency to average rather than sum reflects a domain-general cognitive bias.

At the same time, it is important to recognise that the present design was not optimised to detect all possible affective mechanisms. The affective measures employed here assessed general evaluative tone (valence and arousal) rather than specific moral or self-relevant emotions such as guilt, regret, or perceived compensatory virtue, which have been emphasised in prior research on moral licensing and environmental decision-making. Moreover, the estimation task involved abstract judgments of total environmental impact rather than concrete consumption or choice scenarios, contexts in which affective responses may be weaker or less behaviourally consequential. Consequently, the present findings should be interpreted as demonstrating a limited role for general affective evaluation in predicting susceptibility to the NFI under estimation conditions, rather than ruling out the possibility that more specific moral emotions may influence environmental judgments in other contexts.

Importantly, the same dissociation extended to behavioural intentions. Although recyclable items were rated as substantially more likely to be purchased than non-recyclable items, individual differences in buying-likelihood did not predict susceptibility to the NFI. This finding indicates that preferences and intentions toward environmentally friendly products are not incorporated into the cognitive computation that underlies impact estimation. Behavioural intentions therefore appear to reflect downstream evaluative responses rather than upstream determinants of how environmental information is integrated. In this respect, the NFI arises prior to choice, at the level of judgment formation, rather than as a motivated justification of pro-environmental behaviour.

Although affect did not predict the magnitude of the illusion, this does not imply that affect is irrelevant to environmental judgment. Rather, affective evaluations may arise as secondary appraisals layered onto a more fundamental cognitive process (see, e.g., Asutay et al., 2024). Participants may first form an integrated estimate of a set's environmental friendliness through attribute substitution, with affective tone attaching only afterwards. This would explain why recyclable items may elicit positive feelings without modulating the NFI itself. In this sense, affect colours judgments without influencing the computational mechanism that produces the illusion.

At an applied level, the results point to a systematic cognitive obstacle in everyday environmental reasoning. Mixed waste streams are common, and the present findings show that people reliably underestimate their combined impact. This bias may contribute to public overconfidence in recycling as a mitigation strategy (Barnett et al., 2023) and to organisational tendencies to highlight minor eco-initiatives alongside conventional practices. Interventions that counteract the NFI should therefore target the integration rule itself, for example, by prompting summation, presenting cumulative metrics such as total kilograms of CO_2 or waste volume, or visualising additive outcomes. Such approaches may help consumers and policymakers achieve more accurate environmental assessments.

Several limitations warrant consideration. Although category structure and affective tone were manipulated, all materials involved recycling contexts. Future research should test whether similar averaging mechanisms govern judgments in more abstract or large-scale settings, such as household energy use, carbon-offset schemes, or mixed sustainability portfolios. The use of explicit rating scales may also obscure dynamic influences such as attention or motivation that would emerge in more ecologically valid environments. Finally, although affect did not modulate the illusion here, subtle affective processes may still interact with moral framing under real-world conditions; combining behavioural tasks with physiological or neurocognitive measures may help clarify such interactions.

In conclusion, the present work shows that the NFI extends to recycling judgments and that its emergence depends on conditions that promote heuristic averaging. Once engaged, the mechanism generalises across categories and affective contexts and reflects a high-level cognitive averaging process rather than an affective or moral bias. By situating the illusion within broader theories of heuristic information integration, the findings reveal how the tendency to average mixed cues can lead to systematic underestimation of environmental harm. Understanding this mechanism provides both a theoretical advance in cognitive-judgment research and a practical foundation for improving sustainability communication.

References

Anderson, N. H. (1981). Foundations of Information Integration Theory. New York: Academic Press.

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.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*(2), 157–162.

Asutay, E., Karlsson, H., & Västfjäll, D. (2024). Affect and impact neglect in sustainable decision-making. *Emotion Review*, *16*(4), 276-278. https://doi.org/10.1177/17540739241278933

Barnett, M. J., Hancock, P. I., Klotz, L. E., & Attari, S. Z. (2023). Recycling bias and reduction neglect. *Nature Sustainability*, *6*(11), 1418-1425.

Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes, 94*(2), 74–85.

Birnbaum, M. H. (1973). Morality judgment: Test of an averaging model with differential weights. *Journal of Experimental Psychology, 99*(3), 395-399.

Birnbaum, M. H. (1974). The nonadditivity of personality impressions. *Journal of Experimental Psychology, 102*(3), 543-561.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49-59.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*(3), 223–241.

Faul, F., Erdfelder, E., Lang, A., & 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.

Fetherstonhaugh, D., Slovic, P., Johnson, S., & Friedrich, J. (1997). Insensitivity to the value of human life: A study of psychophysical numbing. *Journal of Risk and Uncertainty, 14*(3), 283–300.

Gorissen, K., & Weijters, B. (2016). The negative footprint illusion: Perceptual bias in sustainable food consumption. *Journal of Environmental Psychology, 45*, 50–65.

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.

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision, 9*(11), 1.1-113.

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.

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.

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*, 823.

Hsee, C. K. (1998). Less is better: When low-value options are valued more highly than high-value options. *Journal of Behavioral Decision Making, 11*(2), 107–121.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. 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.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS One, 2*(9), e943.

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience, 4*(1), 89–94.

MacCutcheon, D., Holmgren, M., & Haga, A. (2020). Assuming the best: Individual differences in compensatory “green” beliefs predict susceptibility to the negative footprint illusion. *Sustainability, 12*(8), 3414.

Nadricka, K., Millet, K., & Aydinli, A. (2024). Are consumers more or less averse to wasting organic food? *Journal of Environmental Psychology, 93*, 102222.

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review, 72*(6), 407.

Pelham, B. W., Sumarta, T. T., & Myaskovsky, L. (1994). The easy path from many to much: The numerosity heuristic. *Cognitive Psychology, 26*(2), 103–133.

Sörqvist, P., Colding, J., & Marsh, J. E. (2020). Psychological obstacles to the efficacy of environmental footprint tools. *Environmental Research Letters, 15*(9), 091001.

Sörqvist, P., Haga, A., Langeborg, L., Holmgren, M., Wallinder, M., Nöstl, A., Seager, P. B., & Marsh, J. E. (2015). The green halo: Mechanisms and limits of the eco-label effect. *Food Quality and Preference, 43,* 1–9.

Sörqvist, P., & Holmgren, M. (2022). The negative footprint illusion in environmental impact estimates: Methodological considerations. *Frontiers in Psychology, 13*, 990056.

Sörqvist, P., & Marsh, J. E. (2024). Conceptual and methodological considerations to the negative footprint illusion: A reply to Gorissen et al. (2024). *Journal of Cognitive Psychology, 36*(8), 954–963.

Ternerot, S., & Piccardo, J. (2023). The negative footprint illusion: När bedömningar av klimatpåverkan missleds av det miljövänliga. [The negative footprint illusion: When judgment of climate impact is misled by the environmentally friendly]. Dissertation retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:ltu:diva-98530

Threadgold, E., Marsh, J. E., Holmgren, M., Andersson, H., Nelson, M., & Ball, L. J. (2022). Biased estimates of environmental impact in the negative footprint illusion: The nature of individual variation. *Frontiers in Psychology, 12*, 648328.

Todd, P. M., & Gigerenzer, G. (1999). What we have learned (so far). In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 357–365). Oxford University Press.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

van Doorn, J., & Kurz, T. (2021). The warm glow of recycling can make us more wasteful. *Journal of Environmental Psychology, 77*, 101672.

Westfall, J. (2015). PANGEA: Power analysis for general ANOVA designs. *Unpublished Manuscript. Available at Http://Jakewestfall.Org/Publications/Pangea.Pdf, 4*

Whitmarsh, L., & O'Neill, S. (2010). Green identity, green living? the role of pro-environmental self-identity in determining consistency across diverse pro-environmental behaviours. *Journal of Environmental Psychology, 30*(3), 305–314.

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology, 69*(1), 105–129.

Figure 1

Non-Addition Condition (Houses):
Illustration of the stimulus used in the non-addition condition for houses. Participants viewed an image of 15 conventional houses, representing standard housing with no environmental certification, and estimated their environmental impact on a 1-100 scale.

15 conventional houses

Figure 2

Addition Condition (Houses):

Illustration of the stimulus used in the addition condition for houses. Participants viewed an image of 15 conventional houses plus 15 environmentally certified houses, where the latter were designed with sustainable materials and energy-efficient technology. Participants estimated the overall environmental impact.

15 conventional houses + 15 environmentally certified houses

Figure 3

Non-Addition Condition (Plastic Bottles):
Illustration of the stimulus used in the non-addition condition for plastic bottles. Participants viewed an image of 15 yellow bottles, representing single-use, non-recyclable plastic bottles, and were asked to estimate their environmental impact on a 1-100 scale.

15 single use non-recyclable plastic bottles

Figure 4

Addition Condition (Plastic Bottles):
Illustration of the stimulus used in the addition condition for plastic bottles. Participants viewed an image of 15 yellow bottles (single-use, non-recyclable plastic bottles) plus 15 green bottles (recyclable plastic bottles) and estimated the overall environmental impact.

15 single use non-recyclable plastic bottles + 15 environmentally certified plastic bottles

Figure 5

Illustration of the stimulus used in the cross-category cans 10 group. Participants viewed 10 non-recyclable cans, followed by 10 non-recyclable cans plus 20 recyclable plastic bottles and estimated their environmental impact on a 1-100 scale.

First estimation

10 Cross-category cans

10 non-recyclable cans

Second estimation

10 Cross-category cans

10 non-recyclable cans + **20 recyclable plastic bottles**

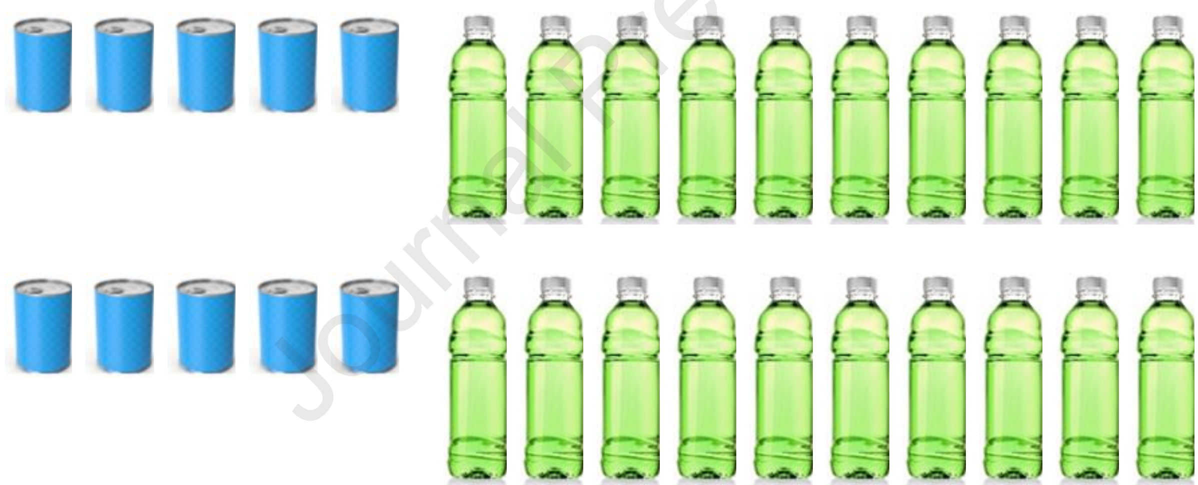
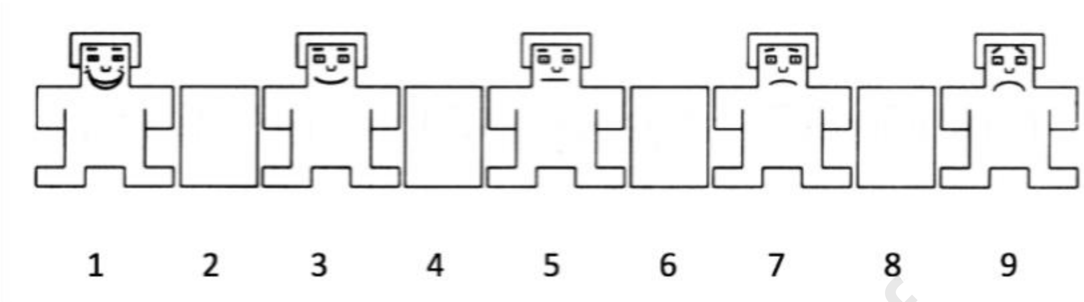


Figure 6

Valence.



Arousal.

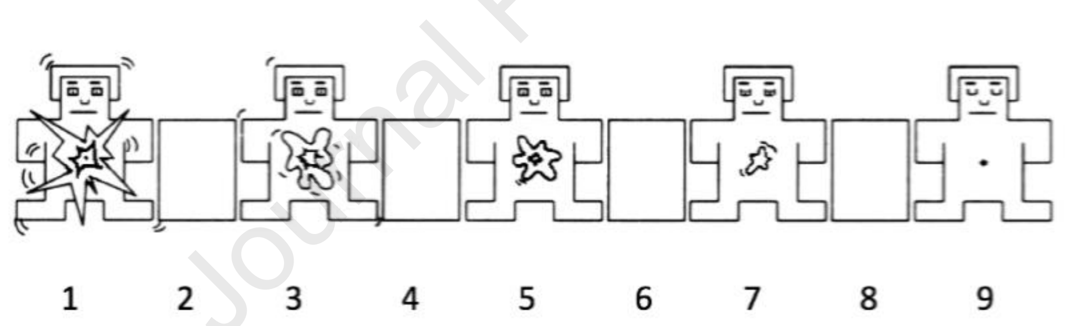


Table 1

The group name, control addition (first estimate) and addition condition (second estimate) for each presentation

Group	Control (1st estimate)	Addition (2nd estimate)
Cans 10	10 non-recyclable cans	10 non-recyclable cans + 20 recyclable cans
Cans 20	20 non-recyclable cans	20 non-recyclable cans + 20 recyclable cans
Cans 40	40 non-recyclable cans	40 non-recyclable cans + 20 recyclable cans
Bottles 10	10 non-recyclable plastic bottles	10 non-recyclable plastic bottles + 20 recyclable plastic bottles
Bottles 20	20 non-recyclable plastic bottles	20 non-recyclable plastic bottles + 20 recyclable plastic bottles
Bottles 40	40 non-recyclable plastic bottles	40 non-recyclable plastic bottles + 20 recyclable plastic bottles
Cross-category cans 10	10 non-recyclable cans	10 non-recyclable cans + 20 recyclable bottles
Cross-category cans 20	20 non-recyclable cans	20 non-recyclable cans + 20 recyclable bottles
Cross-category cans 40	40 non-recyclable cans	40 non-recyclable cans + 20 recyclable bottles
Cross-category bottles 10	10 non-recyclable bottles	10 non-recyclable bottles + 20 recyclable cans
Cross-category bottles 20	20 non-recyclable bottles	20 non-recyclable bottles + 20 recyclable cans
Cross-category bottles 40	40 non-recyclable bottles	40 non-recyclable bottles + 20 recyclable cans

Table 2

The means and standard deviations for Quantities 10, 20, 40, for plastic bottles and cans in the same category and cross category conditions for the non-recyclable (first estimate) and addition recyclable condition (second estimate) for each presentation in Experiment 2.

IVs	Estimations and quantity											
	Non-recyclable (1st estimate)						Recyclable (2nd estimate)					
	10		20		40		10+20		20+20		40+20	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Same Category Cans	31.23	18.93	42.66	16.81	68.75	20.54	31.50	16.67	45.54	17.63	64.13	21.56
Cross-category Cans	29.83	17.79	44.71	18.10	65.72	21.67	32.47	17.61	44.04	16.95	62.26	22.25
Same Category Bottles	40.76	19.02	58.77	18.15	80.67	19.52	43.06	18.45	56.13	19.45	74.96	20.85
Cross-category Bottles	39.66	18.15	57.04	16.50	79.33	17.21	40.90	17.21	56.15	18.21	73.98	18.48

Table 3

The means and standard deviations for Item type, Category Type, Estimate Type and Quantity Type for groups 10, 20, 40 in Experiment 2.

	Independent variables								
	Item		Estimate		Category		Quantity		
	can	bottle	1st	2nd	Same	Cross	10	20	40
Mean	46.90	58.45	53.26	52.09	53.18	52.17	36.18	50.63	71.23
SE	1.34	1.20	1.03	1.29	1.10	1.10	1.24	1.11	1.35

Table 4

Results from a repeated measures ANOVA in Experiment 2.

Factor	*F*	*MSE*	*p*	η_p^2
Item type	57.49	1391.74	<.001**	.367
Set type	1.08	760.66	.302	.011
Category type	1.44	421.63	.233	.014
Quantity of base set	421.25	589.19	<.001**	.810
Item type × Set type	2.94	92.74	.090	.029
Item type × Category type	0.07	363.26	.789	.001
Item type × Quantity	2.90	161.48	.058	.028
Set type × Category type	0.03	174.02	.874	< .001
Set type × Quantity	25.76	83.51	<.001**	.207
Category type × Quantity	0.71	163.22	.493	.007
Item type × Set type × Category type	0.61	78.44	.806	.001
Item type × Set type × Quantity	2.18	58.43	.115	.022
Item type × Category type × Quantity	0.86	129.19	.427	.009
Set type × Category type × Quantity	0.73	59.22	.483	.007
Item type × Set type × Category type × Quantity	3.34	75.07	.037*	.033

Note: * $p < .05$, ** $p < .001$

Highlights

- Negative footprint illusion extends to everyday recycling judgments
- Mixed recyclable sets judged less harmful than non-recyclable alone
- Effect emerges when set size makes summation difficult
- Averaging occurs across categories (e.g., bottles and cans)
- Affective responses and buying intentions do not explain the effect

Author Contributions (CRediT)

Sarah Ormesher: Conceptualization; Methodology; Investigation; Data curation; Formal analysis; Visualization; Writing – original draft.

Kathleen C. McCulloch: Conceptualization; Methodology; Supervision; Project administration; Writing – review & editing.

Beth H. Richardson: Methodology; Writing – review & editing.

Patrik Sörqvist: Conceptualization; Methodology; Writing – review & editing.

John E. Marsh: Conceptualization; Methodology; Formal analysis; Supervision; Project administration; Writing – review & editing.