
Article:
Purvis, B., Calzolari, T. and Genovese, A. orcid.org/0000-0002-5652-4634 (2025) Consensus and contestation: reflections on the development of an indicator framework for a just transition to a circular economy. Ecological Economics, 230. 108476. ISSN 0921-8009

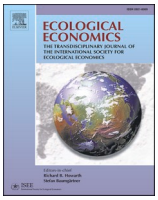
https://doi.org/10.1016/j.ecolecon.2024.108476

Ecological Economics 230 (2025) 108476

Contents lists available at ScienceDirect

Ecological Economics

journal homepage: www.elsevier.com/locate/ecolecon



Consensus and contestation: Reflections on the development of an indicator framework for a just transition to a circular economy

Ben Purvis *, Tommaso Calzolari, Andrea Genovese

Sheffield University Management School, University of Sheffield, Sheffield, UK

ARTICLE INFO

Keywords:
Indices
Philosophy of measurement
Pluralism
Ecological justice
Quantification
Circular economy

ABSTRACT

We explore an attempt to derive a set of indicators reflecting a just transition to a circular economy (CE) at a supply chain level. Here we build upon the theoretical work presented in Purvis and Genovese (2023) with an account of an empirical exercise following the standard methodological steps outlined for the creation of a measurement dashboard. A literature review of existing CE indicators for supply chains was therefore followed by a Delphi approach which sought to understand and incorporate the expertise of CE scholars and practitioners. The 3 round Delphi incorporated a survey, and an individual, and group Analytical Hierarchy Process, as a standard technique to derive consensus from experts in terms of suitable indicator categories. Yet contestations observed during the consensus building exercises cast doubt on the suitability of our nominally consensus-driven approach, as well as the use of indicators themselves for our critical purposes. We describe the dilemmas precipitated by this failure of consensus, with reference to the inherent challenges to indicator frameworks and a series of questions for better research design. The paper also reflects on the fundamental contradictions related to the use of indicators for inducing transformational dynamics, and problematises the desire for consensus, thereby paving the way for further research avenues.

1. Introduction

The use of indicator-based methods, such as the dashboard of 231 Sustainable Development Goal (SDG) indicators for measuring progress on the UN's 2030 Agenda for Sustainable Development (United Nations, 2017), are widespread as tools for purportedly assessing and enabling progression towards a sustainable future. Yet indicators remain contested, and their limitations are well documented, including technical issues such as data availability (Kinyondo and Pelizzo, 2018; Mitchell, 1996), and epistemological challenges such as trade-offs and risks of reductionism (Böhringer and Jochem, 2007; Gasparatos et al., 2008). Metricisation can also be criticised for its perpetuation of a neoliberal institutionalising logic that reduces social and ecological justice to externalities on a balance sheet. Nevertheless, such assessment approaches are considered by many to be vital for communicating a broad set of complex information to policy-makers, facilitating better decision making (Sébastien and Bauler, 2013).

This present work represents an empirically-driven follow up to our previous theoretical discussion on the challenges of developing indicator frameworks, presented in Purvis and Genovese (2023). In this initial theoretical article, a typical methodological approach to framework development was presented as three iterative steps: selection, theoretical framing, and implementation. Purvis and Genovese (2023) illustrate each of these steps with a series of technical and epistemological challenges presented within the critical literature on indicator assessment, reproduced in Table 1 below. The work concludes by problematising the ability of indicator based approaches for catalysing transformative change, framing this in terms of whether we should be making better and more informed use of the same methods, or consider different approaches entirely. A series of questions are subsequently outlined to support better research design for indicator based studies, and a number of different approaches are suggested.

This present article significantly builds on this by contextualising the previous theoretical exploration of indicator approaches with an empirical example of our own attempt to develop a 'better' indicator set. By situating the prior theoretical reflection, we are able to evidence, enrich, and further develop our reflections on new ways forward in terms of critically informed assessment and alternative methods. As part of a European Commission funded project, we were tasked to develop a 'dashboard of indicators' relating to a reconceptualisation of the

* Corresponding author.
E-mail address: b.purvis@sheffield.ac.uk (B. Purvis).

https://doi.org/10.1016/j.ecolecon.2024.108476
Received 8 February 2024; Received in revised form 28 October 2024; Accepted 26 November 2024
Available online 4 December 2024

Table 1
Challenges of indicator based approaches identified within the academic literature. Reproduced from Purvis and Genovese (2023).

Step	Technical Issues	Episto-ontological Challenges
1) Selection	● Determining problem scope ● Getting the right coverage ● Different understanding of terms ● Quality criteria ● Which stakeholders? How can they input?	● Reductionism: can complex socio-environmental factors be reduced to a series of indicators? ● Measuring the immeasurable: marginalisation of social / qualitative aspects
2) Theoretical framework	● Absence of theory ● Handling complexity ● Difficulty of comparing disparate factors ● Weighting & normalisation	● Trade-offs as inherent? ● The whole is bigger than the sum of its parts ● Depoliticisation ● Arbitrariness
3) Implementation & outcomes	● Data constraints ● Limited resources ● Lack of political will ● Failure of consensus	● Lack of pluralism ● Difficulty of observing and measuring outcomes

Circular Economy (CE) paradigm through the lens of an inclusive and socially just transition (Pansera et al., 2021), specifically focusing on supply chains and production networks as our unit of analysis. Our initial research design for this task intended to follow the methodological steps outlined for the creation of an indicator dashboard within Purvis and Genovese (2023), following what is arguably the dominant approach within the literature. In the selection of indicators, we began with a systematic literature review of CE indicators for supply chains (published as Calzolari et al., 2022). This was followed by a three-round Delphi study consisting of a survey, and a hybrid co-production workshop, which were intended to incorporate the views of CE scholars and practitioners on suitable indicators and framing.

Whilst critical of reductionist methods, we took an open and experimental approach aiming to build consensus across this group of experts, utilising existent critique to address potential challenges. By maintaining a critically reflective stance with respect to both the possibility and indeed desirability of a consensus driven approach, we were able to interrogate the inherent assumptions and difficulties involved. Despite this openness, the discussion that took place during the third Delphi round, echoed many of the objections and challenges to an indicator approach observed within the literature, and cast serious doubt on its suitability for our purposes, as well as the nature of our consensus building exercise. Ultimately our indicator-based approach was largely abandoned, deemed unsuitable for the critical perspective desired by the wider project's goals.

This paper is thus both a presentation of an empirical approach to develop an indicator dashboard within a critically oriented project, and a reflection on the challenges experienced. Similar to Berry et al. (2022), we consequently dispute that consensus is necessary and desirable for operationalising the CE concept. We emphasise the inherently political nature of the CE, and reject the possibility of an objective methodological approach. At a theoretical level, this paper and its discussion furthers the debate on measuring the transition towards a CE, unfolding in greater contextualised detail epistemological and ontological challenges which have often been overlooked in the extant literature. It holds wider relevance for the use of indicator methods more widely within the sustainability paradigm, and approaches which take a critical orientation to transformative social change.

The remainder of this paper is organised as follows: Section 2 provides background on the current debate on the development of indicator systems for evaluating the transition to a CE. Section 3 describes the methodological approach we employed for the development of the indicator dashboard. Section 4 illustrates the results of this exercise, which are discussed in Section 5 with reference to the challenges outlined within Purvis and Genovese (2023). Section 6 presents our reflections on the nature of consensus, indicator frameworks, and some thoughts on why our approach failed to have the desired outcomes, before some final conclusions are given in Section 7.

2. Background

Over the past decade national and supranational bodies have revived the concept of the Circular Economy (Pearce and Turner, 1990) as a key principle of their industrial and environmental policies (Bleischwitz et al., 2022; European Commission, 2020). In its mainstream conceptualisation, a CE is intended to stimulate the development of new production and consumption systems where materials and products are reused, remanufactured and recycled. Within this techno-optimistic view, enhancing markets of secondary products and materials, and promoting servitisation, is expected to lead to positive environmental, social, and economic outcomes (Lowe and Genovese, 2022). Yet, a growing body of literature has criticised mainstream ecomodernist assumptions (Corvellec et al., 2022), and alternative framings have been proposed from various sources (Genovese and Pansera, 2021; Purvis et al., 2023b; Savini, 2023; Schultz and Pies, 2023, Pansera et al., 2024). Far from having a univocal and agreed definition, the CE represents an umbrella concept, under which competing paradigms exist (Homrich et al., 2018; Korhonen et al., 2018; Oliveira et al., 2021).

In this context, an increasing amount of scholarly work has developed assessment approaches, with the goal of either designing more circular production and consumption systems, or tracking progress towards them. A drive towards 'holism' and the need to capture multiple diverse phenomena within a single assessment framework, means the 'indicator dashboard' pervades the assessment regime as a curated selection of singular indicators (Purvis and Genovese, 2023). Thus, within the CE paradigm, various 'circularity metrics' (Corona et al., 2019) find themselves complemented with indicators purposed towards incorporating additional aspects relating to risk and impact (Ellen MacArthur Foundation and ANSYS Granta, 2019), as well as social, environmental, and economic dimensions (Roos Lindgreen et al., 2020; Vinante et al., 2021; Walzberg et al., 2021).

Whilst the necessity of assessment of or towards a CE is largely accepted uncritically, a number of authors have pointed out the lack of commonly accepted indicators or methods (Harris et al., 2021; Moraga et al., 2019; Vinante et al., 2021), described by Roos Lindgreen et al. (2020) as a "barrier to transitioning to a CE" (p2). It is in this context that CE assessment approaches have been reviewed in terms of their coverage of social, environmental, and economic dimensions, with a recurrent finding being a lack of attention to social dimensions (Calzolari et al., 2022; Roos Lindgreen et al., 2020; Vinante et al., 2021; Walzberg et al., 2021). A lack of attention to social dimensions within CE assessment paradigms is a recurrent observation that has been critiqued across other fields (see e.g. Boyer et al., 2016; Desiderio et al., 2022; Opp, 2017). In particular, CE literature and practice seems to have not benefited much from the clearer focus on the social dimension from related concepts at various scales, such as the Doughnut Economics approach (Raworth, 2017), which relies heavily on the idea that sustainability is about respecting planetary boundaries while granting social minima, or the UN SDGs, which consider a variety of dimensions which integrate both social and environmental factors. Despite narratives relating to corporate social responsibility (CSR), and environmental, social, and governance (ESG) at micro- and meso- scales, it is unclear to what extent these are enabling positive social and environmental change in the CE domain (Todaro and Torelli, 2024). The lack of engagement with the social dimension of CE measurement approaches belies two core considerations. First, what can be measured and how, is the subject of contestation, with the selection of indicators based on readily available data marginalising dimensions which are not easily

quantifiable. Second, the selection of indicators is inherently political and reflects underlying normative value systems: selection communicates what is (and by omission, what is not) important.

Purvis and Genovese (2023) ask "to what extent indicator frameworks are fundamentally limited in terms of transformative potential" (p4), reflecting on the social construction of indicators, and arguing them to be as much political artefacts that create and shape new realities, as technical tools for measurement. Their discussion offers up a number of ways in which a better methodological approach to indicators can be actioned and embedded in the research design. Here, *better* means exploring a series of questions relating to the purpose of the indicator framework, the theory it is embedded in, and a clear approach to indicator selection. They end by questioning the transformative potential of indicator frameworks, referencing the lack of empirical evidence of 'successful' application, and suggesting that better indicator frameworks should be complemented by an underpinning methodological pluralism. In this present work, through the Delphi approach outlined in the next section, such theoretical perspectives are tested against an empirical exercise which adopts the methodological steps outlined for the creation of *better* indicator dashboards.

3. Selection of indicator categories using a literature review and Delphi study

As described in Section 1, our empirical approach was shaped by the demands of a wider project for a dashboard of indicators relating to a just transition to a CE, using the supply chain as its unit of analysis. We opted to use the sequential steps to dashboard development outlined in Purvis and Genovese (2023): "1) The selection process: which indicators to select as relevant, who should decide and how; 2) The assessment framework: including theoretical framing, and how the indicators are combined, aggregated, or contrasted; and 3) Implementation: how assessment is carried out in practice" (p4). For the selection phase, we followed a common approach (El Gibari et al., 2019; Turcu, 2013) whereby a preliminary literature review of relevant factors was followed by external input from stakeholders or 'experts' (in our case, through a Delphi method). It was intended that this would result in a refined selection of core indicator categories[1] that could be populated with measurable indicators as appropriate.

Delphi methods are an established set of processes used to survey and collect expert opinion on a particular subject, allowing for the structured ranking and prioritisation of a set of items (Yousuf, 2007). Several adaptations of the Delphi method have already been demonstrated for surveying CE-related issues (see e.g. Mahanty et al., 2021; Prieto-Sandoval et al., 2018). Our objective here was to gather and incorporate 'expert' opinion in order to fine-tune the most common indicator categories identified by a systematic literature review. Three Delphi rounds followed the preliminary literature review: a survey, an individual Analytical Hierarchy Process (AHP) round, and a group AHP round with discussion. Each of these rounds were supplemented by a number of opportunities for group discussion amongst the expert panel, allowing for a reflexive process. The Delphi was intended to identify a list of indicator categories, based on the initial literature derived set, and provide a ranking of their relative importance as judged by expert stakeholders, which could also inform any potential weighting and aggregation of indicators within the framing step. Fig. 1 outlines our methods of indicator category selection and their interrelationships.

The remainder of this section describes each of these rounds in detail. For our expert panel, we engaged members of the project's extended consortium network, consulting 35 individuals in round 1, and 25 in subsequent rounds. All members of the panel are actively engaged in CE research and practice (both academic and non-academic), and are both aware and largely sympathetic to the critical perspective to CE outlined in Section 2. Table A.1 gives some details of the panel composition, including background, seniority, location, and whether English (the sole language which the Delphi took place in) is a first language. Whilst we made an effort to engage participants from multiple geographical areas to achieve variation in terms of respondents' background and cultural environment, as well as ensuring a wide interdisciplinary coverage, we acknowledge the inevitability of panel bias (Keeney et al., 2001). Particular areas to highlight include drawing solely on individuals actively engaged with the wider project's network, a skew towards academic participants, a skew towards the global north, and potential language barriers despite all participants having a good grasp of the English language. We therefore followed Steinert's (2009) principle of caution, framing Delphi as an explorative research tool rather than a quantitative test or predictor.

3.1. Preliminary literature review

The initial phase of identifying a dashboard of indicators took the form of a systematic literature review, published as Calzolari et al. (2022). This review focused on CE indicators for supply chains, analysing tools and indicator sets within the academic literature, alongside content analysis of industrial practice reports. After being systematically collated and analysed, a corpus of indicator categories was assessed on its coverage of economic, environmental, and social dimensions, and the most commonly used categories within each of these dimensions were compiled. These more generic dimensions, rather than explicit conceptualisation of a 'just transition' at this stage, reflect a lack of explicit acknowledgement of this framing within the literature. Coverage of these dimensions was therefore considered to be a preliminary organising framework for conceptualisation of CE from the theoretical perspective of the wider project.

As noted elsewhere in the literature (e.g. Souza Piao et al., 2023), the CE indicator studies examined were observed to focus primarily on economic and environmental dimensions with relatively little attention given to social dimensions. Only 18 % of studies included any social dimension, compared to 80 % and 66 % covering environmental and economic dimensions respectively. Even where social dimensions were present, they were often simplistic, focusing on quantitative rather than qualitative aspects, lying relatively far from concepts of social justice and distributional conflicts, with the most popular social metrics relating to very simple estimates concerning issues such as 'jobs created', and 'health & safety compliance'. We nevertheless sought to balance coverage across these three dimensions in developing a preliminary list. The full compiled list of the most commonly employed metrics across each dimension is reproduced in Table B.1, where the frequency of occurrence across the 203 analysed papers is also recorded, and each category is illustrated with example indicators. The reader may also refer to Table 2 in the main text where the categories following round 1 of the Delphi are described. The preliminary selection thus resulted in 19 indicator categories, organised across 3 dimensions: 6 for both the economic and social dimension, and 7 within the environmental dimension.

3.2. Delphi round 1: survey

An online survey was circulated to the expert panel, eliciting 35 usable responses. The survey questions were designed to build on and refine the preliminary list of indicator categories derived from the literature. This was primarily intended to incorporate the views of a set of critical experts into this set of categories, in particularl encouraging them to reflect on the aspects relating to justice we were unable to sufficiently account for in the literature review step. Following an initial

[1] Note, following Calzolari et al. (2022) we use 'indicator categories' to refer to a thematic grouping of similar operationalisable indicators or metrics, for example the category "virgin resource usage" covers a number of discrete metrics including abiotic depletion of resource; and mineral, fossil, & renewable resource depletion.

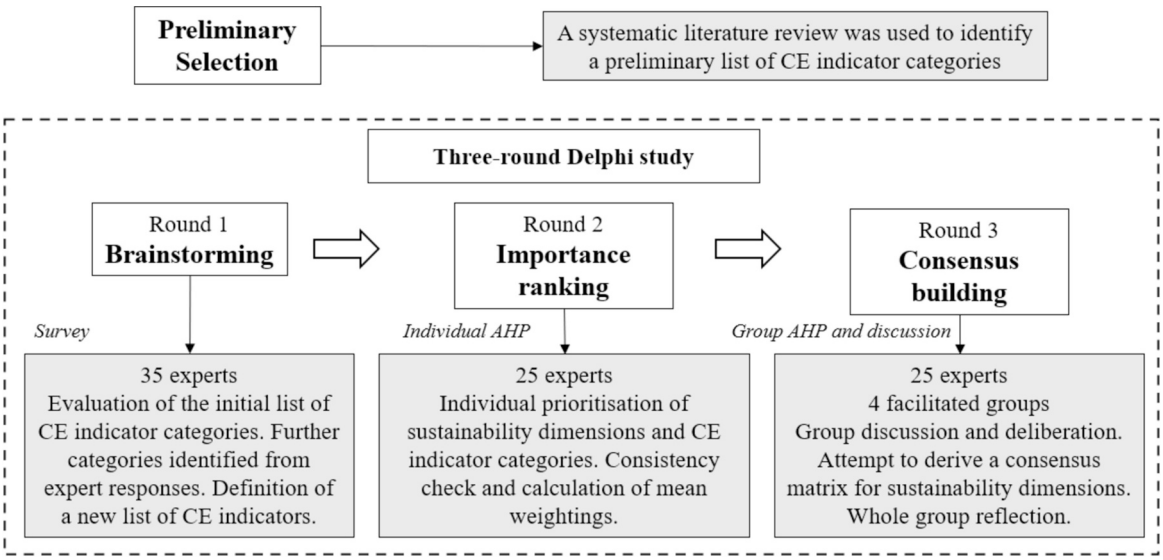


Fig. 1. Outline of indicator category selection methods.

Table 2
The revised list of indicator categories following round 1 of the Delphi study. The descriptions listed were provided to the expert panel for rounds 2 and 3.

Dimension	Category	Description
Economic	Cost of production	Cost of production at company and supply chain level
	Supply chain risks	e.g raw materials availability, reliance on critical materials
	Products quality	e.g. defect rates of end products
	Equitable investments[1]	e.g. ethical investment practices
	Surplus distribution[1]	e.g. worker & community share in profit
Environmental	Waste produced	Waste produced across the supply chain
	GHG emissions	Emissions produced by supply chain activities
	Energy usage	Energy usage across the supply chain
	Air pollution	Air pollution produced by supply chain activities
	Water consumption	Water used and contaminated across the supply chain
	Virgin resource usage	Virgin resource usage across the supply chain
Social	Jobs created	Jobs created across the supply chain
	Customer environmental awareness	e.g. information on disposal methods, incentives for recovery
	Social cost of waste[2]	e.g. community impact, personal health impacts
	Participatory planning[1,3]	e.g. worker participation in production management, involvement of key stakeholders and communities, participation in decision making
	Gender & equalities	e.g. gender balance, opportunities for marginalised groups
	Quality of work	e.g. percentage of employees on open-ended contracts, unionisation rates

[1] Derived from expert suggestions.
[2] Combining Calzolari et al.'s (2022) category of 'social cost of waste' with the coded theme 'social/community benefit' from expert suggestions.
[3] Combining the coded themes of 'worker participation' and 'stakeholder/customer/community participation' from expert suggestions.

prompt which asked the respondent to articulate their understanding of a 'just transition to a circular economy', participants were asked to select from the initial 19 indicator categories those which they judged *'relevant for assessing the contribution of an organisation and its supply chain to a just circular economy'*. This was done across three questions where each sustainability dimension, with its constituent categories, was presented separately. Each category was presented with a short description. In addition, for each dimension, a written response box was provided to solicit further categories that respondents would add to each of the three dimensions. Our aim through the prompts in this round was to be able to remove categories that were unpopular with respondents, as well as allow for the addition of new categories suggested by the experts.

The free text box for suggesting additional indicator categories collected a significant number of responses. These were of varying quality in terms of their clarity, and thus we decided to use a thematic analysis (Braun and Clarke, 2006) in order to analyse and aggregate these responses with reference to the existing categories that were created following the literature review. This was done as a discursive exercise between the research team. The pre-existing indicator categories were used as an initial set of themes to which some of the coded responses were judged to relate (e.g. the written response of 'employee satisfaction' was judged to relate to the initial indicator category of 'quality of work'). Other coded responses prompted the creation of new categories (for example worker participation, democracy, and agency recurred across the responses, leading to the new category of 'participatory planning'). Once we were happy with a set of themes for each dimension we decided to eliminate some of the least popular categories so that we had a manageable number to compare for the subsequent AHP phase.

We followed the principle to preserve the most frequently observed literature-based categories, and include the most frequently proposed categories within the survey responses, with the aim to end up with 6 categories within each dimension, a suitable number for an AHP exercise (Ozdemir, 2005). We preferenced recurrent expert suggested themes, and used consensus thresholds for each of the initial categories ($\gtrsim$ 70 % of participants saying a category is relevant) to derive a new list of categories under each sustainability dimension. Only categories that were judged independent, and had been suggested by multiple participants, were added to the new list. The revised list of indicator categories following this round is displayed in Table 2. Section 4.1 provides further detail on the survey results, and how this new set of categories was constructed.

3.3. Intermission: feedback between Delphi rounds

According to Boulkedid et al. (2011), feedback between rounds represents an essential aspect of the Delphi process. The analysis of the survey findings and refinement of indicator categories were thus

presented to the expert panel at the beginning of the hybrid workshop in which rounds 2 and 3 took place. This was intended to give participants an understanding of the breadth of views across the group, with areas where responses diverged and converged being emphasised. Space was given for discussion, and any questions or reflections that participants wished to raise.

3.4. Delphi round 2: individual AHP

The remainder of the workshop took the form of an AHP, split into an individual and then group round. AHP is a quantitative technique, commonly employed to structure and facilitate complex decision-making processes (Saaty, 2008), the reader can refer to Appendix C for a more detailed outline of AHP. The technique employs 'pairwise comparison matrices' which ask the participant to compare components and assign a numerical weighting to quantify their relative importance. The rationale of using AHP was not to assign quantitative weights to each category within the dashboard, as is sometimes how frameworks have been developed within the literature (Gan et al., 2017), but to better understand the level of consensus that experts had relating to the derived indicator categories.

After an explanation of the activity, participants were given three pairwise comparison matrices for each respective dimension, comparing the indicator categories derived in round 1. Additionally, a fourth matrix was given to participants which compared the three sustainability dimensions themselves. Participants were asked to individually fill in these matrices answering *'which of the two elements is more important in the context of the transition towards a just circular economy, and how strongly?'*, using a scale of 1 to 9. A number of experienced facilitators were on hand to aid participants with this activity.

3.5. Delphi round 3: group AHP and discussion

Workshop participants were split into groups (3 in person, and 1 online), each with a facilitator. The groups were asked to derive a collective group matrix based on their individual matrices comparing the three sustainability dimensions. This was done by deliberation and group discussion guided by the facilitator. Each facilitator also made notes on the group discussion that took place during the production of the collective group matrix, focusing specifically on *'areas of conflict, and areas of consensus'*. Due to time constraints, this round was only able to focus on the high-level matrices comparing each dimension, rather than the matrices which compared individual indicator categories. Instead, prioritisation was given to wider discussion and reflection amongst all the workshop participants.

4. Findings from the Delphi process

The core findings from the Delphi process as described above may be delineated into the survey findings from round 1, the consistency of the matrices obtained from round 2, the resultant weightings and analysis of these matrices, and the discussion that was facilitated amongst the experts in round 3.

4.1. Round 1: survey results

Table 2 presents the list of revised categories which resulted from this Delphi round. Of the social categories which were identified through the literature review, the selection of those deemed relevant by respondents was broad without strong agreement. None of the initial categories were unanimously selected by all participants, with the most frequently picked category, 'jobs created', selected by 78 % of respondents, followed by 'customer environmental awareness' 76 %, 'social cost of waste' 76 %, 'quality of work' 67 %, 'worker training' 64 %, 'H&S compliance' 61 %, and 'employee benefits' 61 %. Over half of participants provided write-in responses which were thematically coded as described in Section 3.2. Of these, the theme of 'social/community benefits' (5 responses) was added to the existent theme of 'social cost of waste', and 9 responses were judged to relate to the existing themes of 'quality of work' and 'worker training'. Two new distinct categories were identified for inclusion relating to 'gender & equalities' (6 responses, e.g. 'company gender balance'), and 'participatory planning' (14 responses, e.g. 'the existence of mechanisms for employees to participate in decision making'). The two least popular original categories were removed.

The selection of categories in the environmental dimension amongst participants displayed a significant amount of agreement relative to the social dimension. All categories were selected with a high frequency: 'waste produced' 100 %, 'GHG emissions' 97 %, 'energy usage' 94 %, 'air pollution' 91 %, 'water used or contaminated' 88 %, 'virgin resource usage' 82 %, 'acidification' 79 %. Fewer write-in responses were received, which were more diverse. Notably, 'circularity' was mentioned by only one respondent. Themes that were created from multiple coded responses encompassed 'global supply chain factors' (5 responses, e.g. 'the distance of travel for raw materials'), 'land use aspects' (3 responses, e.g. land footprint), and 'displacement of primary production' (2 responses, e.g. 'share of raw materials in production'). Ultimately due to low coverage, none of these write-in themes were included as categories in the new set. Acidification was removed as the least popular category to take the total down to six.

For the economic dimension the selection of relevant categories displayed some agreement, though it appeared several of the presented categories were unpopular. The selection rate was as follows: 'cost of production' 97 %, 'supply chain risks' 84 %, 'quality' 81 %, 'profits' 68 %, 'time responsiveness' 65 %, 'return on investment' 55 %. Recurrent themes within the write-in responses related to 'ethical and equitable investment' (5 responses, e.g. 'the existence of ethical investment policies'), and 'profit distribution' (5 responses, e.g. 'the share of profits distributed to workers'). The three least popular categories were removed and these two new categories were included.

4.2. Round 2: AHP consistency

Consistency adjustment is a routine analysis within the AHP paradigm, whereby the matrices are checked for areas of logical inconsistency in assigning numerical values (Karapetrovic and Rosenbloom, 1999). The standard method employs an acceptability threshold, associated with a Consistency Ratio (CR), which if the matrix meets is judged 'sufficiently consistent' (Saaty, 2008); details of this process are provided in Appendix C. Whilst we must be careful drawing strong conclusions from these quantitative checks, consistency may be interpreted as a proxy for how much difficulty participants had comparing elements. Thus a high consistency represents easier choices within comparison, and a lower consistency represents a more difficult choice. The presence of logical inconsistencies as a standard observation within the AHP process highlights some of the inherent challenges relating to comparing and ranking indicators, a clear manifestation of incommensurability.

The individual matrices from the second Delphi round were checked for consistency, with matrix consistency observed to vary by sustainability dimension. In particular, the social matrices were the least consistent: only 64 % of the social matrices met the CR acceptability threshold, compared to 68 % for the economic matrices and 88 % for the environmental matrices. Differences in judgement consistencies across dimensions was also checked with a Chi-Square test, whose results are reported in Table 3, highlighting the existence of some level of statistical significance in these differences, despite the small sample size.

These scores align with the results of the survey round, in which the social dimension observed the most disagreement over the importance of categories, whereas the environmental dimension saw strong agreement. It is notable that a number of participants filled in their environmental matrices in a manner indicating that they viewed all indicator categories as equally important. Again, we wish to stress caution in the

Table 3
Chi-Square test results comparing inconsistency levels across dimensions (* significant at $p < 0.10$ level; ** significant at $p < 0.05$ level).

Chi-Square (p-value)	Social	Environmental	Economic
Social		3.947 (0.047)**	0.089 (0.765)
Environmental			2.914 (0.088)*
Economic			

overinterpretation of statistical values here, and it is worth noting that the social matrices were presented to users as the first 6 × 6 pairwise comparison matrix after the 3 × 3 dimensions matrix (followed by the environmental and economic matrices respectively).

Having identified inconsistent matrices, we used an algorithmic method to reduce inconsistency to allow for further standard analysis. Whilst the preferred approach for this involves a deliberative process between the decision maker and a facilitator, this was not possible due to time constraints. Instead, we used an online calculator (Goepel, 2018) to determine consistency of each numerical entry, before manually adjusting the most inconsistent entry 1 point on the scale to improve consistency. This process was repeated until the overall consistency of each matrix met the consistency threshold. This algorithm is intended to maximise consistency whilst minimising the change to the matrices in terms of the judgments made, thus aiming to preserve the overall preference structure, rather than forcing respondents to alter their matrices in a hurry. This approach also helped us preserve, as much as possible, the judgement formulated by the participants, without forcing them to achieve an artificial consistency by significantly altering their preferences, which could have jeopardised the objective of our analysis.

4.3. Round 2: ranking importance of indicator categories

Once all individual matrices met the consistency threshold, we were able to calculate weightings (Render et al., 2017) of each indicator category, and dimension, and thus determine the relative importance of each category, as judged by the expert panel, after this round. By calculating weights for each individual matrix we were able to examine divergences and convergences across the panel (see Appendix C for more details on this approach). Table 4 displays the weights derived from each dimensional matrix averaged over all participants, we also display the standard deviation, and minimum and maximum individual values for comparison. Note that the standard deviation is included mostly as a rough indication of the numerical variation of responses, as we do not expect the distribution of weights to fit against a normal distribution. The weightings for each indicator category (prior to their further weighting with respect to dimensional results) are given in Table B.2.

Averaged over the panel, social and environmental dimensions appear to be judged roughly equivalent in importance, with the economic one clearly judged the least important, both in terms of its average value and the variation. Apart from two participants that judged all dimensions to be of equal importance, all other participants indicated the economic one to be the least important, often by a large amount (i.e. using the maximum value). Of the remaining participants, an even split was observed across those who judged the social dimension to be the most important dimension, those who selected the environmental one, and those who judged the two to be equally important. Thus whilst average scores indicate parity between social and environmental dimensions, this erases the spectrum of responses across participants, of whom nearly 60 % did not identify the two as equal. It is interesting to compare these findings to our preliminary literature review, in which the frequency of each dimension's occurrence across the literature was observed to be 49 % for the economic dimension, and 11 % for the social (Calzolari et al., 2022). This emphasises the need to be critical of interpreting the frequency of occurrence within the literature as an indication of importance.

Table 4
Weights of each dimension averaged over all participants.

	Mean	S.D.	Max	Min
Social	43.8 %	10.4 %	63.7 %	27.8 %
Environmental	45.4 %	11.5 %	66.3	25.8 %
Economic	10.8 %	8.6 %	33.3 %	5.1 %

The resultant weightings of the individual indicator categories (Table B.2) are more difficult to interpret. A high standard deviation makes it difficult to draw clear conclusions on preferences across the expert panel. Nevertheless, the categories added following round 1 were observed to perform well within their respective dimensions. Notably, the added categories of 'surplus distribution' and 'equitable investments' received the highest weightings within the economic dimension, again revealing the limitations of drawing directly from frequency of occurrence within the literature. A lower standard deviation, and more equal distribution of means across indicator categories within the environmental dimension aligned with the results from round 1, revealing more agreement amongst the panel in the choice of important environmental indicator categories. This again contrasted with higher divergence observed across the social dimension. Such an observation may be influenced by the permeation of environmental indicators, particularly GHG emissions, across popular discourse. The social dimension itself, on the other hand, is notably underconceptualised, and there are few prominent indicators that have become the subject of global policy initiatives.

By multiplying the calculated average weights of each dimension (Table 4) with the average weights of each category (Table B.2), we were able to develop a list which ranks indicator importance as judged across the expert panel. This list is displayed in Table 5, alongside the standard deviation, calculated as that of the product of two independent random variables. We observe that all economic categories scored lower than any other category; social categories representing more quantitative metrics relating to the activities of firms, namely 'jobs created' and 'customer environmental awareness' also performed relatively poorly. Of course, once again we caveat this with our caution in relation to overinterpreting these statistics, and note given the large standard deviation, an overlap of categories.

4.4. Round 3: group AHP and discussion

Qualitative observations were collected from the AHP group facilitators, as well as the final group discussion. Most participants found the AHP task of rounds 2 and 3 challenging. Partly this was due to

Table 5
Ranking of each indicator category following round 2 based upon the average weights of each category multiplied by the dimensional weights.

Indicator Category	Dimension	Average Weight	S.D.
GHG Emissions	Environmental	9.8 %	4.7 %
Social cost of waste	Social	9.5 %	4.7 %
Participatory planning	Social	8.2 %	5.1 %
Quality of work	Social	8.1 %	4.4 %
Gender & equalities	Social	8.0 %	3.5 %
Water consumption	Environmental	7.8 %	2.7 %
Virgin resource usage	Environmental	7.5 %	5.0 %
Air pollution	Environmental	7.4 %	3.0 %
Energy Usage	Environmental	6.3 %	3.0 %
Waste produced	Environmental	6.3 %	3.1 %
Jobs created	Social	5.8 %	4.2 %
Customer Environmental Awareness	Social	4.4 %	4.2 %
Surplus distribution	Economic	3.3 %	3.1 %
Equitable investments	Economic	2.5 %	2.3 %
Supply chain risks	Economic	2.0 %	1.8 %
Products quality	Economic	1.9 %	1.7 %
Cost of Production	Economic	1.3 %	1.2 %

unfamiliarity with what was perceived as a technically demanding task, but also related to the challenge of comparing a diverse set of elements, and translating thoughts into numerical terms. It is notable that almost none of the matrices were fully consistent, even those that met the acceptability threshold. These difficulties are inherent to AHP as a method (see e.g. Ozdemir, 2005, who suggests limits to participant processing), and relate to broader epistemological challenges to indicator paradigms relating to the commensurability of qualitatively different categories (Martinez-Alier et al., 1998), which the difficulty in achieving consistent judgments seems to point to.

Participants articulated frustration at the relatively vague definitions given for each dimension and category, leaving descriptors somewhat arbitrary and open to divergent interpretation. The difficulties this presented at an individual level were compounded in the group discussion section when participants discovered that members of the group had competing interpretations of several terms. Ambiguity was also observed in the normative dimension of various categories, for example a discussion took place relating to 'energy usage' in relation to renewable versus non-renewable energy. Within round 3, a discussion also arose relating to the meaning of the economic dimension, and its relative importance for developing nations contrasted to the dominant positionality of their group within the global north. Here participants noted the limitations of this discussion when their conversation and thereby choices were being influenced by the location of their voices with a particular dominant positionality. Indeed, the particular issue of 'economic development' and the global south has been debated extensively within the literature on degrowth (see e.g. Escobar, 2015, and Dengler and Seebacher, 2019).

One area where groups struggled to reach consensus was whether the environmental dimension was more important than the social dimension. The differing views here appeared to be as much epistemological and axiological, with participants raising the dependence of society on its environment. Others in opposition argued that the reason we want a healthy environment is for society, and thus the social dimension should be judged more important. Regardless of the individual values articulated, it was clear that there was broad consensus that the environmental and social dimensions are closely linked, and that both should be judged much more important than the economic. These debates reflect a wider discourse surrounding anthropocentric versus ecocentric views relating to sustainability (Costanza et al., 2017), a socio-environmental dialectic (Harvey, 1993), and issues pertaining to value incommensurability (Martinez-Alier et al., 1998).

The third Delphi round aimed to collect qualitative data from the panel; this was intended to a) sense check the quantitative findings and garner feedback on the process, and b) feed into the theoretical framing of the indicator set. From the discussions that took place, it was clear that a notable number of participants were unhappy with the use of the approach to ranking the importance of indicator categories, and various panellists objected to ranking entirely. Whilst it is somewhat unusual within a Delphi process to present the participants an opportunity to directly discuss with each other their responses (Keeney et al., 2001), we judged this important from a validation perspective as well as allowing for broader critical reflection on the construction of the intended indicator framework. The resultant list of indicator categories (Table 5), whilst operationally representing an aggregate 'consensus' of the expert panel, found itself in contestation. This presented a dilemma for the research team in how to proceed with these findings, leading to the reflection presented in the next sections.

4.5. Limitations

Despite our attempts to minimise bias, and select indicators in a critical and rigorous manner, our methodology retains various limitations to reflect upon, and maybe improve in future work. Firstly, due to the nature of the panel (which we acknowledge being associated with a single wider project introduces its own biases in its very particular selection), we assumed a rough consensus in terms of the panel's understanding of CE. As our 'failure' to find consensus on indicator categories shows, this assumption is questionable. An alternative approach could have been to use the Delphi process to derive consensus on a definition, though in many respects this is what we did do (see e.g. Goertz and Mahoney, 2012 for an outline of varying interpretations on the relation between conceptualisation and measurement), and it is thus unclear whether this would have circumvented the problems we faced. We also recognise the limitations of our goal to develop a dashboard of indicators, which despite its intention to distil information, can remain complex and difficult to allow for informed decision making by themselves. Future work might benefit from instead focusing on areas of dissensus, where areas of disagreement arise, with a particular focus on the positionality of the participants here, e.g. do participants from the global south and global north have different views on the economic dimension? Why? How does gender, or academic discipline shape responses? Unfortunately we did not collect this information from participants within this study.

The first Delphi round is usually structured to have open ended questions with the objective of generating ideas. In this study, we chose to provide experts with preliminary information (the results of the literature review), to let them focus on refining an existing list of indicator categories. Rather than starting from zero, we built on pre-existing CE assessment frameworks, encouraging reflection on aspects that are usually overlooked. Nevertheless, this starting point provided a clear steer to the panel, and we would have likely ended up with a very different set of indicator categories if we began with an open discussion with no predefined information.

Rounds 2 and 3 were both held during the annual meeting of the project consortium, where most of the panel were present in the same location. Due to some participants attending online however, we decided to have an online group. Whilst allowing for a wider breadth of participants, this nevertheless suffered in terms of the differences in online vs in person interaction, and a bias in terms of the panellists who were unable to attend in person (many of these were located in Africa). In round 3, we allowed participants to directly discuss responses, relaxing the constraint of anonymity typical of Delphi studies (Keeney et al., 2001). This choice might have introduced some biases: knowing the identity of the person expressing a certain opinion could reduce the equal chance of expression and decrease the opportunity for less experienced participants to share their ideas. Nevertheless, we considered fostering moments of participatory and open discussions and decision-making to be more important. To minimise potential biases, we tried to ensure that the most experienced researchers respected the discussion space, allowing equal opportunity for all participants to contribute.

5. Reflection on doing better

In this section, we reflect upon the apparent failure of consensus, with reference to the challenges to indicator frameworks, and the questions to aid better indicator framework design presented in Purvis and Genovese (2023), and reproduced in Table 6. Our contribution in this section therefore acts to illustrate and contextualise the framework of challenges presented in Purvis and Genovese (2023) with the empirical example of our experience.

5.1. Implementation & outcomes

The initially conceived purpose of the study was to develop a dashboard of indicators to be used within a web tool which would allow stakeholders to assess a *just* transition to a CE within the context of supply chains. This goal was loosely conceived in the project's grant proposal, offering strategic ambiguity to allow for refinement as the project progressed. As such, it contained unknowns which demanded refinement: who are the stakeholders, and what do we mean by a just transition to a CE. We thus attempted to refine this purpose through

Table 6
Challenges and research design questions for indicator methods, reproduced from Purvis and Genovese (2023).

Step	Critical Challenges	Questions to aid research design
1) Implementation & outcomes	What outcomes are desired? Stakeholder engagement Data collection	What is the purpose of the study? What is the context of application? What is the novelty? What changes should be catalysed by the study? Are these realistic? What is the political orientation?
2) Theoretical framework	What theory? Handling complexity Confronting trade-offs	What literature strands is the study embedded in? How are key concepts understood? How is the epistemology of measurement conceived? What is the axiological orientation?
3) Selection	Which indicators are relevant? Says who? How should indicators be selected?	Does the concept to be measured define the selection or vice versa? How will arbitrariness be minimised? Will measurability or coverage be prioritised? How does selection reflect the political/axiological orientation and the theoretical framework?

collective workshops and discussions within the research team and members of the expert panel. Contradictions and competing priorities, as well as anticipating the users who might be engaged, meant that we ultimately opted for a general audience. Whilst good for maximising engagement, this lack of precision in the user profile for the dashboard may have contributed to the lack of consensus in which indicators were deemed important. Indeed, this was apparent from discussions, and each member of the panel would have their own understanding of for whom, how, and to what purpose the resultant framework would be applied.

Where we did expect to observe consensus across the panel, was in the anticipated novelty of the framework, in particular its critical view of the dominant ecomodernist and techno-centred conceptualisations of a CE. Consensus was largely observed here in how the importance of economic indicator categories were judged, and the more quantitative social categories. Yet, contradictions and contestations were revealed in the round 3 discussion in which a number of panellists questioned an indicator approach in its entirety. Arguments made echoed the discussion of transformative potential within Purvis and Genovese (2023): perhaps this critical approach is novel for an indicator framework, but is an indicator framework itself novel, or indeed useful for a critical study? Does the incrementalism of an indicator approach find itself in contradiction with a study which seeks to be critical and transformative? Yet, to what extent is a transformative approach possible in a research landscape dominated by funders who look towards ecomodernism as a panacea for maintaining a capitalist status quo?

On reflection, the changes intended to be catalysed by the study thus found themselves shaded by various degrees of pragmatism amongst the interpretation of the expert panel and research team itself. Nominally the goals articulated within the grant proposal indicated an ambitious rollout of the indicator framework across a myriad of stakeholder groups in order to 'support decisions' and 'shape policy'. A reassessment of realistic goals however, led to the stated ambition of the indicator framework as part of a "learning tool, encouraging the user to consider and explore new avenues of thinking" (Purvis et al., 2023a) Here we attempted to follow Purvis and Genovese's (2023) articulation of better indicator frameworks in terms of embracing the political, shaping narratives, & uncovering overlooked avenues. Despite this pragmatism, and as a number of panellists raised in the round 3 discussion, we found ourselves questioning whether this was a realistic demand for an indicator framework.

5.2. Theoretical framework

The theoretical orientation of the framework was somewhat clear from the outset of the project, due to its elaboration in the proposal. This took a critical approach to the dominant ecomodernist CE discourse, emphasising the importance of social dimensions. Nevertheless, this identified lacuna meant that the preliminary selection of indicators via literature review required correction to account for this overlook of social aspects. To do this, we adopted the 'three pillar' conception of sustainability dimensions, and aimed to 'balance' coverage across these three dimensions. We also gave preference to expert suggestions collected from round 1 of the Delphi study, over those found within the literature. Whilst aware that this three dimensional approach may be problematic (Purvis et al., 2019), we intended this initial framework to be complemented by concurrent theoretical work within the wider project relating to justice.

The justice dimension was thus only loosely conceived initially, despite a series of discussions, seminars, and readings which were shared and accessible for members of the group. Related to this, it became apparent from discussions amongst the panellists within rounds 2 and 3, that participants didn't possess shared understandings of various terms that were used in the description of dimensions/categories to be compared. Whilst effort was made to provide a short description of each indicator category, various debates nevertheless played out in discussions, related e.g. to the terms and geography of application, and perceived necessity of economic development in the Global South. This recalls the warning in Ramsey's (2015) discussion around not defining sustainability: "we cannot define our way to clarity" (p1085). Even when we felt there was a shared understanding of the literature in which we assumed the framework to be embedded in, this turned out not to be the case, and the broad interdisciplinarity of the group brought in other unexpected conflicts, such as what justice means and how it should be understood. Whilst a concurrent work package was working on developing a theoretical understanding of justice within the context of CE (Pansera et al., 2024), this took a literature based rather than consensus driven approach.

The indicator dashboard did not evolve to a sufficient degree to explicitly address the handling of complexity, and confrontation of trade-offs. Nevertheless, the choice of a dashboard over a composite indicator was selected as we felt the latter was reductionist, and reduced the transparency for the user. Thus a dashboard approach in many ways offloads the issue of trade-offs to the user to make sense of, presenting a broad range of information for the user to interpret how they wish. Whilst the Delphi approach did lead to a ranking of the relative importance of indicators as perceived by the expert panel, this was done for purposes of selection rather than intended for instrumentalising weighting and comparison within the final framework itself.

5.3. Selection

The selection process formed the bulk of our endeavours, and we seemingly addressed all of the questions in the second column of Table 6. As described in Section 3, we used a literature review and Delphi process to determine which indicators were relevant and how to select them. We thus enriched literature findings with expert consultation to select more appropriate indicators. We chose to consult a narrow expert panel drawn from the wider project team, which we believed would share some theoretical coherence and common values. The level of consensus reached by the expert panel varied, both across the dimensions, and across specific indicator categories, as outlined in Section 4.4. Whilst consensus was mechanistically reached in a quantitative way, we, and the expert panel itself, were unsatisfied with this supposed consensus.

The third column of Table 6 provides various more philosophical questions that were engaged with to different extents. In many ways there was a disconnect between the concept being measured and our

selection. This has already been alluded to in Section 5.2 where we note the biases of the initial literature, and detail our use of the 3 sustainability dimensions whilst a more theoretically driven understanding of justice was being developed concurrently. Thus in this initial iteration of indicator development, the selection (of indicators) would have defined the concept (a just transition to a CE), though we anticipated that this may have been refined to reflect later theoretical developments. Thus the political/axiological dimension was largely reflected through the positionality of the expert panel, and the focus of the consensus building was framed as indicator selection rather than explicitly developing a shared understanding of the concept. In terms of selection based upon measurability versus coverage, our decision to prioritise indicator categories over specific indicators necessarily favoured coverage. Whilst each of the categories that were created following subsequent rounds were created with the eye to serving as an umbrella for a number of specifically measurable indicators, with some examples provided, we ultimately did not move to a position where we began to break down categories into specific measurable indicators.

Arbitrariness in indicator selection is in many ways unavoidable, as it is inherent in a selection of a small number of indicator categories from an infinite set of possibilities. Whilst aware of this, we endeavoured to minimise arbitrariness, through a definite selection process consisting of narrowing down a literature based selection based on expert consideration. Despite this, the results of the consistency analysis in the AHP exercise show that whilst such a quantitative approach is aimed at bringing rigour, participants still struggled to choose between the importance of various categories. This was confirmed by a number of participants who directly suggested that some of the choices they made in prioritising one category over another was largely arbitrary. It should be noted that arbitrariness is a subjective concept which doesn't lend itself to quantification.

6. Failure of consensus: contesting an indicator dashboard

The Delphi approach was specifically identified as an instrument to attempt to bridge some of the technical and episto-ontological challenges outlined in Purvis and Genovese (2023), through an approach that sought consensus, and approval, from a panel of experts. This was intended to 1) foster a shared understanding amongst the project and democratise the approach taken, and 2) draw on specific expertise that existed within the panel, relating to e.g. disciplinary orientation, geographical location, and depth of knowledge in particular dimensions. Such a democratically conceived approach is somewhat at odds with a top down, more theoretically driven one in which a single researcher (or team of researchers) maps indicators to existing theory, though as identified above we attempted to combine both theory and consensus. Despite our best intentions, the discussion within round 3 of the Delphi led us to conclude that whilst we had succeeded in drawing on varied expertise, we had failed to reach a shared understanding or reach a democratic consensus to indicator selection.

6.1. *Contesting consensus*

Consensus is typically normatively conceived as the realisation of a radical democracy (Polletta and Hoban, 2016), a means of conveying authority (Beatty and Moore, 2010), or a pragmatic necessity (van den Hove, 2006). The desire for consensus is often maintained despite acknowledged challenges to achieving it, including the paralysis of decision making (Mintz and Wayne, 2016), the alienation of those who are not included in discussions, and dominance of more powerful or vocal actors (Keeney et al., 2001). In academic pursuits, consensus may typically be pursued to lend an authoritative voice, or to provide a semblance of objective truth to policy makers or publics (Pearce et al., 2017). As described by Barrios et al. (2021), the Delphi process is widely recognised as a "valuable technique for reaching consensus about specific issues when empirical evidence is scarce or contentious" (p1). Our pursuit of consensus in developing an indicator framework was informed by these goals, primarily to sense check a framework by 'experts', as well as to foster buy-in from actors we wish to use and disseminate the resultant framework.

It is worth reflecting on the quantitative slant of our chosen method of AHP and Delphi methods to reach consensus. Whilst arguably we did reach a 'consensus' within the paradigmatic framework of these approaches, in the literal sense of deriving a set of quantitative weights for each indicator category, this did not cohere with the qualitative understanding of consensus of the participants present. It is notable too that the averaging involved in the AHP process disguises the variance which was seen across respondents (see Tables 3 and B.2 for the max and min weightings for each indicator category). Whilst we may conclude that the panel ranked the indicator categories in the way presented by the final outcome, an analysis of the variance would reveal that the placing of specific categories is subject to contestation. It is also worth reflecting on how this variance can relate to contextual factors, e.g. the respondent's experiences and background, rather than being subject to a more standard statistical distribution.

As detailed above, our bottom-up expert consensus driven approach holds an uneasy relation with a more theoretically driven approach. Whilst we intended the derived set of indicators to be further refined to respond to a concurrent theoretical strand of the project, we did not reach this stage due to the perceived consensus failure. Arguably the theoretical framework would be a more appropriate target for attempting consensus than the selection of indicators. This is something that has been explored to various extents elsewhere, such as in the study of Van Schoubroeck et al. (2022) which uses a Delphi approach to develop a "shared vision on a circular food economy (CFE)". Here, the first Delphi round prompts participants to articulate their understanding of a CFE in terms of production and consumption, and thematic coding is developed from responses which inform the development of indicator themes. Such an approach arguably presents a more robust theoretical framework than our use of sustainability dimensions as organising structures. Yet it is also notable that this study does not include a step, such as our round 3, that allows for reflexive responses from the participants in relation to the consensus building exercise, thereby allowing for a consensus to be determined by the research team without problem. It is therefore unclear if moving the target of consensus further up the road would circumvent the challenges we present, or is merely shifting the bounds of a more fundamental problem. Indeed, whilst formally focusing on theory development as a target for consensus did not take place in our study, discussions did take place amongst the research team and various subsets of the expert panel, and the questionnaire from round 1 did prompt participants to provide their understanding of a just CE, even if this wasn't formally used for theory development.

It is thus pertinent to revisit the question of why we wished to reach consensus within this study in the first place. The answer to this touches on all three motivations stated at the start of this subsection. From a very pragmatic sense, there came the need to progress a work package within a structured project with deadlines and time limited funding, and a desire to include an attempted democratic steer through a participatory approach. We also sought to provide an amount of academic authority in the selection of indicators that would be presented to users. The consensus building thus was anticipated to build a more authoritative and concrete academic voice that could bring to the fore important and often forgotten (social) justice aspects linked within the sustainability transition. Given the very political nature of the problem at hand however, i.e. a transformation or displacement of the capitalist mode of production, it is perhaps unsurprising that we should fail to reach consensus. Indeed, we may not wish to pursue a consensus which compromises the fundamentals of a vision that is radical and transformative, something Polletta and Hoban (2016) emphasise in terms of an uneasy relationship between ideological purity and movement building.

What does this mean for the direction of CE research? We concur

strongly with the thoughts of Berry et al. (2022) here, that convergence around a singular framework or common understanding is neither necessary nor desirable. Instead the authors call for divergence, pluralism, and an *ecology of practices* (Stengers, 2005), whilst remaining highly critical of sanitised and depoliticised visions of a CE. Thus more debate and highlighting of difference is needed. Here we hark to Korhonen et al., who in 2018, described CE as an "essentially contested concept", and argued that CE scholars were too busy in the practice stage developing indicators, tools, and techniques, rather than debating the essence of CE at a paradigmatic level. What has happened in the meantime? Not too much, but we can now point to critical discussion of CE at the paradigm level, as outlined in Section 2, and the critique of mainstream CE discourse of techno-optimistic and eco-modernist. But have we yet reached the moment of developing practice level knowledge for a paradigmatically different CE? We might argue that the empirical study we have presented here was such an attempt, yet its adherence to an indicator set in many ways locates it as an uncertain break.

6.2. Contesting the transformative potential of indicators

Purvis and Genovese (2023), despite suggesting ways to improve indicator methods, are ultimately very critical about their transformative potential. They suggest that despite their abundance within the academic literature, indicators provide at best an "incrementalist adjustment to the status quo" (p10). As Kaika (2017) has argued, 30 years after the Brundtland Commission we are still using the same failed methodological, political, and technological frameworks, questions, and methods. There is indeed some irony in suggesting that our novelty is in our critical reconceptualisation of a technocratic and ecomodernist CE, but then fall back to using the same technocratic tools & narrow methods. We observe a fundamental contradiction in indicators which are literally measuring incremental changes within the frames of the current system, and transformation or transition, which in our view requires rupture, and the dismantling of structures of systemic injustice. As the reflection within Purvis and Genovese (2023) suggests, we may rethink the purposes of indicator frameworks, beyond that of measurement, in terms of their potential power to shape new narratives: as political artefacts. Merry (2016) also suggests that in order to draw attention to a social problem, it helps to show (through measurement) that the problem is significant (p129). Yet, we should also be realistic about the amount of institutional power required for the mainstreaming of such frameworks (Merry, 2016). We can also claim that despite its hegemonic position in contemporary governance, quantification and measurement is not the only way to shape narratives and convey significance.

Regardless, the epistemological challenges to an indicator approach remain largely unsolved: we can explicitly politicise our framework, and present it as a learning tool for shaping new narratives, yet it remains reductionist, arbitrary, and neglectful of aspects which are not easily quantifiable. It remains unclear how change can be catalysed from a better selection of indicators, and we see little evidence of previous indicator sets having remarkable outcomes (Gahin et al., 2003; Kaika, 2017; Merry, 2016; de Olde et al., 2017). In placing too much emphasis on the indicator, as a quantified representation and proxy for a broader qualitative goal, we mystify and mask this goal itself. Like GDP, or even Carbon emissions, we create abstract idols whilst sweeping a messy concrete reality awash with power differentials, historical inequities, systemic inequalities, and structural violence under the neat rug of numbers on a spreadsheet.

Is there any space for indicator frameworks in the vision for a radically transformative future? Can we dismantle the master's house with the master's tools? We are not suggesting that measurement is counterrevolutionary; it is clear that sustainability assessment tools are value articulating institutions (Gasparatos, 2010), which could also have positive effects, e.g. many of the SDGs, if met, could have significant tangible benefits for many people. We may even consider more dynamic application of indicator frameworks which evolve over time to include new indicators and remove old ones. Yet there are also dangers in selecting the wrong goals, which may shape the world around them in powerful and catastrophic ways (e.g. GDP, see O'Neill, 2012), or lead to unexpected and adverse impacts (e.g. the eradication of sparrows during Mao's Great Leap Forward; Rich, 2003). Indicator frameworks may have their place in the right contexts, but they should not be seen as a substitute for visioning radically alternative futures, and debating how we get there. The 2015 Paris Agreement consensus on the need to reduce carbon emissions was hailed as a monumental moment (Falkner, 2016), and yet in the absence of a shared radical vision of an alternative future, a decade later little progress has been made. Just because we are monitoring variables does not mean we are taking action or changing the system. When we began our empirical study, we were thinking that simply by changing assumptions, values, and worldviews we could measure progress towards a different outcome (from a CE to a just CE), yet we realised that the idea of "what gets measured can be improved" comes from that very same worldview we are critiquing.

6.3. Where to next?

The reflexive bent that we have taken to our attempt to derive indicator categories relating to a critical and just orientation of a CE has nevertheless been generative, particularly in terms of the reflections that we have illustrated relating to the nature of consensus and challenges of indicators. This has uncovered various potential avenues for future research. In particular we think it fruitful to consider further how consensus building in the theoretical framing stage impacts the development of indicator sets, as well as how a dynamic indicator framework that might evolve over time may be implemented. Theoretical framing is important, and as shown within Merry, 2016 ethnographic study, the choice of measurement paradigm has huge implications for how a phenomena is understood and addressed (p108). There are a lot of issues to think about here, and we view deeper consideration around theories of change to be a particularly pertinent area of pursuit here in relation to inducing transformation.

We also emphasise the need to contest consensus, particularly in the mechanistic form of 'consensus' that can arise from Delphi and AHP approaches; consensus which approximates the summed average of views over a range of experts. True consensus should be seen as something more than a mechanistic quantitative exercise. We might consider alternative approaches to addressing value conflicts, such as agonistic models, and dissensus management (Rettschlag, 2024). But again, it is useful to consider our purposes; why is it necessary to address value conflicts, and do we risk erasing marginal perspectives, and diluting our ideological orientation by doing so unreflexively? It might be more useful to drill into areas where disagreement arises and ask why this is, with reference to power dimensions, and historically embedded material conditions. As emphasised, CE imaginaries are inherently political, reflecting at their most transformative, fundamental changes to social systems and relations. Given this, seeking a semblance of 'objectivity' seems a contradiction of terms. Qualitative illustration of messy reality seems a potential antidote here, through case studies, examples and counterexamples, Merry (2016) here champions the power of ethnographic approaches (p222). Contextualisation is important here to resist the potential homogenisation that quantitative indicator approaches might engender. As ever for the critical social scientist, fruitful avenues lie in probematisation and opening up problems, rather than seeking to reduce complexity and offer uncaveated solutions.

7. Conclusions

In this work, we set out to empirically test and explore the challenges to the development of indicator dashboards as laid out within the theoretical contribution of Purvis and Genovese (2023). We furthered these reflections by taking a typical approach to the development of a

dashboard: seeking to build upon a literature review of indicator categories (Calzolari et al., 2022) with a Delphi process that attempted to derive consensus on these categories across experts. This led to a quantitative convergence in the selection of indicators, but the validity of this 'consensus' was brought into question by the qualitative feedback from participants, some of whom rejected the selection process in its entirety. Whilst we started with the best intentions, informed by the various technical and epistemological challenges to such an approach, as articulated within Purvis and Genovese (2023), we ended with a very subjective framework. Even in a close community where we expected to observe similar views, we saw various divergences, and a failure to reach consensus qualitatively. This preceded any later technical issues we may have encountered relating to measurability and data availability.

The challenges we encountered are, we argue, inherent to the derivation of indicators, and whilst there is still room for improvement in our methods, we suggest there is no transformative potential to be found. In our discussion, we have contested both the normative orientation of consensus, and the role of indicators themselves. Within the CE discourse, as well as more broadly, both these aspects are oriented towards depoliticising what amounts to a call for a radically different global system of production. In masking this goal with a focus on quantitative markers, and debates around which aspects are important to measure, we obscure productive visioning of what this new system would look like and how we could get there. Instead of attempting to bound concepts, reduce complexity, and hunt for 'objectivity', we argue that fruitful avenues lie in complementing quantitative approaches with qualitative perspectives which instead illustrate complexity, and contextuality. To transform society we must first look to transform our modes of enquiry.

CRediT authorship contribution statement

Ben Purvis: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Tommaso Calzolari:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Andrea Genovese:** Writing – review & editing, Writing – original draft, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgements

This paper was supported by the following European Union research programme: H2020-SC5-2020-2 scheme, grant agreement number 101003491, JUST2CE (A Just Transition to the Circular Economy) project.

Appendix A. Panel details

Table A.1
Details of the 'expert' panel who took part in all Delphi rounds. Note that we did not collect personal details in the survey round, so are unfortunately unable to provide accurate details on the 10 participants who took part in round 1 but not rounds 2 and 3.

Role	*University Academic: Professorial (6)*	*University Academic: Post-doctoral (10)*	*University Academic: Doctoral (3)*	*Research Institute (1)*	*Research Consultant (1)*	*Policy (2)*	*Entrepreneurial (2)*	
English as a first language				*Yes (6)*				*No (19)*
Country of residence	*South Africa (3)*	*United Kingdom (3)*	*Portugal (3)*	*Greece (2)*	*Spain (7)*	*Zimbabwe (1)*	*Ghana (3)*	*Italy (3)*

Appendix B. Indicator categories

Table B.1
Initial indicator categories as derived from the systematic literature review. Reproduced from Corvellec et al., 2022. The occurrences column indicates the variability of coverage, with costs and emissions indicators being particularly common across all studies.

Dimension	Category	Examples of Measurable Indicators	Description	Occurrences	%
Economic	Costs	● Operational costs ● Facility location costs ● Transportation cost ● Reverse supply chain cost	Cost-based indicators, both at a company and at a supply chain level	112	55 %
	Profits	● Total CSC profits ● Profits from recovery activities including remanufacturing, recycling and disposal	Profit-based indicators, both at a company and at a supply chain level	50	25 %
	Time	● Time responsiveness of the network ● Delivery reliability of suppliers	Time responsiveness-based indicators, both at a company and at a supply chain level	18	9 %
	Quality	● Reliability of supply ● Quality level of the production ● Quality of the returns	Quality-based indicators, both at a company and at a supply chain level	14	7 %
	Risk	● Financial risk ● Value at risk ● Conditional value at risk ● Variability index	Risk-based indicators associated to uncertainty (e.g. of demand, collection)	12	6 %

(*continued on next page*)

Table B.1 (*continued*)

Dimension	Category	Examples of Measurable Indicators	Description	Occurrences	%
		● Downside risk			
	Profitability	● Net Present Value ● Return on Equity ● Return on Assets	Profitability-based indexes, measuring	9	4 %
Environmental	Emission equivalent	● Climate Change ● Greenhouse gases ● Global Warming Potential	CO2 eq. emissions associated with supply chain	90	44 %
	Waste	● Waste Landfilled ● Recycled waste ● Recovered waste ● Recyclability and ease of disassembly	Residual waste produced and landfilled or recovered by supply chain activities	35	17 %
	Energy usage	● Energy use ● Cumulative energy demand ● Renewable energy use ● Energy self-sufficiency	Energy-based indicators associated with supply chain	32	16 %
	Virgin resources usage	● Abiotic depletion of resource ● Mineral, fossil & renewable resource depletion	Virgin resource use associated with supply chain material consumption	26	13 %
	Water	● Water depletion ● Water emissions ● Water use	Water used or contaminated	26	13 %
	Air emissions	● Particulate Matter ● Respiratory inorganics	Other air emissions associated with supply chain	22	11 %
	Acidification	● Terrestrial acidification ● Marine acidification	Acidification potential associated with supply chain processes	19	9 %
Social	CSC jobs created	● Number of fixed and variable jobs ● Number of drivers hired for transportation	Employment opportunities provided by the CSC	15	7 %
	Organisational H&S compliance	● Compliance with the ILO guidelines	Measures of compliance to H&S Guidelines for the jobs created in the CSC	7	4 %
	Quality of work	● Work damages ● Number of accidents, lost ● Employee turnover	Measures of quality of the jobs created	7	3 %
	Training	● Average hours of training ● Training on skills for employability	Indicators of the training provided to workers	4	2 %
	Expenditure on Benefits for employees	● Food ● Transportation ● Pension	Indicators of benefits provided to the workers	4	2 %
	Customer environmental awareness	● Enlightening customers to return end of used product ● Customer incentives for recovery from discarded product	Indicators of environmental awareness of the customers	3	1 %
	Social cost of waste	● Penalty cost of disposal	Social cost of waste produced. Sum of disposal cost and of the cost for the recycler	2	1 %

Table B.2
Weightings of each indicator category following round 2 of the Delphi (prior to normalisation with respect to dimensional results).

Social Categories	Mean	S.D.	Max	Min
Jobs created (JC)	13.1 %	9.4 %	33.2 %	2.6 %
Social cost of waste (SCW)	21.6 %	9.8 %	48.6 %	3.2 %
Customer Environmental Awareness (CEA)	10.0 %	9.8 %	40.5 %	2.0 %
Participatory planning (PP)	18.6 %	11.4 %	50.5 %	4.2 %
Gender & equalities (G&E)	18.2 %	7.1 %	30.0 %	5.2 %
Quality of work (QW)	18.5 %	9.8 %	40.3 %	2.5 %
Environmental Categories				
Waste produced (WP)	14.0 %	6.0 %	25.5 %	3.9 %
GHG Emissions (GHG)	21.8 %	9.3 %	39.5 %	5.3 %
Energy Usage (EU)	14.0 %	5.6 %	26.7 %	4.9 %
Air pollution (AP)	16.4 %	5.3 %	32.1 %	5.9 %
Water consumption (WC)	17.3 %	4.2 %	25.0 %	9.2 %
Virgin resource usage (VRU)	16.5 %	10.0 %	52.9 %	2.6 %
Economic Categories				
Cost of Production (CP)	11.4 %	8.4 %	35.5 %	2.4 %
Supply chain risks (SCR)	17.8 %	12.5 %	43.4 %	3.7 %
Products quality (PQ)	17.6 %	12.7 %	56.5 %	4.1 %
Equitable investments (EI)	22.9 %	13.3 %	46.9 %	2.9 %
Surplus distribution (SD)	30.3 %	19.3 %	65.7 %	7.0 %

Appendix C. AHP generalities

Developed by Saaty (1980), Analytical Hierarchy Process (AHP) is a flexible, yet structured, methodology which enables an individual (or a group of individuals) to define a specific problem and derive a solution based on the individual's (or the group's) own experience of that problem. AHP is a

method that encourages respondents to make subtle trade-offs in non immediately quantifiable attributes. The primary reason for using the AHP is to improve the understanding of how such trade-offs are understood. The end result of AHP is a measure of the decision maker's relative preference of one attribute over another attribute. It is important to remember that AHP is not a statistically based methodology, there is no requirement about the sample size; AHP was originally developed to enable a single decision maker to select an alternative amongst multiple alternatives.

The methodology has since been extended to enable the use of AHP in group decision making where the 'single' decision maker is actually a group of *m* people. In this case, use of the geometric mean of the comparison ratings is typically used in lieu of individual ratings. The AHP was used for its capability of managing, through a hierarchical structure, a very large number of criteria and sub-criteria. Specifically, in our case, the relative degree of importance of sustainability dimensions, and, within each dimension, of identified categories of indicators, were assessed through the application of AHP.

As such, participants were asked to perform two types of pairwise comparisons: (1) pairwise comparisons of the specific categories of indicators within each general sustainability dimension, and (2) pairwise comparisons of the general sustainability dimensions.

The survey group's pairwise comparisons were synthesised into four comparison matrices (one comparison matrix for indicator categories for each sustainability dimension, and one comparison matrix of the general sustainability dimensions).

Each of the 4 matrices assumed the following form: $A = [a_{ij}]$, where a_{ij} represents the pairwise comparison rating for elements *i* and *j*. Below, details about the pairwise comparison scale which is universally used in AHP are provided, and these were provided to participants:

- 1: Equal importance of two elements
- 3: Moderate importance of one over another
- 5: Strong or essential importance
- 7: Very strong or demonstrated importance
- 9: Extreme importance 2, 4, 6, 8: Intermediate Values

Four main axioms underlie the theoretical validity of comparison matrices (Saaty, 1986):

- Reciprocal comparison: if $a_{ij} = x$, then $a_{ji} = 1/x$
- Homogeneity: if characteristics i and j are judged to be of equal relative importance then, $a_{ij} = a_{ji} = 1$ with $a_{ii} = 1$ for all *i*.
- Independence: when expressing preferences under each criterion, each criterion is assumed to be independent of the properties of the decision alternatives.
- Expectations: when proposing a hierarchical structure for a decision problem, the structure is assumed to be complete.

Given the reciprocity property, only $n(n-1)/2$ actual pairwise comparisons are needed for an $n \times n$ comparison matrix. However, it must be mentioned that complete consistency in rating alternatives is rarely the case given that subjectivity is involved in formulating judgments. In order to evaluate consistency of comparison matrices, Saaty (1977, 1980) proposed the eigenvector method, which derives the vector of priority weights and allows testing for inconsistency. In case of perfect consistency, $AW = nW$, where A is the $n \times n$ comparison matrix and $W = (w_1, w_2, \ldots, w_n)$ represents the priority vector. However, when a certain degree of inconsistency exists amongst subjective pairwise comparisons of items, Saaty (1977, 1980) proposed the following redefinition: $AW = \lambda_{max}W$, where λ_{max} is the maximum eigenvalue of matrix A.

The difference deviation of max from n measures the degree of inconsistency within the $n \times n$ matrix A. The consistency index (CI) for an $n \times n$ comparison matrix with largest eigenvalue, λ_{max}, is $CI = [(\lambda_{max} - n)/(n-1)]$. In order to assess the level of inconsistency, Saaty (1980) denotes the consistency index for a randomly generated $n \times n$ matrix as RI. Values for RI are reported in Table C1 below.

n	1	2	3	4	5	6	7	8	9	10
RI	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

This allows defining a consistency ratio as $CR = CI/RI$. Values of $CR \leq 0.1$ are desired, as an acceptable level of inconsistency in human judgement is deemed to be not higher than 10 % of the one of a matrix which is randomly generated without any preoccupation about consistency. Higher CR values imply an unacceptable level of inconsistency and respondents would be asked to revise their pairwise comparison ratings.

Data availability

Data will be made available on request.

References

Barrios, M., Guilera, G., Nuño, L., Gómez-Benito, J., 2021. Consensus in the delphi method: what makes a decision change? Technol. Forecast. Soc. Chang. 163, 120484. https://doi.org/10.1016/j.techfore.2020.120484.

Beatty, J., Moore, A., 2010. Should we aim for consensus? Episteme 7 (3), 198–214. https://doi.org/10.3366/epi.2010.0203.

Berry, B., Haverkamp, J., Isenhour, C., Bilec, M.M., Lowden, S.S., 2022. Is convergence around the circular economy necessary? Exploring the productivity of divergence in US circular economy discourse and practice. Circ. Econ. Sustain. https://doi.org/10.1007/s43615-022-00199-1.

Bleischwitz, R., Yang, M., Huang, B., Xu, X., Zhou, J., McDowall, W., Andrews-Speed, P., Liu, Z., Yong, G., 2022. The circular economy in China: achievements, challenges and potential implications for decarbonisation. Resour. Conserv. Recycl. 183, 106350. https://doi.org/10.1016/j.resconrec.2022.106350.

Böhringer, C., Jochem, P.E.P., 2007. Measuring the immeasurable—a survey of sustainability indices. Ecol. Econ. 63 (1), 1–8. https://doi.org/10.1016/j.ecolecon.2007.03.008.

Boulkedid, R., Abdoul, H., Loustau, M., Sibony, O., Alberti, C., 2011. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. PLoS ONE 6 (6), e20476. https://doi.org/10.1371/journal.pone.0020476.

Boyer, R., Peterson, N., Arora, P., Caldwell, K., 2016. Five approaches to social sustainability and an integrated way forward. Sustainability 8 (9), 878. https://doi.org/10.3390/su8090878.

Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. Qual. Res. Psychol. 3 (2), 77–101. https://doi.org/10.1191/1478088706qp063oa.

Calzolari, T., Genovese, A., Brint, A., 2022. Circular economy indicators for supply chains: a systematic literature review. Environ. Sustain. Indic. 13, 100160. https://doi.org/10.1016/j.indic.2021.100160.

Corona, B., Shen, L., Reike, D., Rosales Carreón, J., Worrell, E., 2019. Towards sustainable development through the circular economy—a review and critical assessment on current circularity metrics. Resour. Conserv. Recycl. 151, 104498. https://doi.org/10.1016/j.resconrec.2019.104498.

Corvellec, H., Stowell, A.F., Johansson, N., 2022. Critiques of the circular economy. J. Ind. Ecol. 26 (2), 421–432. https://doi.org/10.1111/jiec.13187.
Costanza, R., de Groot, R., Braat, L., Kubiszewski, I., Fioramonti, L., Sutton, P., Farber, S., Grasso, M., 2017. Twenty years of ecosystem services: how far have we come and how far do we still need to go? Ecosyst. Serv. 28, 1–16. https://doi.org/10.1016/j.ecoser.2017.09.008.
de Olde, E.M., Moller, H., Marchand, F., McDowell, R.W., MacLeod, C.J., Sautier, M., Halloy, S., Barber, A., Benge, J., Bockstaller, C., Bokkers, E.A.M., de Boer, I.J.M., Legun, K.A., Le Quellec, I., Merfield, C., Oudshoorn, F.W., Reid, J., Schader, C., Szymanski, E., Manhire, J., 2017. When experts disagree: the need to rethink indicator selection for assessing sustainability of agriculture. Environ. Dev. Sustain. 19 (4), 1327–1342. https://doi.org/10.1007/s10668-016-9803-x.
Dengler, C., Seebacher, L.M., 2019. What about the global south? Towards a feminist decolonial degrowth approach. Ecol. Econ. 157, 246–252. https://doi.org/10.1016/j.ecolecon.2018.11.019.
Desiderio, E., García-Herrero, L., Hall, D., Segrè, A., Vittuari, M., 2022. Social sustainability tools and indicators for the food supply chain: a systematic literature review. Sustain. Prod. Consum. 30, 527–540. https://doi.org/10.1016/j.spc.2021.12.015.
El Gibari, S., Gómez, T., Ruiz, F., 2019. Building composite indicators using multicriteria methods: a review. J. Bus. Econ. 89 (1), 1–24. https://doi.org/10.1007/s11573-018-0902-z.
Ellen MacArthur Foundation & ANSYS Granta, 2019. Circularity Indicators: An Approach to Measuring Circularity. Methodology. 2019 Revision. Ellen MacArthur Foundation.
Escobar, A., 2015. Degrowth, postdevelopment, and transitions: a preliminary conversation. Sustain. Sci. 10 (3), 451–462. https://doi.org/10.1007/s11625-015-0297-5.
European Commission, 2020. A New Circular Economy Action Plan for a Cleaner and more Competitive Europe.
Falkner, R., 2016. The Paris Agreement and the new logic of international climate politics. Int. Aff. 92 (5), 1107–1125. https://doi.org/10.1111/1468-2346.12708.
Gahin, R., Veleva, V., Hart, M., 2003. Do indicators help create sustainable communities? Local Environ. 8 (6), 661–666. https://doi.org/10.1080/1354983032000152752.
Gan, X., Fernandez, I.C., Guo, J., Wilson, M., Zhao, Y., Zhou, B., Wu, J., 2017. When to use what: methods for weighting and aggregating sustainability indicators. Ecol. Indic. 81, 491–502. https://doi.org/10.1016/j.ecolind.2017.05.068.
Gasparatos, A., 2010. Embedded value systems in sustainability assessment tools and their implications. J. Environ. Manag. 91 (8), 1613–1622. https://doi.org/10.1016/j.jenvman.2010.03.014.
Gasparatos, A., El-Haram, M., Horner, M., 2008. A critical review of reductionist approaches for assessing the progress towards sustainability. Environ. Impact Assess. Rev. 28 (4), 286–311. https://doi.org/10.1016/j.eiar.2007.09.002.
Genovese, A., Pansera, M., 2021. The circular economy at a crossroads: technocratic eco-modernism or convivial technology for social revolution? Capital. Nat. Social. 32 (2), 95–113. https://doi.org/10.1080/10455752.2020.1763414.
Goepel, K.D., 2018. Implementation of an online software tool for the analytic hierarchy process (AHP-OS). Int. J. Anal. Hierarchy Process 10 (3). https://doi.org/10.13033/ijahp.v10i3.590.
Goertz, G., Mahoney, J., 2012. Concepts and measurement: ontology and epistemology. Soc. Sci. Inf. 51 (2), 205–216. https://doi.org/10.1177/0539018412437108.
Harris, S., Martin, M., Diener, D., 2021. Circularity for circularity's sake? Scoping review of assessment methods for environmental performance in the circular economy. Sustain. Prod. Consum. 26, 172–186. https://doi.org/10.1016/j.spc.2020.09.018.
Harvey, D., 1993. The Nature of Environment: Dialectics of Social and Environmental Change, 29. Socialist Register. https://socialistregister.com/index.php/srv/article/view/5621.
Homrich, A.S., Galvão, G., Abadia, L.G., Carvalho, M.M., 2018. The circular economy umbrella: trends and gaps on integrating pathways. J. Clean. Prod. 175, 525–543. https://doi.org/10.1016/j.jclepro.2017.11.064.
Kaika, M., 2017. 'Don't call me resilient again!': The new urban agenda as immunology … or … what happens when communities refuse to be vaccinated with 'smart cities' and indicators. Environ. Urban. 29 (1), 89–102. https://doi.org/10.1177/0956247816684763.
Karapetrovic, S., Rosenbloom, E.S., 1999. A quality control approach to consistency paradoxes in AHP. Eur. J. Oper. Res. 119 (3), 704–718. https://doi.org/10.1016/S0377-2217(98)00334-8.
Keeney, S., Hasson, F., McKenna, H.P., 2001. A critical review of the Delphi technique as a research methodology for nursing. Int. J. Nurs. Stud. 38 (2), 195–200. https://doi.org/10.1016/s0020-7489(00)00044-4.
Kinyondo, A., Pelizzo, R., 2018. Poor quality of data in Africa: what are the issues? Policy Polit. 46 (6), 851–877. https://doi.org/10.1111/polp.12277.
Korhonen, J., Nuur, C., Feldmann, A., Birkie, S.E., 2018. Circular economy as an essentially contested concept. J. Clean. Prod. 175, 544–552. https://doi.org/10.1016/j.jclepro.2017.12.111.
Lowe, B.H., Genovese, A., 2022. What theories of value (could) underpin our circular futures? Ecol. Econ. 195, 107382. https://doi.org/10.1016/j.ecolecon.2022.107382.
Mahanty, S., Boons, F., Handl, J., Batista-Navarro, R., 2021. An investigation of academic perspectives on the 'circular economy' using text mining and a Delphi study. J. Clean. Prod. 319, 128574. https://doi.org/10.1016/j.jclepro.2021.128574.
Martinez-Alier, J., Munda, G., O'Neill, J., 1998. Weak comparability of values as a foundation for ecological economics. Ecol. Econ. 26 (3), 277–286. https://doi.org/10.1016/S0921-8009(97)00120-1.
Merry, 2016. The Seductions of Quantification: Measuring Human Rights, Gender Violence, and Sex Trafficking. Chicago University Press, Chicago.
Mintz, A., Wayne, C., 2016. The Polythink syndrome and elite group decision-making. Polit. Psychol. 37 (S1), 3–21. https://doi.org/10.1111/pops.12319.
Mitchell, G., 1996. Problems and fundamentals of sustainable development indicators. Sustain. Dev. 4 (1), 1–11. https://doi.org/10.1002/(SICI)1099-1719(199603)4:1<1::AID-SD24>3.0.CO;2-N.
Moraga, G., Huysveld, S., Mathieux, F., Blengini, G.A., Alaerts, L., Van Acker, K., de Meester, S., Dewulf, J., 2019. Circular economy indicators: what do they measure? Resour. Conserv. Recycl. 146, 452–461. https://doi.org/10.1016/j.resconrec.2019.03.045.
Oliveira, M., Miguel, M., van Langen, S.K., Ncube, A., Zucaro, A., Fiorentino, G., Passaro, R., Santagata, R., Coleman, N., Lowe, B.H., Ulgiati, S., Genovese, A., 2021. Circular economy and the transition to a sustainable society: integrated assessment methods for a new paradigm. Circ. Econ. Sustain. 1 (1), 99–113. https://doi.org/10.1007/s43615-021-00019-y.
O'Neill, D.W., 2012. Measuring progress in the degrowth transition to a steady state economy. Ecol. Econ. 84, 221–231. https://doi.org/10.1016/j.ecolecon.2011.05.020.
Opp, S.M., 2017. The forgotten pillar: a definition for the measurement of social sustainability in American cities. Local Environ. 22 (3), 286–305. https://doi.org/10.1080/13549839.2016.1195800.
Ozdemir, M.S., 2005. Validity and inconsistency in the analytic hierarchy process. Appl. Math. Comput. 161 (3), 707–720. https://doi.org/10.1016/j.amc.2003.12.099.
Pansera, M., Genovese, A., Ripa, M., 2021. Politicising circular economy: what can we learn from responsible innovation? J. Responsible Innov. 8 (3), 471–477. https://doi.org/10.1080/23299460.2021.1923315.
Pansera, M., Barca, S., Martinez Alvarez, B., Leonardi, E., D'alisa, G., Meira, T., Guillibert, P., 2024. Toward a just circular economy: conceptualizing environmental labor and gender justice in circularity studies. Sustain. Sci. Pract. Policy 20 (1), 2338592.
Pearce, D.W., Turner, R.K., 1990. Economics of Natural Resources and the Environment. Johns Hopkins University Press.
Pearce, W., Grundmann, R., Hulme, M., Raman, S., Kershaw, E.H., Tsouvalis, J., 2017. Beyond counting climate consensus. Environ. Commun. 11 (6), 723–730.
Polletta, F., Hoban, K., 2016. Why consensus? J. Soc. Polit. Psychol. 4 (1). https://doi.org/10.5964/jspp.v4i1.524. Article 1.
Prieto-Sandoval, V., Ormazabal, M., Jaca, C., Viles, E., 2018. Key elements in assessing circular economy implementation in small and medium-sized enterprises. Bus. Strateg. Environ. 27 (8), 1525–1534. https://doi.org/10.1002/bse.2210.
Purvis, B., Genovese, A., 2023. Better or different? A reflection on the suitability of indicator methods for a just transition to a circular economy. Ecol. Econ. 212, 107938. https://doi.org/10.1016/j.ecolecon.2023.107938.
Purvis, B., Mao, Y., Robinson, D., 2019. Three pillars of sustainability: in search of conceptual origins. Sustain. Sci. 14 (3), 681–695.
Purvis, B., Calzolari, T., Fevereiro, J.B., Genovese, A., MahmoumGonbadi, A., 2023a. D4.1 Exploring a Just Transition to a Circular Economy: Decision Support Tool Specification. JUST2CE.
Purvis, B., Celebi, D., Pansera, M., 2023b. A framework for a responsible circular economy. J. Clean. Prod. 400, 136679. https://doi.org/10.1016/j.jclepro.2023.136679.
Ramsey, J.L., 2015. On not defining sustainability. J. Agric. Environ. Ethics 28 (6), 1075–1087. https://doi.org/10.1007/s10806-015-9578-3.
Raworth, K., 2017. Doughnut economics: seven ways to think like a 21st century economist.Penguin. Random House, London.
Render, B., Stair, R., Hanna, M., Hale, T., 2017. Quantitative Analysis for Management, *Global Edition* (13th edition). Pearson.
Rettschlag, J., 2024. Transdisciplinary research for democracy: How much conflict does transdisciplinary research need?. In: EASST-4S 2024 Conference. Making and Doing Transformations, Amsterdam.
Rich, W., 2003. Historical high-stakes policies relating to unintended consequences of high-stakes testing. Educ. Meas. Issues Pract. 22 (1), 33–35. https://doi.org/10.1111/j.1745-3992.2003.tb00114.x.
Roos Lindgreen, E., Salomone, R., Reyes, T., 2020. A critical review of academic approaches, methods and tools to assess circular economy at the micro level. Sustainability 12 (12), 4973. https://doi.org/10.3390/su12124973.
Saaty, T.L., 1977. A Scaling Method for Priorities in Hierarchical Structures. J. Math. Psychol. 15, 234–281. https://doi.org/10.1016/0022-2496(77)90033-5.
Saaty, T.L., 1980. The Analytic Hierarchy Process. McGraw-Hill, New York.
Saaty, T.L., 1986. Axiomatic Foundation of the Analytic Hierarchy Process. Manage. Sci. 32 (7), 841–855.
Saaty, T.L., 2008. Decision making with the analytic hierarchy process. Int. J. Serv. Sci. 1 (1), 83. https://doi.org/10.1504/IJSSCI.2008.017590.
Savini, F., 2023. Futures of the social metabolism: degrowth, circular economy and the value of waste. Futures 150, 103180. https://doi.org/10.1016/j.futures.2023.103180.
Schultz, F.C., Pies, I., 2023. The circular economy growth machine: a critical perspective on "post-growth" and "pro-growth" circularity approaches. J. Ind. Ecol. https://doi.org/10.1111/jiec.13450 n/a(n/a).
Sébastien, L., Bauler, T., 2013. Use and influence of composite indicators for sustainable development at the EU-level. Ecol. Indic. 35, 3–12. https://doi.org/10.1016/j.ecolind.2013.04.014.
Souza Piao, R., de Vincenzi, T.B., da Silva, A.L.F., de Oliveira, M.C.C., Vazquez-Brust, D., Monteiro Carvalho, M., 2023. How is the circular economy embracing social inclusion? J. Clean. Prod. 411, 137340. https://doi.org/10.1016/j.jclepro.2023.137340.
Steinert, M., 2009. A dissensus based online Delphi approach: an explorative research tool. Technol. Forecast. Soc. Chang. 76 (3), 291–300. https://doi.org/10.1016/j.techfore.2008.10.006.

Stengers, I., 2005. Introductory notes on an ecology of practices. Cult. Stud. Rev. 11 (1). https://doi.org/10.5130/csr.v11i1.3459. Article 1.
Todaro, D.L., Torelli, R., 2024. From greenwashing to ESG-washing: a focus on the circular economy field. Corp. Soc. Responsib. Environ. Manag. https://doi.org/10.1002/csr.2786.
Turcu, C., 2013. Re-thinking sustainability indicators: local perspectives of urban sustainability. J. Environ. Plan. Manag. 56 (5), 695–719. https://doi.org/10.1080/09640568.2012.698984.
United Nations, 2017. Resolution Adopted by the General Assembly on Work of the Statistical Commission Pertaining to the 2030 Agenda for Sustainable Development (A/RES/71/313). https://documents-dds-ny.un.org/doc/UNDOC/GEN/N17/207/63/PDF/N1720763.pdf?OpenElement.
van den Hove, S., 2006. Between consensus and compromise: acknowledging the negotiation dimension in participatory approaches. Land Use Policy 23 (1), 10–17. https://doi.org/10.1016/j.landusepol.2004.09.001.
Van Schoubroeck, S., Vermeyen, V., Alaerts, L., Van Acker, K., Van Passel, S., 2022. How to monitor the progress towards a circular food economy: A Delphi study. Sustainable Production and Consumption 32, 457–467. https://doi.org/10.1016/j.spc.2022.05.006.
Vinante, C., Sacco, P., Orzes, G., Borgianni, Y., 2021. Circular economy metrics: literature review and company-level classification framework. J. Clean. Prod. 288, 125090. https://doi.org/10.1016/j.jclepro.2020.125090.
Walzberg, J., Lonca, G., Hanes, R.J., Eberle, A.L., Carpenter, A., Heath, G.A., 2021. Do we need a new sustainability assessment method for the circular economy? A critical literature review. Front. Sustain. 1. https://doi.org/10.3389/frsus.2020.620047.
Yousuf, M.I., 2007. Using experts` opinions through Delphi technique. Pract. Assess. Res. Eval. 12 (1). https://doi.org/10.7275/rrph-t210.