Visualizing Internal Sustainability Efforts in Big Companies

Chiara Ceccarini, University of Bologna, 40126, Bologna, Italy
Edyta Paulina Bogucka, Technical University of Munich, 80333, München, Germany
Indira Sen, GESIS Leibniz Institute for Social Sciences, 68159, Cologne, Germany
Marios Constantinides, Nokia Bell Labs, CB3 0FA, Cambridge, U.K.
Catia Prandi, University of Bologna, 40126, Bologna, Italy
Daniele Quercia, Nokia Bell Labs, CB3 0FA, Cambridge, U.K.

Internal sustainability efforts (ISE) refer to a wide range of internal corporate policies focused on employees. They promote, for example, work–life balance, gender equality, and a harassment-free working environment. At times, however, companies fail to keep their promises by not publicizing truthful reports on these practices, or by overlooking employees' voices on how these practices are implemented. To partly fix that, we developed a deep-learning framework that scored four fifths of the S&P 500 companies in terms of six ISEs, and a web-based system that engages users in a learning and reflection process about these ISEs. We evaluated the system in two crowdsourced studies with 421 participants, and compared our treemap visualization with a baseline textual representation. We found that our interactive treemap increased by up to 7% our participants' opinion change about ISEs, demonstrating its potential in machine-learning driven visualizations.

Internal sustainability efforts (ISEs) describe a broad range of corporate policies focused on employees, including, for example, work–life balance, gender equality and diversity, and a harassment-free working environment. These ISEs can not only decrease staff turnover but also enhance a company's competitiveness. It comes as no surprise that companies, at times, obfuscate information about how ISEs are actually implemented in their public reports,[a] contributing to a gap between what companies publicize and what they actually do. Also, a study showed that investor reports and annual corporate reports (the gold standard for business assessment) are more of a corporate PR exercise than objective assessments,[b] especially for emerging concepts, such as sustainability. Therefore, as the same study also argued, accountability and verification of corporate claims are very much needed.

To partly close that gap, we developed a DL-driven visualization[c] for surfacing ISEs in big companies and engaging the general public in a debate about them. In so doing, we made three sets of contributions:

1) We collected public employee reviews from a company reviewing site, and, using a DL natural language processing tool, we scored four fifths of the S&P 500 companies in terms of their ISEs.[1]
2) Using these scores, we developed a web-based visualization tool for raising ISEs' awareness.

[a][Online]. Available: https://www.forbes.com/sites/forbesnonprofitcouncil/2021/03/23/businesses-should-be-held-accountable-for-their-esg-claims/

[b][Online]. Available: https://hbr.org/2019/06/business-as-usual-will-not-save-the-planet
[c][Online]. Available: http://social-dynamics.net/sustainability/

0272-1716 © 2022 IEEE
Digital Object Identifier 10.1109/MCG.2022.3163063
Date of publication 29 March 2022; date of current version 8 June 2022.

3) We evaluated the tool in two crowdsourced studies with 421 participants and compared our treemap visualization with a baseline textual representation.

We found that treemap increased by up to 7% our participants' opinion change about ISEs, demonstrating its potential as an alternative representation in ML-driven visualizations.

RELATED WORK

ML4VIS is a new branch of research that uses ML techniques to develop, design, and evaluate visualizations.[2] Cunningham-Nelson *et al.*[3] used a latent Dirichlet allocation algorithm to analyze free-text students' comments obtained from satisfaction surveys, which, in turn, powered a visualization that allowed educators to understand students' concerns on teaching. Corporate sustainability efforts have gained traction within academic circles. For example, a theoretical framework divides companies into ones that report on their own efforts only to manage their brands (symbolic) and to those that genuinely report on actual changes (substantial).[4] To add transparency in this area, Sneha *et al.*[5] developed interactive visualizations for comparing companies' sustainability efforts. Similarly, the OECD's Life Index[d] provides a web-based visualization for comparing sustainability efforts (e.g., health and environment), but it does so at a country level rather than company level.

To summarize, previous works focused on public reports, often overlooking employees' opinions on the practical implementation of ISEs. In addition, ML-driven visualizations often use static, default types of graphs, and exploration techniques (e.g., bar charts).[2] The unmet design challenge is, therefore, how to provide users with dynamic, ML-driven visualizations using a new combination of data engagement mechanisms.

DEEP-LEARNING FRAMEWORK

We collected a dataset of 358,527 reviews published on a popular company reviewing site. On that site, former or current employees share their experiences of their companies as free-form textual reviews, in addition to ratings about different aspects, such as management and culture. We selected 104 U.S.-based companies with at least 1000 reviews between 2008 and 2020, and with a (physical) presence in more than 10 U.S. states. A total of 81% of these companies were in the S&P 500.

[d][Online]. Available: http://do.minik.us/blog/oecd_bli

The reviewing site ensures quality reviews by performing both automatic and manual content moderation (e.g., registered users and those who wrote at least one review have full content access, and a maximum of one review per employee per year is allowed[e]). However, while data could be biased, it is systematically so across companies, making companies and their scores comparable. Therefore, several studies have explored corporate culture at scale using data from the site.[6,7] The site explicitly divides reviews into pros (positive) and cons (negative). As sustainability has a positive valence, we opted for using pros. By manually inspecting a random sample of 500 pros and cons, we found that, on average, 89% of pros mentioned ISEs (and did so with a positive valence) compared to 63% of cons (with mixed-valence). To then operationalize ISEs, we adopted a three-step mixed-method approach:

Step 1—Preselection: Three annotators assessed the UN's Sustainable Development Goals definitions.[f,1] Given their broad scope, not all 17 goals might be relevant to internal corporate practices. The annotators unanimously decided to discard four: "life below water," "life on land," "sustainable cities," and "partnerships for goals." The former three focus on water bodies, land, and cities, which are unlikely to appear in employees' reviews outside of highly specialized companies. "Partnership on goals" was explicitly designed to foster countries' collaboration, and, as we focused on U.S.-based companies, that goal was also excluded.

Step 2—Unsupervised discovery: To find the similarity between reviews and goal definitions, we employed the state-of-the-art DL method called SBERT.[8] Using SBERT, we scored each employee's review against the 13 previously retained goals.[1] For each goal, the three annotators assessed the relevance of the top five most relevant reviews identified by the framework. On average, they reached an agreement as high as Fleiss K equal to 0.83. To take a conservative approach and ascertain that goals were less accurately identified by the framework, we identified which goals had less than four (of the top five) reviews to be marked as relevant by the majority of the annotators. As a result, we dropped five goals, which had to do with environmental sustainability (e.g., "clean water" and "climate change") rather than corporate internal sustainability, leaving us with eight goals.

[e][Online]. Available: https://www.glassdoor.com/research/app/uploads/sites/2/2017/10/Glassdoor_GiveToGet_Oct2017-1.pdf

[f][Online]. Available: https://sdgs.un.org/goals

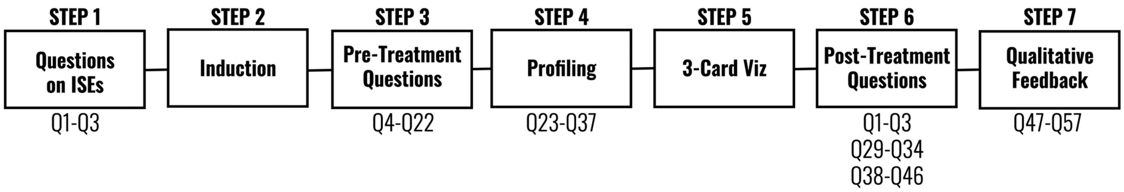

FIGURE 1. Our user study procedure consists of seven steps.

Step 3—Consolidation: The refined goals were assessed for semantic relatedness, and strongly related goals were merged together, ultimately leaving us with six goals. Sustainability goals are not mutually exclusive and a certain degree of overlap might be expected (e.g., work–life balance facilitates both health and gender equality). However, there might be cases where two goals are so strongly related to one another that they cannot be discerned. To systematically tackle this issue, we plotted the content overlap O for each pair of goals by computing the proportion of sentences that two goals j and k have in common. The only overlap higher than 0.5 occurred for the pair "food (no hunger)" versus "health." These have indeed strong conceptual relatedness, thus subsuming "no hunger" under "health." Note that two other goals' pairs exhibited semantic relatedness close to 0.5: these were "supportive environment" versus "supporting infrastructure," and "diversity" versus "gender equality." To decide whether to combine them, the annotators assessed the top five reviews for each goal, and found that "supportive environment" and "supporting infrastructure" covered related yet different concerns; however, they discovered that the "diversity" goal (reducing inequality) was mostly expressed through mentions of "gender discrimination." Consequently, we merged these two goals together.

Scoring companies: After identifying these six goals, the framework computed each company's score $s(c, i)$ of the ith ISE for company c as the fraction of c's reviews that mentioned i

$$s(c,i) = \frac{\sum_{r \in R(c)} \mathrm{sim}_t(v_r, v_i)}{|R(c)|} \tag{1}$$

where $R(c)$ is the set of c's reviews, v_i is the SBERT vector of ISE i [see Figure 2(a)—definitions of the six ISEs], and $\mathrm{sim}_t(v_r, v_i)$ is the *thresholded* SBERT similarity score[8] between the SBERT vector of review r and the SBERT vector of ISE i. More precisely, $\mathrm{sim}_t(v_r, v_i)$ is defined as

$$\mathrm{sim}_t(v_r, v_i) = \begin{cases} \mathrm{sim}(v_r, v_i), & \text{if } \mathrm{sim}(v_r, v_i) > 0.31 \\ & \text{AND} \\ & \mathrm{sim}(v_r, v_i) > 95\%(i) \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where $\mathrm{sim}(v_r, v_i)$ is the cosine similarity between v_r and v_i. We chose the threshold of 0.31 by computing the mean SBERT similarity for the goals. We then paired the fixed generalized threshold of 0.31 with an ISE dimension-specific threshold. Based on our experiments, we chose a 95%(i) threshold value, which is the 95% percentile of the ISE's distribution.

To support a seamless visualization experience, for each company c, we computed its company vector (v_c) *offline* (i.e., the computation was not repeated for every user but was performed only once)

$$v_c = [s'(c,1), s'(c,2), \ldots, s'(c,6), \mathrm{sg}(c)] \tag{3}$$

where $s'(c,i)$ is the ith ISE scaled score for c ($s'(c,i)$ is the value of $s(c,i)$ scaled between 0 and 100), and $\mathrm{sg}(c)$ is the c's internal sustainability gap computed as per (4). As the sum of the scaled scores of a company ($\sum_{i=1}^{6} s'(c,i)$) may not reach the maximum value of 600 (each ISE can take a maximum of 100), we wanted to reflect that information in the company card and, as such, introduced the concept of *internal sustainability gap* (sg) for company c

$$\mathrm{sg}(c) \quad = \max - \textstyle\sum_{i=1}^{6} s'(c,i) \tag{4}$$

where 6 is the number of ISEs, and max is 600, which is the maximum score for the sum of the ISEs.

SCORING THE USER

In our visualization tool, we had two types of cards: company card and user card. The company card is the graphic representation of the company's vector previously collated as per expression (3). The user card is the graphic representation of the user's vector. During our user study (see Figure 1), our participants go through seven steps, some of which consist of answering questions. To collate the user's vector, we relied on the user's answers to the two sets of questions (Q23–Q34), asked before exposing the user to the visualization [see Figure 2(b)]. The first set, Q23–Q28, consists of six questions (one for each ISE) asking the user whether (s)he values a given ISE with a statement: "I see myself as someone who values workplace *[X]*." The second set, Q29–Q34, consists of

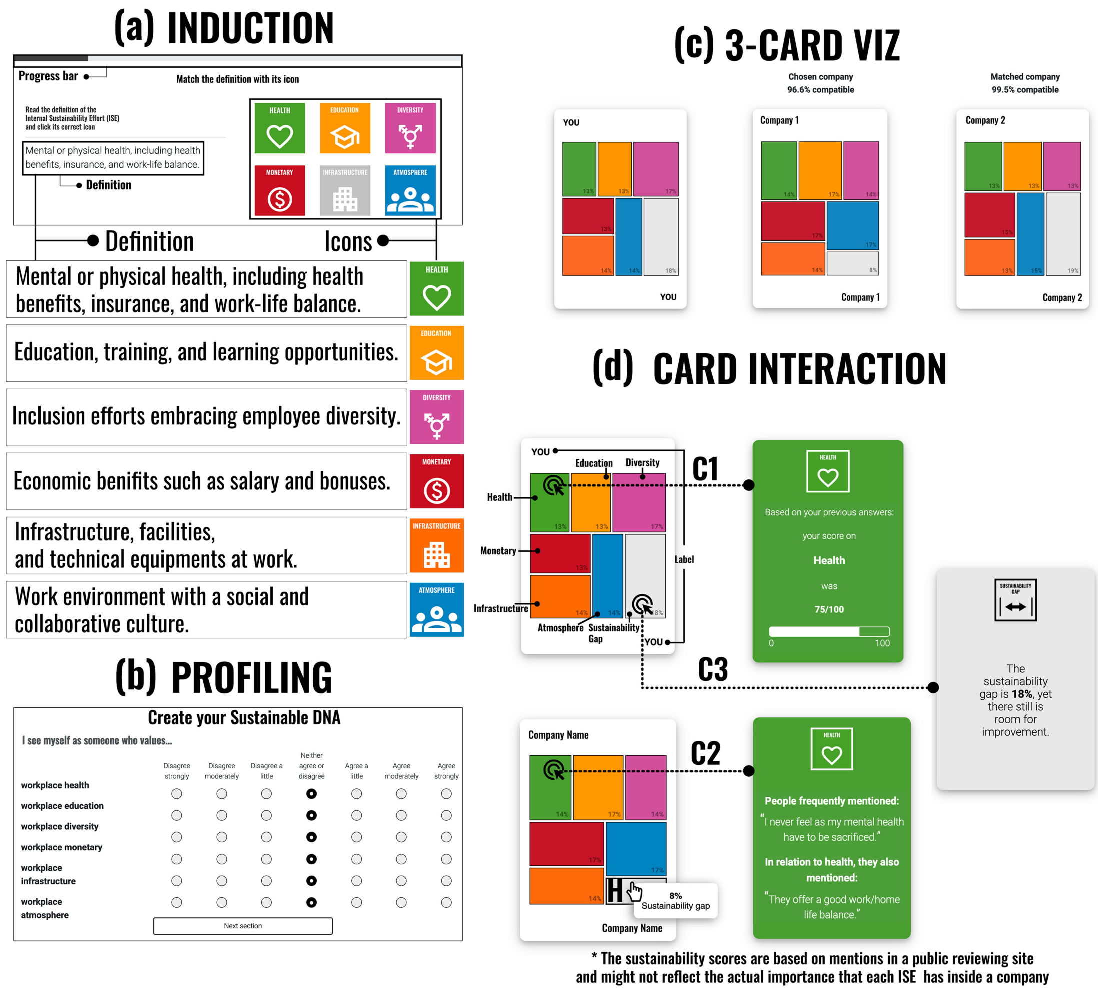

FIGURE 2. (a) User interface for step 2 (induction) consisting of: a progress bar at the top, a definition of one of the six ISEs to be matched on the left, and the six icons to be matched (if the match done by the user is correct, the corresponding icon is disabled and turned gray, as the infrastructure icon is in this picture). (b) User interface for step 4 (profiling) consisting of a statement/question the user needs to answer, typically on a 1–7 Likert scale. All the answers at this step were used to create the user vector v_u. (c) The three-card viz (step 5) consists of three cards (user card, chosen company card, and recommended company card). (d) Interactions with a card included clicking on a colored rectangle to flip the clicked user (C1, C3) or company (C2) card, and hovering on a rectangle to display the tooltip showing the name of ISE and its relative percentage score (H).

six questions asking the user whether (s)he chose an employer based on a given ISE with a statement: "In the past, I chose one employer because it valued workplace *[X]*." To then quantify the extent to which u cared about the ith ISE, we used the two sets of questions. For example, to quantify the extent to which u cared about internal efforts related to *health*, we took u's score for Q23 (his/her disposition to value health) plus u's score for Q29 [whether (s)he decided to work for an employer because it valued health]. In other words, for internal efforts related to *health* ($i = 1$), u's score $s(u, 1)$ is: $\text{score}(u, Q23_t) + \text{score}(u, Q29_t)$, where t is the time before exposing u to the visualization. This procedure was then repeated for all the six ISEs by computing each user u's score $s(u, i)$ for the ith ISE as

$$s(u, i) = \text{score}(u, Q(k+i)_t) + \text{score}(u, Q(k+6+i)_t) \quad (5)$$

where $i \in [1, 6]$, k is 22, t is the time before exposing u to the visualization, $\text{score}(u, Q(k+i)_t)$ is user u's score to question $Q(k+i)$ at time t, which goes through the set Q23–Q28, while $\text{score}(u, Q(k+6+i)_t$ goes through the set Q29–Q34.

Hence, we computed six $s(u,i)$, scored between 2 and 14. Finally, to compare users with companies, we linearly mapped these scores from the range of [2,14] to [0,100], obtaining $s'(u,i)$, and collated these values in u's vector

$$v_u = [s'(u,1), s'(u,2), \ldots, s'(u,6), \mathrm{sg}(u)] \tag{6}$$

where $s'(u,i)$ is the ith ISE scaled score for u, and $\mathrm{sg}(u)$ is the u's internal sustainability gap computed in a way similar to (4) as $\mathrm{sg}(u) = 600 - \sum_{i=1}^{6} s'(u,i)$, where 600 is the maximum score of the six ISEs' sum.

Recommended company: After obtaining the user vector v_u, we knew u's preferences concerning ISEs. We then matched these preferences with the extent to which a company matched the preferences, and did so by computing v_u's cosine similarity values with all companies v_c's, and, as a result, found the company most similar to u, which we call u's recommended company.

THREE-CARD VIZ: AWARENESS AND REFLECTION

Our visualization consisted of a "card game" with three cards: one visualizing the user's vector, and the other two visualizing the vectors of two companies: one company was the *recommended company*, while the other was provided by the user, which we call *chosen company* [see Figure 2(c)-C2]. We chose only one card for the *recommended company* due to the limited screen size. If more cards were to be displayed, alternative interactions techniques would have been used (e.g., scrolling through the card deck,[9,10] or stacking cards into multiple groups). However, such an implementation would have increased the complexity in comparing cards. Therefore, a card of any of those three types was designed with two main characteristics concerning its display and its interactions.

Card display: We relied on the metaphor of DNA. DNA is a structure made of molecules that encode individuals' biological information and, therefore, it can uniquely identify them. Similarly, sustainable behavior can be seen as a structure where the molecules are represented by the six ISEs, and the unique combination of the six ISEs' scores can uniquely identify the user's predispositions or a company's internal initiatives. Despite being an oversimplification, the DNA metaphor likely reflects the popular understanding of dynamic and adaptable patterns. The card was designed as a treemap wherein six rectangles showed the six ISEs and one gray rectangle showed the internal sustainability gap [see Figure 2(c)]. We chose treemaps because they: 1) allow for visualizing fractional values that must be interpreted in a comparative fashion rather than at face value, and 2) have considerable engagement qualities over alternatives (e.g., lists) for the task at hand.[11] Moreover, SDGs are typically visualized in grids,[g] representing each goal as a square. We did the same but with rectangles of different dimensions obtained from the "*squarify*" treemap algorithm. This algorithm—as many other treemap algorithms—creates rectangles approximated to squares that are easier to compare and select regardless of screen size.[12] The total treemap's dimension area depended on the screen size of the device. Since the rectangles inside the card graphically represented a user/company's vector, we computed each rectangle's dimension $d(e, \mathrm{area})$ based on the vector being displayed in the card (either v_u or v_c)

$$d(e, \mathrm{area}) = \frac{v_x(e) * \mathrm{area}}{\sum_{j=1}^{7} v_x(j)} \tag{7}$$

where x is either company c or user u, e is a counter for user/company vector v_x, $v_x(e)$ is the eth values of v_x [e.g., $v_x(1)$ is about health, $v_x(7)$ is the internal sustainability gap sg], 7 is the length of v_x (6 ISEs+sg), and area is the total area of the treemap.

Card interactions: Interactions with card-based visualizations are generally inspired by physical card use,[13] being based on hovering, swiping, stacking, and shuffling.[9,10] In our design, we opted for the minimal set of interactions that balance:

1) fitting the limited screen space;
2) showing extracted patterns; and
3) providing example reviews.

As a result, we opted for two interactions- *card flipping* and *hovering on a rectangle*. On the front of both the user's card and the company's card, the user can view his/her vector and the companies' vectors in the form of seven-colored rectangles. Flipping allows for interacting with both sides of a card—a modality the general public is likely familiar with (e.g., from memory matching games). The card flips when the user clicks on a rectangle corresponding to a specific ISE: if the card is a user card, then the back shows the user's score for that ISE [see Figure 2(d)-C1 and C3]; if it is a recommended/chosen company's card, the back instead shows two reviews of that company related to that ISE [see Figure 2(d)-C2]. These two reviews come from a clustering process. For each ISE,

[g][Online]. Available: https://sdgs.un.org/goals

we indeed clustered reviews based on semantic similarity and took one review from the most frequent cluster, and another review from the second most frequent cluster. The back of the card shows these two reviews under ("*Employees frequently mentioned*") and ("*they also mentioned*"), respectively. In addition to flipping, a user can interact with the front of a card by "hovering on a rectangle" (on an ISE) and being shown a tooltip with the corresponding ISE's name and score [see Figure 2(d)-H].

USER STUDY

We evaluated our tool in an online crowdsourcing study on Amazon MTurk (AMT) wherein our participants followed a 7-step study procedure (see Figure 1), with a completion time of 15 min and compensation of \$0.50. To ensure quality responses, we applied quality controls in the form of two attention questions in steps 3 and 6. To then ensure a comprehensive assessment, the questions focused on whether our visualization contributed to user learning, user opinion change, and increasing transparency between a company and the public.

Step 1—Questions concerning ISEs (Q1–Q3): Before being exposed to the visualization, the user answered three questions (see Figure 3-Q1–Q3), reflecting what (s)he knew about ISEs. We also asked the very same questions after exposing the user to the visualization (step 6). The differences in the before/after answers then reflected whether the user learned anything new about ISEs as a consequence of interacting with our visualization.

Step 2—Induction: After answering those questions, in the form of an induction game, the user had to match a definition of an unnamed ISE with the correct ISE icon [see Figure 2(a)]. For example, the definition of health ISE should be matched with the heart icon. Every time the answer was correct, the right icon got disabled, and the user was able to proceed with the next match. In the case of a wrong answer, the user was encouraged to try again, learning in a trial-and-error fashion. Note that we designed the six icons to be easily matched with an ISE's meaning, and be distinguishable from each other.

Step 3—Pre-treatment questions (Q4–Q22): The user completed the 10 item personality measure questionnaire[14] (Q4–Q13), which measures the Big-Five personality dimensions through 10 sentences, rated on a seven-point Likert scale (1: strongly disagree; 7: strongly agree). The user-provided personal information in Q14–Q16 and Q18–Q22 (i.e., age, gender, education, country of origin, and residence). In between these questions, the user answered a first attention question, which took the form of "Without speculating on possible advances in science, how likely are you to live to 500 years old?," and used as a quality control check (Q17).

Step 4—Profiling (Q23–Q37): The user then answered 15 profiling questions [see Figure 2(b)]. The first 12 questions were used to build the user's vector v_u (Q23–Q34) and asked whether the user valued the six ISEs (Q23–Q28) and whether (s)he chose to work for an employer because it valued the ISEs (Q29–Q34). The three other questions (Q35–Q37) asked for:

1) the industry sector in which the user worked;
2) the industry sector the user would like to explore; and
3) the name of the company the user would like to explore (the so-called chosen company).

Step 5—Three-card viz: The user was then shown the three cards [see Figure 2(c)]: one reflecting his/her ISE vector, and the other two reflecting the vectors of the chosen and recommended companies. This allowed the user to compare his/her own card with those of the two companies.

Step 6—Post-treatment questions (Q38–Q46 + repeated Q1–Q3 and Q29–Q34): The user answered 18 questions (nine new plus nine repeated) to evaluate the visualization's contribution to user *learning* (Q1–Q3 were asked again to test whether the user learned additional information about ISEs after interacting with the visualization); user *opinion change* (Q29–Q34 were asked again to test whether the user changed his/her views on whether (s)he would select an employer based on its commitments to ISEs); and increasing *transparency* between companies and the general public (Q38–Q39 asked whether the user thought that, based on what (s)he learned through our visualization, the two companies effectively communicated their internal efforts). Finally, we asked six questions to test our algorithmic and interaction choices. We asked whether the recommended company could be a good match (Q40), whether the user preferred a specific type of interaction, whether (s)he found the percentage displayed on the card helpful (Q41–Q42), whether the visualization made him/her more aware of what ISEs entailed (Q43) compared to what (s)he knew before (Q45), and whether the visualization helped him/her reflect on ISEs (Q46). In between these groups of questions, the user answered a second attention question (Q44): "what best defines economic benefits," among options describing mental health, inclusion, infrastructure, and salaries and bonuses.

ID	Question	Step
Q1	With which statement do you agree the most? Internal sustainability initiatives at the workplace translate into: 1) multiple ecological and environmental benefits for employees 2) multiple monetary benefits for employees 3) multiple benefits for employees beyond ecological and monetary benefits	STEP 1, STEP 6
Q2	Is internal corporate sustainability more than the natural environment?	STEP 1, STEP 6
Q3	Did people mention ISEs in company's reviews?	STEP 1, STEP 6

ID	Question	Step
Q4 – Q13	Ten Item Personality Measure (TIPI)	STEP 3
Q14	What is your first name or nickname?	STEP 3
Q15	To which gender identity do you most identify?	STEP 3
Q16	What is your age?	STEP 3
Q17	Without speculating on possible advances in science, how likely are you to live to 500 years old?	STEP 3
Q18	What is the highest degree or level of school you have completed?	STEP 3
Q19	In which country do you currently live?	STEP 3
Q20	In which USA state do you currently live?	STEP 3
Q21	In which country were you born?	STEP 3
Q22	In which USA state were you born?	STEP 3

ID	Question		Step
Q23	I see myself as someone who values workplace...	health	STEP 4
Q24		education	STEP 4
Q25		diversity	STEP 4
Q26		monetary	STEP 4
Q27		infrastructure	STEP 4
Q28		atmosphere	STEP 4
Q29	In the past (future), I chose (will choose) at least one employer because it valued workplace...	health	STEP 4, STEP 6
Q30		education	STEP 4, STEP 6
Q31		diversity	STEP 4, STEP 6
Q32		monetary	STEP 4, STEP 6
Q33		infrastructure	STEP 4, STEP 6
Q34		atmosphere	STEP 4, STEP 6
Q35	What is your area of work?		STEP 4
Q36	What is your area of interest?		STEP 4
Q37	Choose a company of your interest		STEP 4

ID	Question	Step
Q38	Did the visualization help you understand whether your chosen company cared about ISEs?	STEP 6
Q39	Did your chosen company care about ISEs as you expected?	STEP 6
Q40	Do you feel your recommended company is a good match?	STEP 6
Q41	Which action did you find the most engaging?	STEP 6
Q42	Did the percentage displayed next to the colored box help you explore the data?	STEP 6
Q43	Did your card make you more aware of your values?	STEP 6
Q44	Based on what you've learned, the best definition for monetary benefits at the workplace is: 1) Mental or physical health, including health benefits, insurance, and work/life balance. 2) Infrastructure, facilities, and technical equipments at work. 3) Economic benefits such as salary and bonuses. 4) Inclusion efforts embracing employee diversity.	STEP 6
Q45	Did you know something about ISEs before?	STEP 6
Q46	Did the visualization help you reflect on ISEs?	STEP 6

ID	Question	Step
Q47	Why the visualization help (or not) you understand whether your chosen company cared about ISEs?	STEP 7
Q48	Did the comparison between the two companies' cards help you better quantify their ISEs?	STEP 7
Q49	Why do you feel your recommended company is (or isn't) a good match?	STEP 7
Q50	Why did you select action [X] as the most engaging?	STEP 7
Q51	Why the percentage displayed next to the colored box help (or not) you explore the data?	STEP 7
Q52	Why your card make you more aware (or not) of your values?	STEP 7
Q53	Which part of the visualization was the most helpful? Why?	STEP 7
Q54	Which part of the visualization was the least helpful? Why?	STEP 7
Q55	What was the most surprising thing you learned about ISEs at the workplace?	STEP 7
Q56	Is there any way you can make other people aware of ISEs?	STEP 7
Q57	Please share any comment you might have on the project.	STEP 7

FIGURE 3. Set of questions/statements presented to the user during the user study sketched in Figure 1 (this figure is best seen in color). The evaluation questions tested our visualization's contribution to user learning (questions in orange asked pre- and post-treatment); user opinion change (questions in yellow asked pre- and post-treatment); and increasing transparency between companies and the general public (post-treatment questions in light blue).

Step 7—Qualitative feedback (Q47–Q57): Finally, the user-provided feedback about the visualization through 11 open-ended questions.

Metrics

To analyze the Likert-scale questions, in line with previous work,[15] we divided our participants into three groups based on their answers' polarity: negatively polarized (NP) participants who answered "disagree strongly" (-3), or "disagree moderately" (-2); neutrally/weakly polarized (NWP) participants who answered "disagree a little" (-1), "neither agree or disagree" (0), or "agree a little" (+1); or positively polarized (PP) participants who answered "agree moderately" (+2), or "agree strongly" (+3). For Likert-scale questions and those asked twice (Q2–Q3 and Q29–Q34), before and after the visualization, we evaluated our participants' opinion change as the percentage growth rate between each of the answers' group (NP, NWP, and PP). In particular, starting from the negative polarized, we calculated the delta between the percentage of NP after experiencing the visualization ($NP_{(t+1)}$) and before experiencing it (NP_t): $\Delta NP = NP_{(t+1)} - NP_t$. The same procedure was repeated for the other two groups: $\Delta NWP = NWP_{(t+1)} - NWP_t$; and, $\Delta PP = PP_{(t+1)} - PP_t$.

We computed an *aggregated opinion change score* $o(u)$ to determine the extent to which our participants changed their opinions after interacting with the three-card viz. The questions that were repeated before/after the viz were nine: Q1 was a multiple choice question, while those in the set (Q2, Q3, Q29–Q34) were on a Likert scale. By aggregating the eight Likert-scale questions, we computed the opinion change score

$$o(u) = \sum_{k \in \{2,3,29-34\}} |\text{score}(u, Qk_{(t+1)} + \\ - \text{score}(u, Qk_t)| \quad (8)$$

where t is the time before exposing u to the visualization, $(t+1)$ is the time after being exposed, and $\text{score}(u, Qk_t)$ is the user u's score to question Qk at time t.

Quantitative Results

User sample (pre-treatment Q4–Q22 asking personal information): We had 244 participants. They scored, on average, on a scale from 1 to 7 (Q4–Q13): high in agreeableness ($\mu = 5.3$ and $\sigma = 1.7$), high in conscientiousness ($\mu = 5.7$ and $\sigma = 1.3$), high in emotional stability ($\mu = 5.0$ and $\sigma = 1.3$), high in openness ($\mu = 5.0$ and $\sigma = 1.5$), and low in extraversion ($\mu = 3.8$, $\sigma = 1.3$). The distributions of these traits were aligned with the normative personality values drawn from a large U.S. population sample.[16] A total of 133 participants were female (Q15), all aged between 18 and 75 years old, with a median age of 40 (Q16). They were well educated (66% held a B.Sc.), and were mostly U.S. citizens (97%), with only 3% being immigrants but residing in the U.S. (Q20–22). In between these questions, participants answered the first attention question (Q17), which led us to filter out the contributions of 29% of the initial participants.

User learning (pre- and post-treatment Q1–Q3): Initially, participants thought that sustainability efforts revolve around ecological and environmental benefits for employees (47%) along with monetary benefits (12%) ($Q1_t$). After interacting with the visualization, 58% of them agreed that sustainability efforts can go beyond ecological and monetary benefits ($Q1_{(t+1)}$). Most of them (73%) were aware that sustainability encompassed more than the natural environment ($Q2_t$); however, only 27% of them had any knowledge of how sustainability efforts could be introduced in the workplace. After interacting with the visualization, 77% of participants recognized that sustainability has many facets ($Q2_{(t+1)}$). Finally, before being exposed to the visualization, 12% of participants knew that employees could mention sustainability in companies' reviews ($Q3_t$); after the visualization, that percentage peaked at 38% ($Q3_{(t+1)}$).

Initial user views on ISEs (pre-treatment Q23–Q28): Before interacting with the visualization ($Q29_t$–$Q34_t$), the ISE most valued for the employer choice was the monetary one (71% PP), followed by atmosphere (66% PP), health (52% PP), education (48% PP), infrastructure (44% PP), and diversity (43% PP) (see Figure 4). To ensure that our participants' answers after interacting with the visualization were not confounded by any previously held opinions, we plotted the distribution of the participants' employer choices before interacting with the visualization. We observed, to a great extent, a normal distribution for the six ISEs (health - ; education - ; diversity - ; monetary - ; infrastructure - ; atmosphere -), suggesting a lack of systematic bias at population level.

User opinion change (pre- and post-treatment Q2–Q3 and Q29–Q34): After interacting with our visualization ($Q29_{(t+1)}$–$Q34_{(t+1)}$), the positive polarized grew in all of the six ISEs (PP in Figure 5), suggesting that the three-card viz persuaded our participants of the importance of all ISEs. Based on the participants who became PP, the opinion change was most remarkable for infrastructure and diversity, and least for monetary (which was already high in the first place). By then computing the Spearman's rank correlation between the Big-Five personality traits (derived from Q4–Q14) and $o(u)$, we found that opinion change did not correlate with any specific personality trait, suggesting that our participants changed their views mainly because of what they reevaluated about a specific ISE because of their interaction with the visualization rather than who they were (their personality traits). Exceptions to this rule were found in weak correlations of $o(u)$ with: agreeableness ($r = -0.11$ and $p < 0.1$), conscientiousness ($r = -0.17$ and $p < 0.01$), and emotional stability ($r = -0.14$ and $p < 0.05$). These results suggest that people who changed their views tended, only to a limited extent, to be less organized and goal-oriented, to put their interests above those of others, and to be less emotionally stable.

Dimensions			ID		Statement		NP	NWP	PP
Profiling user values	PRE-TREATMENT	STEP 4	Q23	S1	I see myself as someone who values workplace...	health	2%	21%	**77%**
			Q24			education	3%	22%	**74%**
			Q25			diversity	4%	30%	**65%**
			Q26			monetary	0%	19%	**81%**
			Q27			infrastructure	1%	31%	**68%**
			Q28			atmosphere	1%	18%	**81%**
Profiling user opinion change	PRE-TREATMENT	STEP 4	Q29	S2	In the past, I chose one employer because it valued workplace...	health	15%	33%	**52%**
			Q30			education	12%	39%	**48%**
			Q31			diversity	17%	40%	**43%**
			Q32			monetary	3%	26%	**71%**
			Q33			infrastructure	12%	**44%**	**44%**
			Q34			atmosphere	6%	26%	**66%**
	POST-TREATMENT	STEP 6	Q29	S3	In the future, I will choose one employer because it values workplace...	health	4%	33%	**63%**
			Q30			education	5%	33%	**62%**
			Q31			diversity	9%	33%	**58%**
			Q32			monetary	0%	25%	**74%**
			Q33			infrastructure	4%	34%	**62%**
			Q34			atmosphere	3%	18%	**79%**
Increasing transparency between company and public	POST-TREATMENT	STEP 6	Q38	S4	Did the visualization help you understand whether your chosen company cared about ISEs?		8%	43%	**49%**
			Q39	S5	Did your chosen company care about ISEs as you expected?		15%	**54%**	31%

FIGURE 4. Percentages of participants who were NP, NWP, and PP on: the importance of a given ISE (S1); having chosen a past employer based on it valuing a given ISE (S2); choosing a future employer based on it valuing a given ISE (S3); and the effectiveness of our visualization in increasing transparency between a company and the general public (S4–S5). For each question, the group with the highest percentage of participants is marked in bold.

To understand whether the differences in polarization values were due to our design choices or by the content itself (scores plus reviews), we conducted a second experiment to compare our treemap visualization to a plain text baseline. The baseline displayed the scores and the reviews in a textual list. We recruited a new set of 177 participants (95 females, median age of 40) from AMT, while ensuring the same participants' characteristics, study duration, and compensation. All participants were well educated (63% held a B.Sc.), and, mostly, U.S. citizens (97%), with only 3% being immigrants but residing in the United States.

By analyzing this new set of participants' opinion change of the six ISEs when interacting with the two visualizations, we found that comparatively speaking, the treemap outperformed the baseline (see Figure 5). The treemap increased the opinion of PP participants in four out of six ISEs compared to the baseline. In particular, we found a 7% increase for infrastructure and atmosphere, 5% for education, and 2% for diversity, whereas we registered no change for monetary benefits and for health. As we shall see from the qualitative analysis, this was explained by the fact that monetary and health are familiar concepts, whereas concepts such as infrastructure or atmosphere were less relatable.

Increasing transparency between a company and the public (post-treatment Q38–Q39): Almost half of the participants (49%) declared that our visualization helped them understand whether their chosen company cared about sustainability (see Figure 4-Q38). The visualization introduced a sense of surprise: 31% of them were not aware of how much the chosen company cared about ISEs (see Figure 4-Q39).

Testing our algorithmic and interaction choices (post-treatment Q40–Q46): Most of our participants found the matched company to fit their own views on ISEs (Q40), adding external validity to our vector-based matching technique. No interaction strategy was preferred (Q41): 30% of participants preferred flipping the cards to get more details, 26% expressed a preference for hovering the colored box to see percentages, 24% preferred comparing their own card with the chosen company card, and 19% with the recommended company card. Showing percentages on

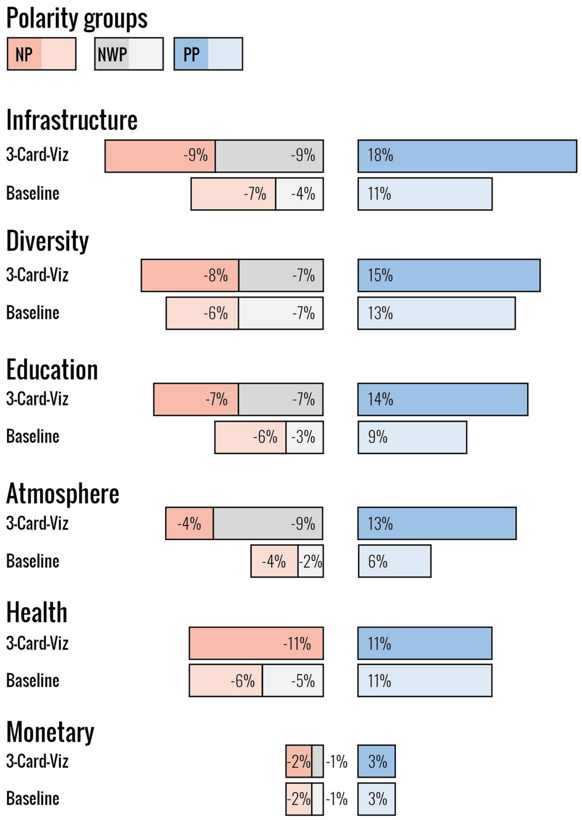

FIGURE 5. Percentage growth rates of three groups—NP, NWP, and PP—toward each ISE answered before the three-card viz and the baseline ($Q29_t$–$Q34_t$) and after them ($Q29_{(t+1)}$–$Q34_{(t+1)}$).

cards helped 47% of participants to explore the data (Q42), while 32% of them became more aware of their own values (Q43), and 61% of them had little prior knowledge about ISEs (Q45). After interacting with our visualization though, half of them became more aware of ISEs (as much as 49% were positively polarized to Q46). In between these questions, participants answered the second quality control question (Q44), which led us to filter out the contributions of an additional 19% of our participants.

Qualitative Results

The answers to the open-ended questions (see Figure 1-Step 7) were broken down into self-contained statements and labeled with concept categories (open coding), and then these concepts were grouped into themes (axial coding).[17] After the two coding steps, three themes emerged.

Theme 1—Communicating ISEs: Participants suggested entering into a dialog with companies for raising ISE awareness, both internally and externally. By internally, they meant *fostering communication between employees and employer* (P58). The discussion could be initiated at the recruitment stage, during team meetings, or through official communication channels. By externally, they meant communicating ISEs to the general public. Participants mentioned several channels on which such a promotion could take place (e.g., social media and companies' websites). Yet, such communication should pay attention to the so-called "crisis of buzzwords." Participants generally observed that current corporate communication tends to obfuscate internal efforts through "lip service, lame ads, and email campaigns" (P81). The challenge would be to convey a genuine tone as people "need actions rather than just words" (P135). Participants also noted that such methodologies could engage "other sustainability actors" (e.g., "public institutions, hackerspaces") in ISEs debate.

Theme 2—Making sense of ISEs: Two orthogonal sense-making strategies emerged—"soft" (emotion-driven) and "hard" (number based).

In the "soft" strategy, participants framed their data experiences as feelings, e.g., P84: "I feel like this company is like me," P171: "I got a better understanding of the company," P149: "I feel that the company engages in environmental practices," P186: "[...] the company believed in sustainability." Flipping one's card was strongly connected with "feeling the data." Participants indicated *joy* while "gaining information" (P146), *empathy* (P25: "reading real people's quotes gives a better feel of the company"), *suspense* (P195: "It is exciting to turn over a card without knowing what it'd reveal"), *curiosity* (P30: "[...] interested in what data was coming next"), and finally *control* (P231: "I like I was in control").

In the "hard" strategy, participants framed their data experiences as visual comparisons, e.g., P101: "numbers are clear and unambiguous," P218: "[...] numbers explain things." Displaying percentages facilitated "seeing the data to believe it." P97 mentioned that she "wanted to be sure the color blocks were based on correct numbers." Other participants found the scoring method reassuring (P206: "I'd never put a physical number to my values"). Others were familiar with certain ISEs (e.g., monetary), e.g., P83: "I care about monetary value," while some participants, using the card visualization, reflected on emerging concepts (e.g., atmosphere), e.g., P107: "Before this experience, I wouldn't have realized that I valued atmosphere so much.").

Theme 3—Evaluating companies against ISEs: Two participants mentioned that their personal values did not necessarily need to align with those of their company's (P103: "My values and my job are not the same thing [...] my job is a means to make money;" P54: "Disingenuous virtue signaling and mindless corporate pandering are not part of my values"). Other participants used the tool to re-evaluate known companies (e.g., P76 was surprised "how low [Brand X] rated"). P204 mentioned: "it makes me want to start looking into companies that I have been loyal for long time"). P50 noticed that "there are top brands which are committed to sustainable ways of doing business."

Interaction with the *chosen company* card helped some participants *reinforce* their opinions (P187: "Interesting to see how a company I would like to work for matched with my values.") or *question* a company's image. (P220: "Comparing my card to my selected company shows me how different we are in many ways.")

DISCUSSION AND CONCLUSION

After interacting with the visualization, 58% of our participants learned that sustainability is not a monolithic concept only tied to environmental resources but could rather be seen as a multifaceted concept (e.g., infrastructure support, workplace atmosphere, and wellbeing).

From a design perspective, our contribution is threefold. First, our DL-driven visualization created familiar and playful interactions (e.g., card games) through a treemap representation. Second, our methodology demonstrates the integration of four ML4VIS processes:[2]

1) *"data processing4VIS":* extracting mentions of ISEs from reviews;
2) *"data-VIS mapping":* automatically updating the cards whenever new reviews are processed;
3) *"style imitation":* generating dynamic cards with similar layouts to SDGs; and
4) *"user profiling":* analyzing user's quiz answers and providing his/her best matching company.

Third, we partly tackled the common distrust for black-box ML models by validating the effectiveness of interactions based on physical gestures (e.g., card flipping allowing for more content to be displayed), and by designing a user-model interaction that is blended (i.e., participants could not generally distinguish where their interactions with the model ended, and where their interactions with the visualization started).

Our work has two limitations. First, our DL algorithm processed reviews from U.S.-based companies and, as such, our findings may not generalize to wider populations or other organization types; the proposed method could be replicated to analyze other types of reviews and sustainability actors. Second, more research in the emerging field of ML4VIS should go into:

1) supporting the two orthogonal ways individuals typically use to make sense of data: "hard" (number based[15]) and "soft" (emotion-driven)[18];
2) increasing people's trust in ML tools (e.g., making algorithms more transparent[19]); and
3) how to avoid reinforcing incorrect views users may invariably hold because of their confirmation bias[20] (i.e., the tendency to believe only the information that confirms one's prior beliefs).

REFERENCES

1. I. Sen, D. Quercia, L. Capra, M. Montecchi, and S. Šćepanović, "Insider stories: Analyzing internal sustainability efforts of major US companies from online reviews," 2022, *arXiv:2205.01217*.
2. Q. Wang, Z. Chen, Y. Wang, and H. Qu, "A survey on ML4VIS: Applying machine learning advances to data visualization," *IEEE Trans. Vis. Comput. Graphics*, early access, Aug. 26, 2021, doi: 10.1109/TVCG.2021.3106142.
3. S. Cunningham-Nelson, M. Baktashmotlagh, and W. Boles, "Visualizing student opinion through text analysis," *IEEE Trans. Educ.*, vol. 62, no. 30, pp. 305–311, Nov. 2019.
4. G.-M. Emma and M.-F. Jennifer, "Is SDG reporting substantial or symbolic? An examination of controversial and environmentally sensitive industries," *J. Cleaner Prod.*, vol. 298, 2021, Art. no. 126781.
5. C. Sneha, E. Sarangh, V. Vindhya, R. R. Hegde, and V. A. Lakshmi, "Comprehensive analysis of CSR data using interactive reports," in *Proc. IEEE Conf. Energy, Commun., Data Analytics Soft Comput.*, 2017, pp. 1460–1463.
6. V. D. Swain *et al.*, "Modeling organizational culture with workplace experiences shared on glassdoor," in *Proc. ACM CHI Conf. Hum. Factors Comput. Syst.*, 2020, pp. 1–15.
7. N. Luo, Y. Zhou, and J. Shon, "Employee satisfaction and corporate performance: Mining employee reviews on glassdoor.com," in *Proc. Int. Conf. Inf. Syst.*, vol. 1, 2016, pp. 387–402.
8. N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," in *Proc. ACL Conf. Empirical Methods Natural Lang. Process.*, 2019, pp. 3982–3992.

9. M. Dittus, L. M. Aiello, and D. Quercia, “Community Engagement triage: Lightweight prompts for systematic reviews,” *Proc. ACM Hum.-Comput. Interaction*, vol. 1, pp. 1–22, 2017.
10. B. Bach *et al.*, *Narrative design patterns for data-driven storytelling, in Data-Driven Storytelling*. New York, NY, USA: Taylor & Francis, 2018, ch. 5, pp. 107–133.
11. C. Fiedler, W. Scheibel, D. Limberger, M. Trapp, and J. Döllner, “Survey on user studies on the effectiveness of treemaps,” in *Proc. ACM Symp. Vis. Inf. Commun. Interaction*, 2020, pp. 1–10.
12. M. Bruls, K. Huizing, and J. J. Van Wijk, “Squarified treemaps,” in *Data Visualization*. Berlin, Germany: Springer, 2000, pp. 33–42.
13. S. Werning, “Making data playable: A game co-creation method to promote creative data literacy,” *Media Lit. Educ.*, vol. 12, no. 3, pp. 88–101, 2020.
14. S. D. Gosling, P. J. Rentfrow, and W. B. Swann Jr, “A very brief measure of the big-five personality domains,” *Res. Pers.*, vol. 37, no. 6, pp. 504–528, 2003.
15. A. V. Pandey, A. Manivannan, O. Nov, M. Satterthwaite, and E. Bertini, “The persuasive power of data visualization,” *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 12, pp. 2211–2220, Dec. 2014.
16. C. J. Soto, O. P. John, S. D. Gosling, and J. Potter, “Age differences in personality traits from 10 to 65: Big five domains and facets in a large cross-sectional sample,” *Pers. Social Psychol.*, vol. 100, no. 2, pp. 330–348, 2011.
17. V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qualitative Res. Psychol.*, vol. 3, no. 2, pp. 77–101, 2006.
18. H. Kennedy and R. L. Hill, “The feeling of numbers: Emotions in everyday engagements with data and their visualisation,” *Sociology*, vol. 52, no. 4, pp. 830–848, 2018.
19. T. Cerquitelli, D. Quercia, and F. Pasquale, *Transparent Data Mining for Big and Small Data*. Berlin, Germany: Springer, 2017.
20. E. Graells-Garrido, M. Lalmas, and D. Quercia, “Data portraits: Connecting people of opposing views,” 2013, *arXiv:1311.4658*.

CHIARA CECCARINI is a Ph.D. candidate at the University of Bologna, 40126, Bologna, Italy. Her research interests include human–computer interaction and data visualization. She is the corresponding author of this article. Contact her at chiara.ceccarini6@unibo.it.

EDYTA PAULINA BOGUCKA is a Ph.D. candidate at the Technical University of Munich, 80333, München, Germany. Her research interests include visual storytelling and urban cartography. Contact her at e.p.bogucka@tum.de.

INDIRA SEN is a Ph.D. candidate at the Gesis Leibniz Institute for Social Sciences, 68159, Cologne, Germany. Her research interests include computational social science, natural-language processing, and measurement theory. Contact her at indira.sen@gesis.org.

MARIOS CONSTANTINIDES is a senior research scientist at Nokia Bell Labs, CB3 0FA, Cambridge, U.K. His research interests include human–computer interaction and ubiquitous computing. Contact him at marios.constantinides@nokia-bell-labs.com.

CATIA PRANDI is an assistant professor at the University of Bologna, 40126, Bologna, Italy and a faculty fellow at the Interactive Technologies Institute (ITI/LARSyS), Funchal, Portugal. Her research interests include human–computer interaction and interactive storytelling for social good. Contact her at catia.prandi2@unibo.it.

DANIELE QUERCIA is the department head at Nokia Bell Labs, CB3 0FA, Cambridge, U.K. and a professor of urban informatics at Kings College London, London, U.K. His research interests include social computing, urban studies, and data science. Contact him at quercia@cantab.net.