Blueprints: Systematizing Behavior Change Designs—The Case of Social Comparison Theory

ROELOF A. J. DE VRIES, Biomedical Signals and Systems, University of Twente, The Netherlands
MAILIN LEMKE, Human-Centered Design, Delft University of Technology, The Netherlands
GEKE D. S. LUDDEN, Design, Production and Management, University of Twente, The Netherlands

11

To improve people's lives, human-computer interaction researchers are increasingly designing technological solutions based on behavior change theory, such as social comparison theory (SCT). However, *how* researchers operationalize such a theory as a design remains largely unclear. One way to clarify this methodological step is to clearly state which functional elements of a design are aimed at operationalizing a specific behavior change theory construct to evaluate if such aims were successful. In this article, we investigate how the operationalization of functional elements of theories and designs can be more easily conveyed. First, we present a scoping review of the literature to determine the state of operationalizations of SCT as behavior change designs. Second, we introduce a new tool to facilitate the operationalization process. We term the tool *blueprints*. A blueprint explicates essential functional elements of a behavior change theory by describing it in relation to *necessary* and *sufficient* building blocks incorporated in a design. We describe the process of developing a blueprint for SCT. Last, we illustrate how the blueprint can be used during the design refinement and reflection process.

CCS Concepts: • **Human-centered computing** → **HCI theory, concepts and models**;

Additional Key Words and Phrases: Behavior change theory, social comparison theory, intermediate-level knowledge, scoping review, methods, operationalization, behavior change design, multidisciplinary research

ACM Reference Format:
Roelof A. J. de Vries, Mailin Lemke, and Geke D. S. Ludden. 2023. Blueprints: Systematizing Behavior Change Designs - The Case of Social Comparison Theory. *ACM Trans. Comput.-Hum. Interact.* 31, 1, Article 11 (November 2023), 32 pages.
https://doi.org/10.1145/3617364

1 INTRODUCTION

The attempt to influence a person's behavior by nudging or by providing motivation or support for the individual's and societal benefit has attracted increased interest from design research [59, 80]. **Human-Computer Interaction (HCI)** research has also started to explore designing

This research was funded by the Pride and Prejudice project funded by the 4TU. Federation in the Netherlands (grant no. 4TU-UIT-346) (https://www.4TU.nl).
Authors' addresses: R. A. J. de Vries (Corresponding author), Human Experience and Media Design, Utrecht University of Applied Sciences, The Netherlands; e-mail: rajdevries@gmail.com; M. Lemke, Human-Centered Design, Delft University of Technology, Landbergstraat 15, Delft, The Netherlands; e-mail: m.lemke@tudelft.nl; G. D. S. Ludden, Design, Production and Management, University of Twente, Drienerlolaan 5, Enschede, The Netherlands; e-mail: g.d.s.ludden@utwente.nl.

1073-0516/2023/11-ART11
https://doi.org/10.1145/3617364

technology to help people change behaviors [41]. In a similar vein as behavioral science—where it is common practice to base behavior change interventions on theories or models of behavior change (e.g., [16, 55, 67, 86])—HCI researchers are also basing their technological interventions on behavior change theory or models. Ideally, with theory-based interventions in behavioral science, an inference loop is established where theory can help guide and ground the process of designing an intervention, and in turn, the evaluation of the intervention can help evaluate and inform the theory. Unfortunately for behavioral scientists and HCI researchers alike, behavior change theories are usually not sufficiently specific to offer a basis for an intervention, let alone for the functional basis of a design [14, 55, 61]. In turn, these designs or implementations are usually not informative enough to help in evaluating or reflecting on the theory.

A common practice in HCI literature is to include a section outlining the *implications for design* based on a specific study's findings [26, 73], but there is often a lack of explanation on *how* a design was grounded in a specific theory in the first place and how this influenced the evaluation process [49, 64]. The lack of explanation on operationalization practices and the lack of specificity in behavioral theories lead to the process of basing **Behavior Change Designs (BCDs)** on theories seemingly taking place in a kind of black box (Figure 1).

Moreover, BCD is a multidisciplinary research field that requires integrating and synthesizing currently mostly separated research disciplines. Establishing a shared framework or tool that different disciplines can understand and use is essential to allow multidisciplinary teams to work jointly on a shared goal and learn from each other [15]. For example, in the context of ubicomp applications where different disciplines work together (e.g., designers, architects, and programmers), specific tools such as personas or system maps have been used to try to bridge the disciplinary division to develop, refine, and create suitable design applications [27]. However, in the context of BCD, the use of theory poses a challenge. On the one hand, theories are often too abstract for design and HCI practitioners to use as they provide little concrete guidance in design decisions. On the other hand, design instances are too concrete and situated to contribute to the development and refinement of a theory effectively. Tools and methods that could facilitate the design and reflection process seem to be scarce.

A shared foundation needs to be established to facilitate multidisciplinary and interdisciplinary research. This shared language, method, tool, framework, or even research paradigm might likely feel unfamiliar to a theory-focused researcher as well as the practitioner. However, a clear description of and a systematic approach to the grounding process of design elements in a specific theory would allow researchers to replicate or reproduce relevant components, open up the design space for the formation of more abstract knowledge, and hopefully support the inference loop for the development of more effective and functional BCDs. With this article, we aim to address this gap by providing a tool that can be used in the design development and reflection process to point out which and how functional elements of a design are based on a specific theory.

2 THEORY AND PRACTICE

In the domain of HCI and design, the metaphorical *gap* or *black box* between practice and theory has been investigated. Discussions center on what the value of a theory is in this particular field and how to bridge the gap [8]. The current discourse around the value of theory can be broadly separated into two main groups: the *scientific camp*, which emphasizes that design decisions should be based on theoretical principles and models, and the *pragmatic camp*, which highlights the designerly ways of knowing that are often tacit and that the role of theory is merely to inspire, reflect, and help a designer in practical choices (and not to serve as a basis) [66, 72].

Wiberg and Stolterman [88, p. 533] point out that HCI research and many of its methods have a dominant focus on developing methods that help generate new ideas rather than supporting

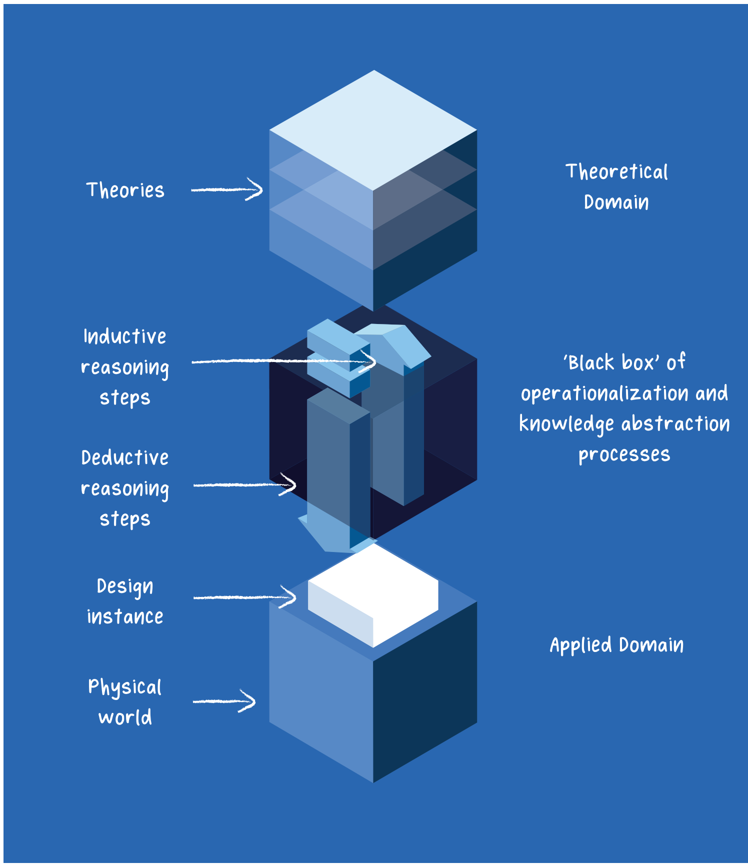

Fig. 1. The current operationalization process of using a theory in a design seems to take place in a kind of black box. As part of the deductive reasoning process, theories are operationalized into a design instance which itself is placed into the physical world. With inductive reasoning, findings are abstracted from such concreted design instances and fed back into theory. However, concrete descriptions of how these steps are performed seem to be missing in most research projects.

attempts to trace back ideas in a design. However, there have been attempts to address the need to communicate interaction research-specific knowledge clearly [43] and bridge the gap between practice and theory in the form of so-called **Intermediate Levels of Knowledge (IMLK)** [8, 22, 44, 52, 53] (Figure 2). IMLK was introduced as the space between these two poles and entailed "knowledge that is more abstracted than particular instances, without aspiring to be at the scope of generalized theories" [44, p. 1]. Several forms of IMLK have been introduced over the years, such as "Bridging concepts" [22], "Strong concepts" [44], "Annotated portfolios" [33], or "Patterns" [10, 24], to name just a few (see Figure 2). It appears that most IMLK often has a bottom-up perspective, focusing on the pragmatic view of abstracting findings from prototypes [32, 33, 44, 52, 90], and just a few have focused on the scientific view of theory and design simultaneously to develop more abstract levels of knowledge [22]. Approaches such as concept-driven design suggested by Stolterman and Wiberg [76] use theory as a main starting point but focus explicitly on novel and futuristic designs rather than pointing out what functional elements of a design operationalize concepts of a specific theory.

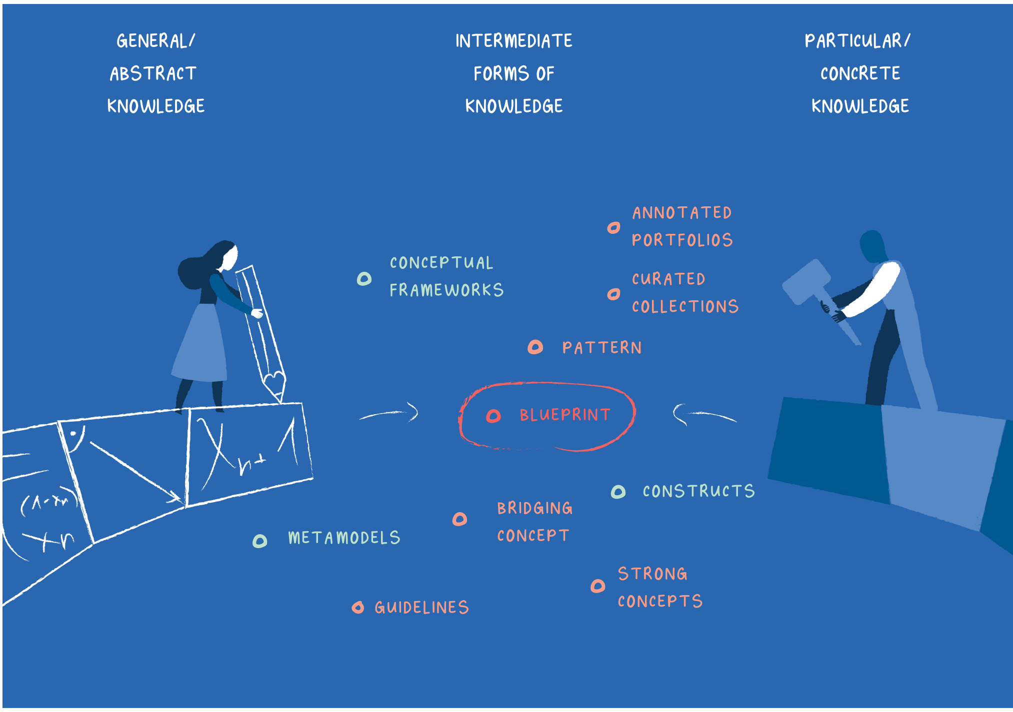

Fig. 2. Different IMLK have been introduced in the literature to bridge the gap between theoretical and practical knowledge. Metamodels, conceptual frameworks, and constructs originate out of reflections on characteristics of behavior change theories. Annotated portfolios, curated collections, patterns, bridging concepts, strong concepts, and guidelines were introduced by design and HCI researchers. Our proposed blueprint resembles the approach of patterns and bridging concepts but also differs in its focus on functional elements needed for a representative BCD. © Vectorarte.

2.1 Behavior Change Theory and Practice

Basic psychological research has shown that human behavior follows predictable patterns [54]. It has been noted that to develop interventions that aim to evoke a specific behavioral response in the user, it is necessary to understand the theoretical underpinning, including the moderators of change and mechanisms of action. To facilitate the development of suitable BCD applications, researchers can draw from a large repertoire of theories of behavior and theories of behavior change [23].

It could be argued that such theories used in the context of BCD and HCI are a "special use case" as the development process is based on the premise that a *specific theory or behavior change technique* will evoke a *specific behavioral response*. From this point of view, theories in this context cannot be reduced to just a source of inspiration if the specific behavioral response is still expected to be evoked. Theories instead should be seen as a basis for design requirements to enable the intended effect. However, it can be challenging for practitioners to use a theory, which is by definition abstract, to apply and justify specific design decisions. We argue that the specific case of BCD calls for a methodological focus, specifically on the functional relation between theory and the functional building blocks of a BCD.

In this article, we introduce a tool to facilitate this exchange between behavior change theory and practice (see Figure 2). We termed our tool *blueprints* as a metaphor for a common blueprint needed to construct a specific object. A typical blueprint is a technical drawing that contains all the *essential elements* to reproduce the depicted design rapidly and accurately without containing unnecessary information. They are signified by details that can be described as *sufficient* and *necessary*. A blueprint specifies the essential or minimal building blocks needed to represent the theory. We define *minimal building blocks* as functional elements that are simultaneously sufficient and necessary. Sufficient means that the defined building blocks are enough to embed the functional elements of a specific behavior change theory (as there might be different definitions of a theory, we focus on it being based on a particular definition of the theory). Necessary means that the building blocks are all separately needed to allow the intended effect of the theory. Approaches with a similar focus include "Bridging concepts" and the notion of "Patterns" and aim to facilitate exchange between theory and practice and focus on best practice examples [2, 10, 22, 84]. Moreover, *generic design thinking* [88] shows a close resemblance to our approach and follows a process of analyzing, comparing, and contrasting designs, making it a tool for analysis as well design. However, the approaches lack the notion of specific functional elements that are *required* in a design to evoke an intended behavior in the user.

We focus for the remainder of this article on **Social Comparison Theory (SCT)** [31] to point out current gaps in regard to the operationalization of a behavior change theory as a BCD and to illustrate the development and use of a blueprint. We chose SCT due to its potential positive and negative influence on people's behavior, which requires careful consideration in the design process. Moreover, social comparison is a ubiquitous social phenomenon, and the theory has a long-standing tradition in the context of psychology. Last, social comparison is also outlined as a concrete behavior change technique that is frequently used in persuasive research applications [62, 78]. The blueprint we present in this article is based on and restricted to SCT. However, we argue that the concept of blueprints we propose can be applied to other theories of behavior change. Due to the varying degrees of complexity of some of these theories, it seems likely that some of them are more challenging to attempt to create a blueprint for than others. We approached this development of a blueprint in two steps. First, on the one hand, we wanted to understand the breadth and depth of SCT and, on the other hand, also the way SCT is currently operationalized in HCI BCDs. Second, we tried to determine the essential building blocks of said designs and theory to develop a blueprint of SCT that other researchers can refer to. We therefore focused on two different research questions:

- *RQ1*: How is social comparison theory being applied in designing behavior change\break technologies?
- *RQ2*: What are the minimal building blocks needed to describe a social comparison design?

To address our research questions, we structured this article as follows. We address RQ1 with a scoping review to gain an understanding of how SCT is used as part of BCDs in HCI and design literature and how the operationalization process is described. To address RQ2, we followed an iterative design process to develop a blueprint that sufficiently represents social comparison for behavior change. We conclude our work by discussing the implication of our proposed SCT blueprint and the concept of blueprints in general.

Our main contribution is the new intermediate level of knowledge (i.e., blueprints) that bridges the areas of behavior change theory and BCD grounded in such theories. Our critical analysis of the current use of SCT in BCD offers an overview of the current operationalization process and inclusion of social comparison constructs in HCI behavior change projects. Our blueprint of social comparison outlines essential building blocks to incorporate into a design and can be used as a foundation for future research. The concept of a blueprint and outlined steps that we report on to determine essential building blocks for a BCD can be applied to other behavior change theories commonly used in HCI.

With this tool, we aim to facilitate design-oriented researchers who want to use theories as part of their process. We see such designers leaning toward a conservative account of the discipline rather than perceiving the level of ‘creative geniusness’ of a designer as the main guarantee for successful designs. The conservative account of design is signified by a rational search process starting with determining requirement specification and resulting in the development of a concrete artifact [30]. In addition, we focus in this work on the role of artifacts produced by design-oriented researchers who focus on producing knowledge, whereas design practitioners focus on creating commercially successful designs [91]. We acknowledge that some researchers use theory to support and enable specific behaviors, for example, through automating interaction or using artificial intelligence. Although such applications might benefit from the tool we propose, they are not the focus of this work. We also acknowledge that design practitioners working outside a research context have differing needs and requirements regarding tools and methods [17, 20, 60], and they tend to perceive theory as a source of inspiration in the design process [18]. The tool we propose is therefore more suitable for research projects rather than design practice [20].

3 REVIEWING SCT IN BCDS

A recent scoping review on persuasive technology indicated that many steps are left unclear in the operationalization process of using SCT as part of a BCD [49]. The review also highlighted a diversity of design implementations which arguably complicates attempts to compare the different designs as well as the empirical findings. Furthermore, this diversity of designs is likely to impact the possibility of abstracting knowledge from the design instances to understand which design elements and features are essential in evoking a social comparison effect. The review pointed toward a general lack of operationalization descriptions but was in its scope limited to conference papers published as part of the Persuasive Technology conference. To gain a broader perspective on the operationalization process of researchers and designers involved in creating BCDs, we conducted a scoping review of the literature focusing on the CHI and DIS conference proceedings. With this review, we address RQ1: How is social comparison theory being applied in designing behavior change technologies? We chose to scope this review on the CHI and DIS conferences because these venues are well known for publishing multidisciplinary work in behavior change theory, HCI, and design, and this type of work is arguably well suited to tackle the problem of operationalizing theory into designs. We chose a scoping review to “examine how research is conducted on a certain topic or field [and] to identify and analyze knowledge gaps” [58, p. 2]. In the next section, we outline the core elements of SCT followed by a description of the scoping literature review process and results.

3.1 SCT and Its Use

SCT was proposed in 1954 by Leon Festinger [31] and has been an active research topic in social and personality psychology to date [35]. The theory centers around the belief that in the absence of objective and non-social norms, people tend to self-evaluate personal abilities and opinions (comparison dimensions) by comparing them to other individuals or groups (comparison targets). The way this comparative effort is performed can be automated as well as deliberate [37] and take place in an upward, lateral, or downward direction depending on the perceived position of the comparison target to the individual [4, 12]. The motive to conduct social comparison can vary and help to self-evaluate, and self-improve as well to enhance some aspects of the self [12].

Despite being a relatively complex and abstract theory (Figure 3 presents an overview of influencing factors), social comparison is also used as a behavior change technique included in different frameworks, including the “Persuasive System Design Model” [62] and the “Behavior Change Taxonomy” that includes 93 different behavior change techniques [56]. In these frameworks, the

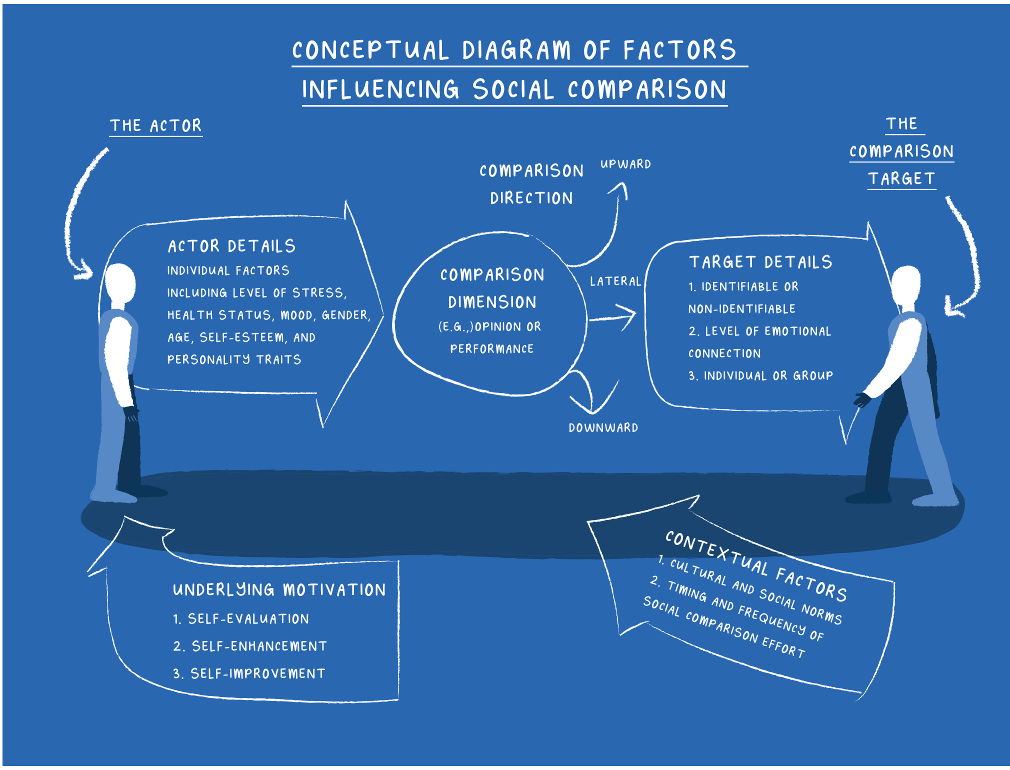

Fig. 3. Since its introduction in 1954, multiple studies have broadened our understanding of factors influencing social comparison efforts. Such factors include the characteristics of the actor, the comparison target, and the underlying motivation to engage in social comparison. Further factors relate to the comparison dimension and direction. Last, contextual factors can have an influence.

definitions of social comparison seem to be more straightforward and functional. For example, the definition of Michie et al. [56, p. 9] of social comparison as behavior change technique is to "draw attention to others' performance to allow comparison with the person's own performance. Note: being in a group setting does not necessarily mean that social comparison is actually taking place." In addition, the definition of Oinas-Kukkonen and Harjumaa [62, p. 495] focuses on an implementation version of social comparison: "System should provide means for comparing performance with the performance of other users." Social comparison embedded in persuasive designs has been pointed out to be perceived as more persuasive than the related techniques of *cooperation* and *competition*. Its effectiveness has been linked to peer pressure because it allows being a role model for others and contributes to reaching a goal due to group enforcement. However, several negative influences were also recorded in this context. Examples include privacy concerns due to the invasive character, a potential reduction in self-confidence, the rise of depression, risk of alienation, and the potential of backfiring and reinforcing unhealthy behavior if the comparison target displays harmful behavior, which is replicated [63].

It has been pointed out that when social comparison is included in BCDs, there seems to be a lack of awareness of key aspects and consequences of social comparison [5]. Moreover, it is often unclear how the comparison effect has been embedded in the design [3] and how SCT has been operationalized [49]. These findings are of concern as it complicates attempts to claim unique scientific contributions and develop theories that others can use due to a lack of rigor and providing convincing evidence [13]. To gain a broader perspective on the operationalization process of researchers and designers involved in creating BCD, we conducted a scoping review of the literature focusing on the CHI and DIS conference proceedings.

3.2 Scoping Literature Review

To investigate whether the multidisciplinary work in HCI and design, similar to the scoping review by Lemke and de Vries [49], suffers from "grounding" issues in the operationalization of social comparison, we specifically looked at venues where this type of multidisciplinary work is published, such as the CHI and DIS conferences. Similar to Lemke and de Vries [49], we refer to the term *grounding* as the process of how the theory was included in a specific design prototype. We refer to design prototypes in the context of this study as "a representation of a design idea, regardless of medium" [45, p. 369].

3.2.1 Identification and Screening. This scoping review followed the PRISMA Extension for Scoping Reviews guidelines [79]. We used two sources for the review: the CHI conference proceedings[1] consisting of 20,990 papers published between 1981 and 2020, and the DIS conference proceeding[2] consisting of 1,850 papers published between 1995 and 2020. This scoping review exclusively focused on peer-reviewed research articles and excluded demonstrations, late-breaking work, workshop proposals, and posters.

In January 2021, we conducted a computerized search using the search term "social comparison." Studies had to include the search term in the title, abstract, or main manuscript. All papers were assessed for eligibility and coded by two reviewers (the first and second author). Abstracts and titles were screened first for eligibility, and in a consequent step, eligible papers were read in full. Any discrepancies were resolved through discussion. We used seven inclusion criteria: (1) SC is part of the design, (2) a visual of the design is included, (3) it is clear which design part/feature relates to SC, (4) the paper is published as part of DIS or CHI, (5) the paper is not retracted, (6) the paper does not discuss a design from another paper, and (7) the design is evaluated with participants. We chose these inclusion criteria to be able to code the grounding and operationalization choices explained in the papers.

3.2.2 Data Extraction. For the data charting process, we captured relevant information electronically. For the coding, we chose categories based on a scoping review investigating the use of SCT in design [49]. We coded the data for the following aspects (Table 1):

- The *Ref.* column refers to the reference to the foundational social comparison definition—for example, did the author simply state they included social comparison or added a reference to a paper defining what social comparison entails?
- The *Grounding* column refers to the explanation of how the theory was used to justify specific design implementations—for example, the design of specific interactive elements that allow sharing one's performance with friends.
- The *Check* column refers to the use of manipulation check as part of the evaluation process to assess if the SCT was recognized as such.
- In the *Direction* column, a description of the use of social comparison in an upward, lateral, or downward direction is provided. Upward social comparison direction refers to the user perceiving other users as superior, lateral direction refers to the user perceiving others on the same level, and downward direction refers to others being perceived as inferior.

[1] https://dl.acm.org/conference/chi
[2] https://dl.acm.org/conference/dis

Table 1. Summary of Dimensions of Social Comparison Information Extracted from Papers Included in the Scoping Review Relevant for Operationalization

Paper	Ref.	Grounding	Check	Direction	Comb.	Target	Outcome	SC eval.
[69]	Yes	Unclear	Partially	Unclear	Yes	Neighbors (first application of social comparison), friends (second use of social comparison), and users of the application that have similar relevant attributes such as household size (third use)	Qualitative feedback	Yes
[19]	Yes	Unclear	Unclear	Upward	No	First condition (leader, unknown person), and second and third condition (similar player, unknown)	Video game self-efficacy, social comparison, self-esteem, and an intrinsic motivation inventory, as well as open-ended questions about their goals for playing and why they left the game	Yes
[46]	Yes	Unclear	Unclear	Study 1: unclear Study 2: all	No	Other participants	Test score of the social intelligence test; reading the mind in the eyes test	No
[25]	Unclear	Unclear	Yes	Unclear	Yes	Other healthy adults (within group), other overweight	Changed calories	Yes
[50]	Yes	Some	Unclear	Unclear	Yes	Two other users of the system	Qualitative feedback	No
[82]	Yes	Some	Unclear	Unclear	Yes	Other users of the system sharing their energy consumption; average energy consumption of the city district; average energy consumption of the whole city	Qualitative feedback	No
[87]	Yes	Some	Unclear	Unclear	Yes	Like-bodied mothers	Qualitative feedback	No
[81]	Unclear	Unclear	Unclear	Unclear	Yes	Other user	Qualitative feedback	No
[74]	Yes	Some	Unclear	Unclear	Yes	Other user/s of the interactive installation	Qualitative feedback	Unclear
[29]	No	No	No	Lateral	Yes	Similar households	Logs, survey, interviews	No
[28]	No	No	No	Lateral	Yes	Similar households	Logs, survey, interviews	Unclear
[65]	Yes	Yes	Yes	Unclear	No	Friends	Perceived persuasiveness, open-ended question	Yes

- The *Comb.* column refers to social comparison being combined with other strategies and theories in the BCD, such as by including SCT and goal-setting theory in a BCD.
- The *Target* column refers to provided information concerning the target for SC—for example, outlining that the target is a neighbor living next door.
- The *Outcome* column entails any information about the evaluation process of the design—for example, if qualitative or quantitative methods were used.
- The *SC eval.* column states if a post hoc evaluation of the social comparison effect was conducted.

We also captured design-specific information relevant to the context of operationalization (Table 2):

- The *Design* column refers to the form the design incorporating social comparison took on—for example, was the design an interactive installation or a game concept?
- The *Study context* column refers to the type of behavior targeted—for example, did the design intend to contribute to healthier food consumption or decrease the carbon footprint?

Table 2. Summary of Dimensions of Social Comparison Information Extracted from Papers Included in the Scoping Review Relevant for Design Implementation

Paper	Design	Study context	Medium	Type	Dimension
[69]	Online application	Energy saving	1: Text based 2: Text and visualization based 3: Text based	1: You vs unknown neighbor 2: You vs friend 3: You vs group	Conserving energy (KWh and KW)
[19]	Online game and comparison feedback screen	Playing an online game	1. Text and visualization based 2: Text and visualization based 3: Text and visualization based	1: You vs the 'leader' 2: You vs 'similar player' 3: You vs the 'leader'	High score of the game
[46]	Online test	Doing a test, social media sharing the results	Text based	You vs all other participants	Test score
[25]	Caloric overview of shopping basket	Choice in calories during grocery shopping	Text based	You vs other healthy adults and you vs other overweight adults	Reduction of calories
[50]	Online chat program and different visualizations	Visualizing language use as part of an online chat program	1: Visualization based 2: Visualization based 3: Visualization based	You vs two other identifiable users (name displayed)	Study 1: emotion words, word count, self-reference; study 2: conversational activity (measured by word count) and agreement with the group (measured by the proportion of agreement words); study 3: agreement words
[82]	Public, interactive sunburst diagram of energy consumption in a city	Energy consumption of individuals	Visualization based	You vs non-identifiable person (pseudonym); you vs city district (neighborhood) usage; you vs city-wide usage	Energy consumption
[87]	Digital app	Fostering healthy (and peaceful) behavior among expecting mothers	Visualization based	You vs like-bodied (anatomized group of mothers)	Symptoms, weight, nutrition, activity level, hydration, and sleep
[81]	Google Docs application	Online analytic tool for exploring writing behavior and reflection	Visualization based	User 1 vs User 2 (who "you" are is unclear)	Revision events, insertion and deletion events
[74]	Interactive installation and app	Climate impact of dietary choice	Visualization based	You vs 7 unidentifiable individuals (users of the same system) + the average of all	Climate impact (relative carbon emission)
[29]	Online water consumption dashboard with a weekly game challenge and 'neighbors like you'	Supporting voluntary reductions of water consumption	Text and visualization based	You and your group (4–6 households) vs another group; you vs a similar neighbor	Water usage
[28]	Online energy consumption dashboard with comparisons via text and graphs	Supporting voluntary reductions of electricity consumption	Text and visualization based	Similar households	Electricity usage
[65]	Storyboard	Persuasive health games and gamified systems	Scenario based	You vs friends (represented through one score)	BAC/alcohol usage

- The *Medium* column includes the characteristics of the implementation form of social comparison—for example, was the social comparison effect included in a text of visual diagram?
- In the *Type* column, details about the comparison target are provided that explain who "you" are compared against—for example, did the BCD allow to decipher particular other users of the system, or was the target included as an anonymous group average?

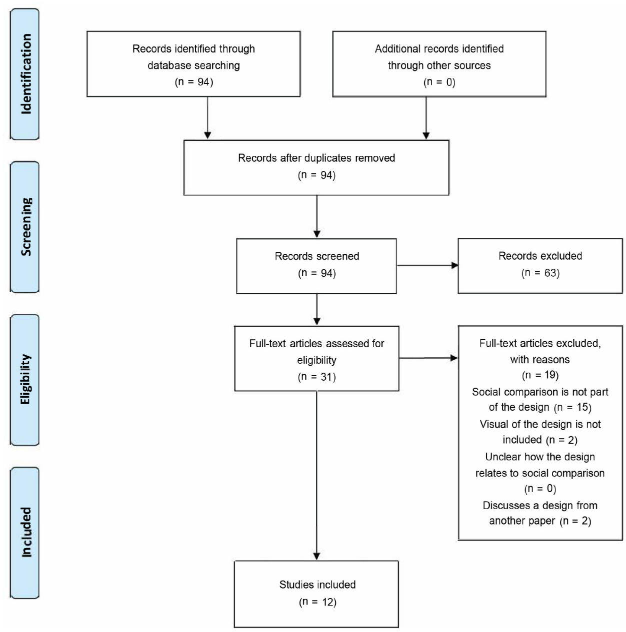

Fig. 4. PRISMA flowchart of the scoping review process.

- The *Dimension* column includes details of what exactly the user can compare themselves on, such as the number of steps taken during the day.

The extracted data were summarized in a narrative report. We discuss the results in relation to the research question and relevant literature findings in the following section.

3.2.3 Results and Discussion. The search term "social comparison" led to 94 unique articles. A total of 63 papers were excluded in the first screening step. Of the 63 papers excluded, 52 did not use SCT as part of the design prototype but rather mentioned SCT in the introduction or discussion to contextualize their work. Two articles were excluded because they did not contain a visual of the BCD, and 6 were excluded because it remained unclear how the BCD related to SCT. For example, some papers would mention SCT to be an essential feature, but in the description of the design, it was unclear which element of the BCD was conceptualized to allow and facilitate SCT. This does not mean that the excluded papers were unsuccessful in incorporating SCT but just that it remained unclear what design feature was supposed to contribute to the intended SC effect. Two papers were excluded because they were not officially published as part of CHI or DIS, and 1 paper was not available anymore on the database.

In a second step, 31 articles were read in full, of which 19 were excluded with reason (Figure 4). Of those 19 papers, 15 papers were excluded because SCT was not part of the design, and 2 papers were missing visuals of the BCD. The inclusion of the visuals of the final design prototypes in open data depositories could help mitigate this effect in future studies and help disseminate design-specific knowledge that is difficult to capture in text form. Two articles were excluded because they discussed a design from another paper already included in the scoping review. In the end, we included 12 papers in the qualitative synthesis. The coding results of the included 12 articles are reported in Tables 1 and 2.

Eight of the 12 papers included a *reference* as part of the study description. Six references cited the original paper by Festinger [31], one mentioned the work by Webb et al. [86], and one work by Lehto et al. [48]. This is an interesting finding since the original paper by Festinger [31] focuses on comparative efforts in the context of abilities and opinions and lacks a reflection on the underlying motives of engaging in social comparison, including the wish to self-evaluate, self-improve, and self-enhancement [12]. Furthermore, a more functional description of SCT as included in "Persuasive System Design Model" [62] or the "Behavior Change Taxonomy" [56] might seem more straightforward to be applied in a design context but was missing in the scoped papers. A detailed understanding of a particular behavior change theory or behavior change technique could help reflect on whether the operationalization is appropriate. For example, research findings have allowed painting a more detailed picture of social comparison beyond Festinger's initial description.

Five of the 12 papers included some kind of *grounding*, whereas 7 papers left it unclear if some sort of grounding was used for the study design. SCT was not referred to in a detailed predictive or prescriptive form in that the theory was used to guide the design process and justify decision making about the design of products and services [9]. This observation is not surprising as tools to explain the operationalization process of a theory into a BCD seem to be missing.

The *manipulation check*, which entails including one or more questions geared toward assessing if the condition to which participants were exposed was successful [42], was included as part of two studies. One additional study had partially included a manipulation check, and the remaining studies left it unclear if such an evaluative step had been carried out. This observation seems to be somewhat surprising as manipulation checks can be relatively easily implemented in a study design as part of the evaluation process and serve as a way to show that the design intentions were successful.

The *comparison direction*, which relates to one's perceived position in relation to the comparison target and can take place in upward, downward, and lateral comparison directions, was rarely explicitly defined and assessed. One study used upward social comparison, two studies used lateral comparison, and one included all three directions. The remaining studies did not specify if the comparison direction was purposefully conceptualized as part of the study design. This is an unexpected result since the direction of social comparison can have a significant influence [12] and is commonly assessed in psychology studies focusing on the daily experience of social comparison [4]. The context of use might also be an essential consideration. For example, people with severe illnesses might compare themselves with others. Still, they might be motivated to contrast themselves with people who are coping worse than they do following a downward direction [21]. However, some users might experience an increased motivation to change a specific behavior (e.g., running a marathon) when they compare their performance with people who seem alike (e.g., similar age and level of fitness) but perform better (upward comparison direction). Adding a justification and description of the anticipated comparison direction could help clarify research findings, and one could potentially even add measurement tools to check and evaluate the intended effect.

Nine of the 12 papers *combined* SCT with other theories or constructs as part of the study design. The combination of multiple strategies has been pointed out to limit the possibility to establish a causal link between a specific behavior change technique and the success of the BCD [41]. A focus on sophisticated statistical techniques or focus on certain behavior change strategies in isolation has been proposed to address this shortcoming [64].

The social comparison *target* that the user compared themselves with included a range of different people, including neighbors living in the same area as the user, generic users, and specific users of the same system. A detailed assessment of the effect that the chosen comparison target has as part of the design could help to clarify research findings since the perceived similarity to the comparison target (e.g., a specific person) can significantly influence the assimilative (e.g., wanting to be like that person) or contrastive (not wanting to be like that person) effect of SC [77]. In this context, some comparison targets might be more beneficial than others depending on the chosen comparison dimension [5], and it has been pointed out that users of persuasive systems prefer targets that they can relate to [89]. For example, designing a website that allows users to compare their energy consumption with their friends might be more effective than letting users compare themselves with unidentifiable users or even user groups.

The *outcome* of the designs was mainly assessed with qualitative feedback (including interviews and open-ended questions) in 10 out of 12 papers. The evaluation of qualitative data as an analytical strategy in theory-driven studies has been indicated as a highly relevant alternative in HCI compared to the current gold standard of randomized controlled trials in the context of BCD [41]. In addition to this, have mediational/path and moderation analyses, as well as alternative experimental designs such as N-of-1 experimental setups, been pointed out as alternative study designs [41]?

The *social comparison evaluation* was specifically focused on in 4 out of 12 papers. This limits the abstraction of research findings as it remains unclear if the remaining 8 design papers successfully induced social comparison in the user. Even if they were not successful, this could still be informative in developing future, more efficient ways of designing BCD based on SCT.

The focus on the relative short-term effect of the applications observed in the different studies might not come as a surprise to many readers and seems to reflect statements that the assessment of a long-term successful behavior change is unreasonable in the context of HCI research [47]. In this context, it also needs to be considered that some people do not wish to compare with others and that measurement tools ideally need to capture individual personality differences of the participants when assessing social comparison [12]. Additionally, tools such as the "Scale of Social Comparison Orientation" [36] or the scale to measure persuadability [11]—which are deemed suitable for applied research and interventions—lack items that would allow assessing if a participant was influenced *due* to an intervention that included social comparison features. Furthermore, social comparison can also occur in an automated as well as deliberate form [37], which can further complicate measurements. The development of measurement tools that can detect different aspects relevant to social comparison could help clarify design effects.

The *design* of the different social comparison artifacts included numerous applications such as online games, storyboards, dashboards, applications, tests, apps for specific users, chat programs, Google Docs plugins, and virtual shopping baskets. Two papers presented interactive installations taking place in a public space. As part of the installations, digital data was visualized and shared with the audience. We observed a lack of reflection on these different design implementations. This was somewhat unexpected as it is likely that social comparison included in public interactive installations might evoke a different experience than an online application where users can compare themselves with anonymized group performances. We would argue that HCI practitioners can contribute to the multidisciplinary area of BCD, especially in this context, as they engage in developing and reflecting on appropriate design implementations.

The *study context* was as diverse as the designs and included visualizing and influencing several consumption behaviors, shopping habits, climate impact due to food choices, and fostering healthy behavior among expecting mothers. Other study contexts included doing a test, playing an online game, and visualizing language use in an online chat program.

The designs included social comparison in a text-based *medium*, visualization based, and a mix of text and visualization. It remains unclear which kind of medium to exchange social comparison-related information is the most effective one (e.g., raw data, text, visualizations, photos, videos) and if some media simply creates an opportunity for social comparison rather than inducing it in the user [5]. However, design specifications such as medium can influence the way the design is experienced and the kind of responses it evokes in a user [51]. For example, in the context of data visualization, it has been argued that creating aesthetic visualizations rather than purely functional ones could potentially facilitate engagement with the data [38, 57] and help remember data and details easier [7]. In this context, a rift between academic practice and commercial applications has been pointed out, with academic applications primarily focusing on soundness and utility and neglecting the exploration of the aesthetic aspect [57]. Further research is required to determine what kind of design aesthetic and display of the comparison dimension is efficient in evoking the social comparison effect. This would allow broadening our knowledge of *how* designs should ideally display social comparison next to *what* needs to be included.

The *type* of social comparison describes details that were included in relation to the comparison target. Types included details that revealed individually identifiable users or users sharing certain aspects, such as living in the same area or using the same system. Based on the scoped paper, it remains unclear which type of comparison target (e.g., friend vs stranger; individual vs group) might be more efficient. A reflection on this aspect and the inclusion of evaluation tools that help shed light on this aspect could benefit HCI and theory-focused research efforts.

The social comparison *dimensions* entailed consumption-related behaviors, scores in the context of a game or test, reduction of calories, word choice behavior, health-related behavior among expecting mothers, writing behavior online, and climate impact based on one's food choice. These dimensions' effect on one's behavior is likely to depend on the perceived value and attached importance. For example, focusing on reducing climate impact might have little impact on someone who believes that climate change is a hoax.

3.3 Conclusion

As we argued in Section 1, there is no clear description and methodology for the operationalization and grounding process for theory-based BCD. To check whether this issue is also evident in the multidisciplinary areas of HCI and design, we performed a scoping review on the CHI and DIS literature. It seems from the results that we can draw a similar conclusion. The reviewed CHI and DIS literature also miss an explanation of design decisions and shared terminology that would allow comparing different design implementations of social comparison. Arguably, we then do not really know how the designs have operationalized the theory, nor whether the results presented in the different studies relate to theory-based design elements. This stops us from further abstracting or generalizing the meaning of the results.

In answer to our research question (RQ1: How is social comparison theory being applied in designing behavior change technologies?), we can say that the explanation of the operationalization process of social comparison as part of HCI and design literature is very limited. Of course, this does not mean that researchers did not follow a specific process for operationalization. It merely means that it was not clearly described in said papers. This might be due to the challenge of explaining design decisions concisely, the use of theories as a kind of inspiration rather than as concrete guidance in the process, and a lack of a shared "*lingua franca that different parties in research and design can use to point to common referents*" [72, p. 134]. To this end, when we could not find specific information relevant for our review in the papers, we classified it as "unclear" to highlight that *we* could not uncover said data from the paper. Moreover, there was a high diversity of design implementation choices in many of the coded categories.

Taking these results plus the previous results from related work together, it seems that an operationalization process for BCD based on SCT (and perhaps other theories as well) is not explained in sufficient detail, but also that, for the categories that we coded, we cannot detect clear patterns (or a shared language) in the designs examined. This does not mean that the designs are insufficient. Our argument is instead that being explicit about design-related decisions and having a shared language to explain how social comparison is embedded in a design could clarify and ground what is needed—in terms of functional design elements—to evoke a comparison effect. This leads us to look at the problem differently. Since there is a lack of explanation on operationalization practices, a lack of specificity in behavioral theories, and a variety of elements in design examples, can we try to determine what is *minimally* needed for social comparison designs? This leads us to our next question—RQ2: What are the minimal building blocks needed to describe a social comparison design? We address this question in the next section.

4 DEVELOPING A BLUEPRINT

To address our second research question (RQ2: What are the minimal building blocks needed to describe a social comparison design?), we argue that a BCD should contain *functional elements* that need to be embedded in a design to allow the comparison effect to take place. As can be seen from the previously discussed scoping review of HCI and design literature, there are numerous dimensions along which BCD can vary (e.g., in which direction the comparison goes or what the relation to the target of the comparison is). It would be an arduous task to systematize all these dimensions and evaluate their relative value in social comparison design terms. To this end, we argue that we should first focus on what is essential, or in other words, what is simultaneously *necessary* and *sufficient* to allow a person to engage in social comparison. Sufficient, in this case, implies that the elements specified are enough to enable SC to take place (based on a chosen definition of said theory). Necessary implies that the parts are all separately needed to evoke the intended influence. These functional elements, which we refer to in this article as *building blocks*, form the blueprint for SCT.

4.1 Process to Develop a Blueprint for SCT

The process of creating a blueprint was conducted in multiple steps (Figure 5 presents a schematic overview). We began the process by choosing a simple definition of social comparison as our starting point: “System should provide means for comparing performance with the performance of other users” [62, p. 495]. Although this definition might not represent the full extent of the high-level theory that is SCT, it does provide a concrete starting point to discuss whether designs do or do not represent the intended functionally minimal definition. Additionally, this definition focuses explicitly on an implementation version of social comparison.

After defining our theoretical basis, our process continued by reviewing the design examples we included in the scoping review. Initially, we considered using an established IMLK as a source of guidance to structure our approach—specifically the notion of “patterns,” which arguably is very close to blueprints. Patterns are based on best practice examples [2, 10, 84]. However, in the scoping review, we could not determine if any of the included examples could be labeled a “best” example to operationalize social comparison as essential details about the process, design decisions, and effect of the incorporated social comparison feature were often missing. Furthermore, as indicated previously, many of the design specifications (e.g., medium of the design) were not reflected on, leaving the rationale needed for developing a pattern impossible to determine. As part of the blueprint process, we started instead to identify essential building blocks needed for social comparison based on scoping review examples and the theoretical definition we had chosen.

While reviewing the scoped design examples, we recognized that even though the scoped design examples showed significant diversity in implementation and refinement, certain functional aspects were evident in most of the designs that also seemed theoretically necessary. For example, the different designs included a comparison target and details about the comparison dimension to allow the user to engage in social comparison. With this reflection in mind, we focused on the first functional elements required for SCT in a design. We used Post-its to describe these functional elements or building blocks we noticed during this step (Figure 6). We then applied the building

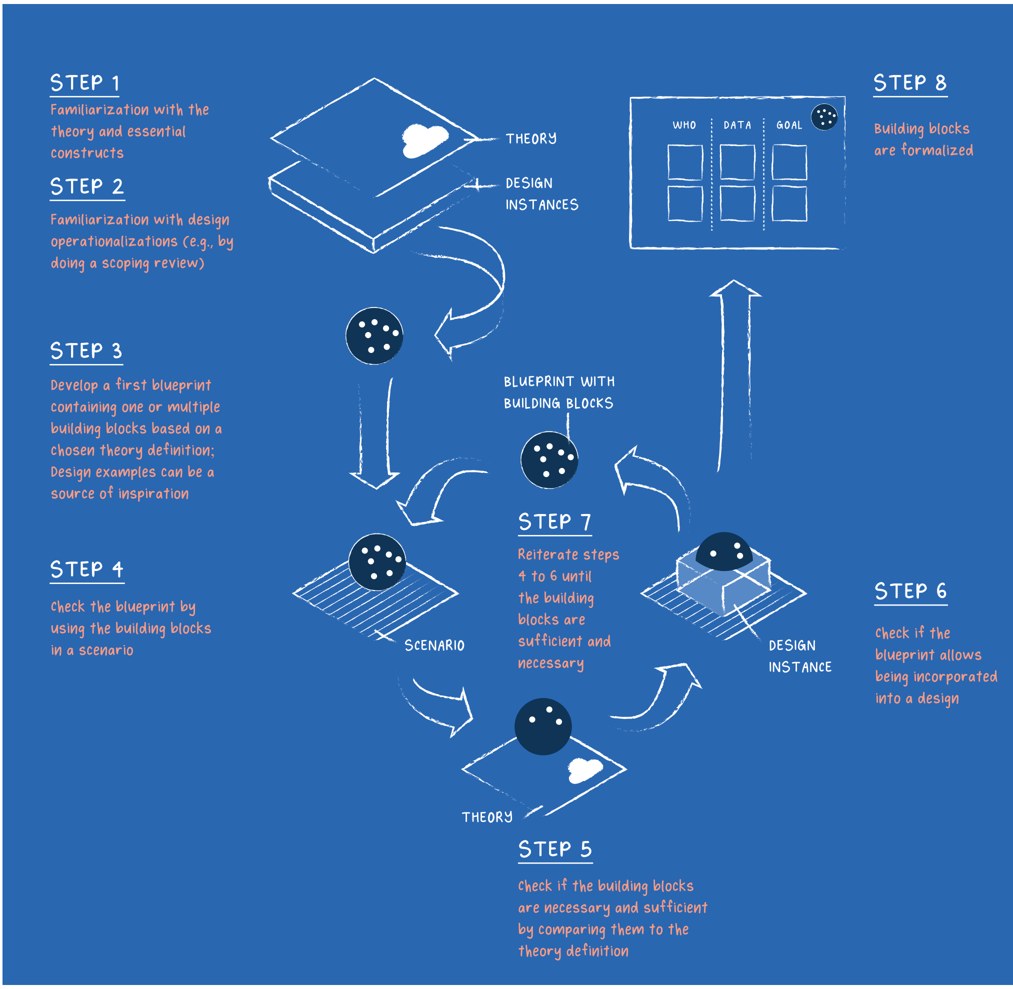

Fig. 5. The iterative development process that was followed to create a blueprint.

blocks to several scenarios to evaluate whether the functional components are necessary as well as sufficient in those scenarios. In a consequent step, we check to see if the blueprint still aligns with the theory definition ("System should provide means for comparing performance with the performance of other users"). We then added additional building blocks based on the reflection (Figure 7; see Figure 6). As an example, with Figures 6 and 7, we use the scenario of weight loss to test some of our initial ideas of necessary building blocks and the conceptual blueprint we imagine (e.g., datum + identifier + comparison dimension = blueprint, tested through Post-its A, B, C). Initial evaluations show that the building blocks (and therefore blueprint) need refinement, and we update our idea for the blueprint (see Figure 6). Consequently, we added new building blocks (see the second row in Figure 6) and checked them against the theory again. We again apply this blueprint to scenarios. Figure 7 incrementally tests the necessity of each building block. Post-its A through F illustrate some of the iterations we went through as part of the process.

The process of adding and removing different building blocks was repeated until a point where the functional elements were deemed sufficient and necessary (see Figure 5). We then applied the

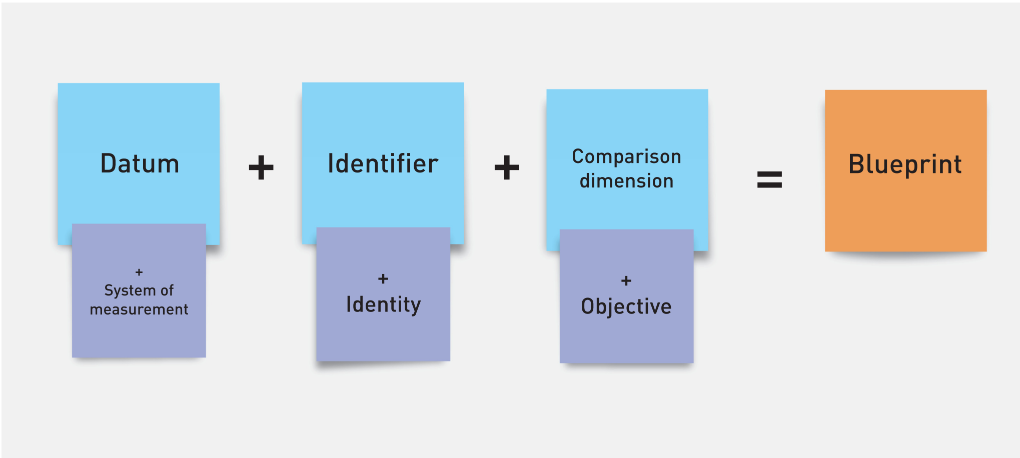

Fig. 6. The building blocks for the blueprints were developed in an iterative manner. For example, we initially used the three building blocks of “datum,” “identifier,” and “comparison dimension.” In consequent steps, we added building blocks as part of the reflection and refinement process. The names of the building blocks displayed in the figure were interim names that were refined in the final step of the process.

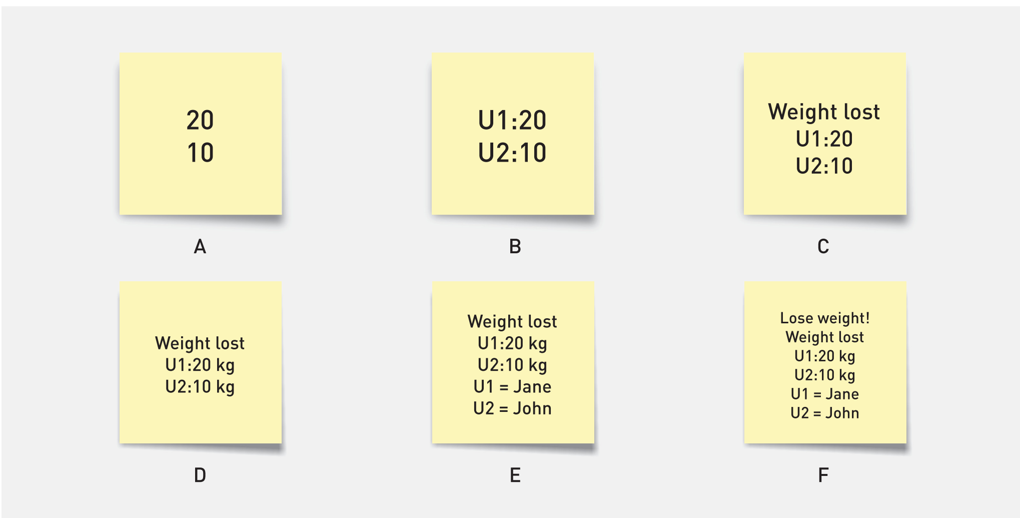

Fig. 7. We applied the building blocks to different scenarios to refine them and evaluate the building blocks, as illustrated here with an example of a BCD to encourage weight loss. Post-it A ‘tests’ the building block “datum.” Post-it B adds the building block “identifier.” Post-it C adds “comparison dimension.” With Post-its D, E, and F, we explored the need for additional building blocks such as “system of measurement,” “identity,” and “objective.”

building blocks to additional scenarios. For this purpose, we also created a visual prototype to evaluate if the blueprint for SCT we had conceptualized could be embedded into a design. This step was repeated for different scenarios as well (e.g., supporting weight loss, increasing online learning performance, reducing carbon footprint). We concluded the process once we perceived all the building blocks to be *necessary* and *sufficient*.

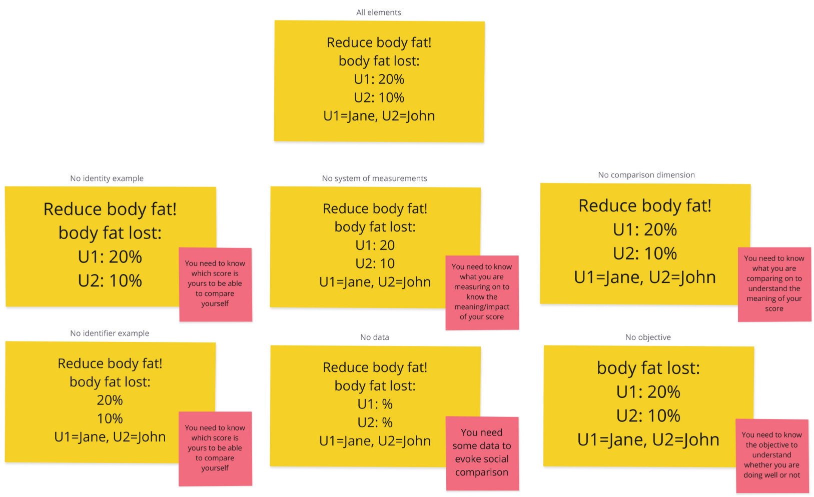

Fig. 8. The six building blocks of the social comparison blueprint checked for necessity.

4.2 Elements of the SCT Blueprint

The first essential building block for social comparison we set was the need for a *datum* (which in the final version of our blueprint we labeled "data points"), which is, in its most minimalist version, two data points to allow for comparison (see Figure 7(a)). The following essential building block we determined was *identifiers*, to link the datum to a user (see Figure 7(b)). The third essential building block we determined was some form of a goal, later termed *comparison dimension* (see Figure 7(c)), to link the data to a goal.

When we assessed the design's (see Figure 7(c)) intended effect, we noticed that such a design allows the user to compare, but the data is still uninterpretable. A design in this state leaves unclear what unit is used for these scores. For example, simply stating 20 as a data point as part of the design leaves it unclear how the value was measured (e.g., kilograms or miles). Moreover, it is still unclear who the two identifiers (i.e., U1 and U2 next to the data points) relate to. Therefore, the subsequent two building blocks added were a clear indication of the *system of measurement* (later labeled "unit") to allow the user to interpret the data points (see Figure 7(d)) and an *identity* for the two identifiers (e.g., clearly stating which identifier is the user, see Figure 7(d) and (e)).

Reflecting on this iteration (see Figure 7(e)), we argued that the user could engage in basic comparison. Still, the anticipated behavior change goal is left unclear. Using social comparison to induce a behavior change as part of a design needs to make these components explicit to guide and influence the user's behavior in an intended way. We therefore differentiate in this article between *basic social comparison* and *social comparison for behavior change*. To facilitate behavior change as part of BCD, we added the additional building block *objective* (see Figure 7(f)). In this way, the user knows the dimension of the data (e.g., the context of the 20 kg vs 10 kg) and the overall objective.

To check whether these building blocks are all truly necessary, we separately left each one out to see if the blueprint breaks down (Figure 8). However, if you leave out *identity*, it is unclear who the scores belong to. If you leave out *identifier*, it is also unclear who the score belongs to. If you leave out *system of measurement*, it is unclear what the score means, and if you leave out *data*, then you have no score to look at. These building blocks all seem clearly necessary. The last two building blocks are a bit more complicated. If you leave out the *comparison dimension*, the example might work, but it could very well be that you misunderstand the meaning of the score (see Figure 8).

Fig. 9. Our finalized social comparison blueprint for behavior change consists of six building blocks.

You might think the score is about the kilos you lost, but the score actually represents the kilos you still have to lose. Similarly, if you leave out the *objective*, the example could also work, but you might think the goal is to lose weight when instead the goal was to gain weight.

Arguably some building blocks were easier to identify than others. For example, *identifier, identity*, and *data* were quickly detectable building blocks for us, whereas the need for *unit* and *comparison dimensions* became apparent within the first iterations of the building blocks. In retrospect, the separation of the building blocks of *objective* and *comparison dimension* was the most difficult one to identify. This was due to many of the examples we scoped missing this aspect as a clear statement as part of the BCD and our own bias in the process. For example, when applying the blueprint to a scenario of *comparing* weight, we discussed how future users of such a system would use it and if *losing* weight should always be seen as the ideal behavioral outcome. It is only when you tease out the details that you realize both blocks are necessary to specify (see Figure 8). We concluded the process by formalizing the different elements and clustering them into three main domains (Figure 9).

Based on the results of the steps presented, we can address RQ2 (What are the minimal building blocks needed to describe a social comparison design?). Our final blueprint consists of six building blocks covering the areas of *who* is involved in social comparison by the two blocks of *identity* and *identifier*, the area of what *data* is used to allow comparison by the two blocks of *data points* and *unit*, and finally the area of the overall *goal* the design seems to aim at by the two building blocks of *comparison dimension* and *objective*. These minimal building blocks all adhere to the simple definition of social comparison that we used as a starting point ("System should provide means for comparing performance with the performance of other users").

4.3 Systematic Steps for a Generic Blueprint

Based on our process of creating a blueprint for SCT, we abstracted the following essential steps. This list of steps is not prescriptive to allow for iterative development and flexibility in application:

- *Step 1*: Familiarize oneself with the theory and essential constructs. Decide what kind of definition of the theory the blueprint will be based on, in case of a variety of definitions such as with SCT (e.g., conduct a literature review on the theory and essential constructs).
- *Step 2*: Familiarize oneself with the designs and different forms of operationalizations of the theory and essential constructs (e.g., conduct a scoping or critical literature review on designed artifacts and different forms of operationalizations, such as in Section 3 of this article).
- *Step 3*: Develop a first blueprint containing one or multiple building blocks. Scoped design examples can be used as a source of inspiration in the process. The building blocks can at this stage be purely text based (see Figure 6).
- *Step 4*: Check if the blueprint allows the function described in the theory definition by using the building blocks in the context of a scenario (e.g., to test if the future user has enough information to perform the ideal behavior (see Figure 7)).
- *Step 5*: Iterate on building blocks while focusing on the minimal (i.e., necessary and sufficient) elements needed to include the theory. This can include adding or removing building blocks to the blueprint (see Figure 6).
- *Step 6*: Check if the blueprint allows being incorporated into a design, for example, by creating a visual layout for a specific scenario.
- *Step 7*: Reiterate steps 4 through 6 until the minimal and essential building blocks remain stable and need no further adjustment. Extrapolate the blueprint to other scenarios of use to check if it holds up to other imaginable examples (see Figure 8).
- *Step 8*: Formalize the different building block elements by naming and defining them (see Figure 9).

5 APPLYING A BLUEPRINT

This section illustrates how researchers can use a blueprint to evaluate the included functional elements and points out additional useful and design-specific features. For this, we apply our developed SCT blueprint to one example of the scoping review and additional *fictional* examples as we created multiple BCD concepts to refine the blueprint. The fictional examples are created to highlight realistic but lacking (in terms of the blueprint) designs and are based on the reviewed designs. In the process, we noticed that some scenarios might be prone to strong normative influences where the explicit statement of an objective might not seem necessary at first sight. For example, developing a design that compares weights between users is likely to be perceived to encourage weight loss rather than gaining weight. Stating an explicit objective clarifying the goal of the BCD might not be seen as necessary in this context. Nevertheless, it is still essential to precisely include the objective because norms are culture dependent and subjective to change. For example, instead of the expected "lose weight," the objective in this context might as well be "do not lose too much weight."

We start with an example included in our scoping review. The public interactive installation *Reveal-it!* allows users to input their energy consumption data and facilitate comparison between individuals and communities [82] (Figure 10). The interface developed as part of the installation asked participants to add data relating to the monthly energy costs, the neighborhood where they were living, and the number of people living in the household. The installation assigned different colors to different communities, and the proportion of a particular color/neighborhood was based on the relative participation rate of said area. The installation included interactive and animated features, such as adding newly added data from visitors and animating the sunburst diagram elements. For example, new data entries were displayed in white to contrast them from the rest and smoothly adjusted to the color of the respective neighborhood.

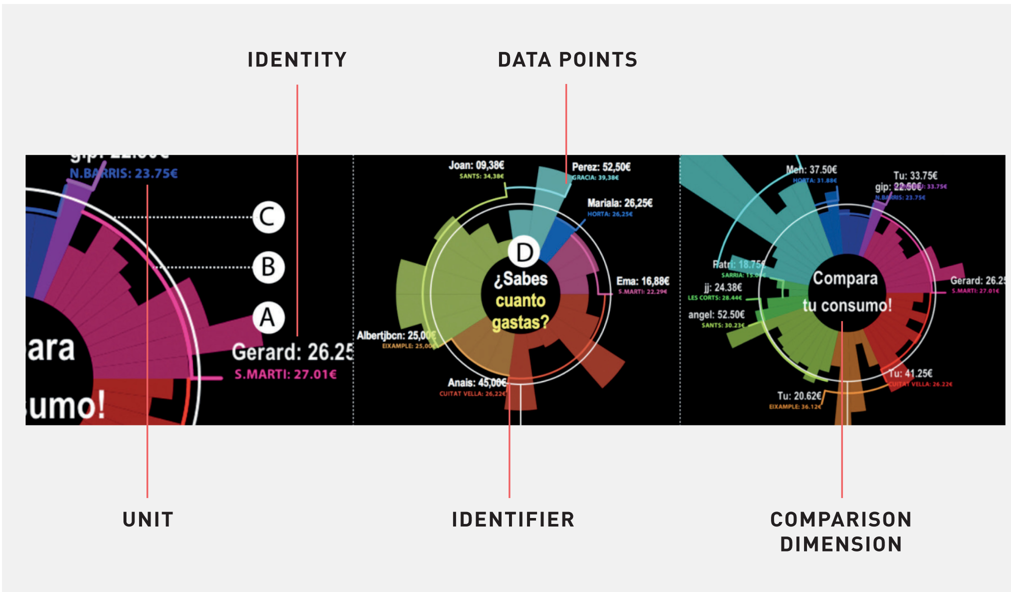

Fig. 10. One of the examples we analyzed as part of our scoping review was the interactive installation by Valkanova et al. [82] allowing visitors to compare their energy cost with others. © Valkanova et al. [82].

Figure 10 indicates the building blocks of our blueprint included in the design example. All building blocks are represented, except the objective which seems implicitly present. After all, the most common way to use energy consumption data is to reduce consumption. We also noticed that the design included further functional, interactive, and visual aesthetic additions. For example, users of the system could compare their spending not just to one comparison target but multiple, including individuals in their neighborhood, other individuals in different neighborhoods, and the average consumption within a particular area. The sunburst visualization also included multiple types of a medium of the comparison dimension (energy consumption). For example, the visualization included different bar graphs that visualized the amount spent on energy and the numerical value at the end of the bar. The installation, furthermore, provided persuasive messaging inviting visitors to add their data to the system. Evaluating and reflecting on these additional features and the overall quality of the design could help clarify not just *if* SC can be evoked but also *how* these additional features increase or decrease the intended effect. For example, a study conducted in the context of designing a hospital environment indicated that the visual quality of a design can evoke a kind of placebo effect. The authors of the study conclude that people seem to anticipate the quality of a treatment based on the quality of the design [71]. Further exploration is required as to if such an effect can be replicated in the context of creating BCD. Inspiration in this context could be sought from the conceptual framework of appropriateness in regard to aesthetic, moral, and contextual factors in the context of BCD [83], as well as general reflections on aesthetic and design [40] and aesthetics in interaction [68].

Our second example is based on the use of a visual metaphor as part of the social comparison design. Figure 11 shows one of the design concepts we developed. The concept aims to allow students to compare their learning performance. The octopuses represent the different students using the system, and their size is an indicator of their learning performance—the more significant the size, the better the performance. We applied the blueprint to check which building blocks are discernible in the design. The building block of "identifier," "identity," and "data points" can be found

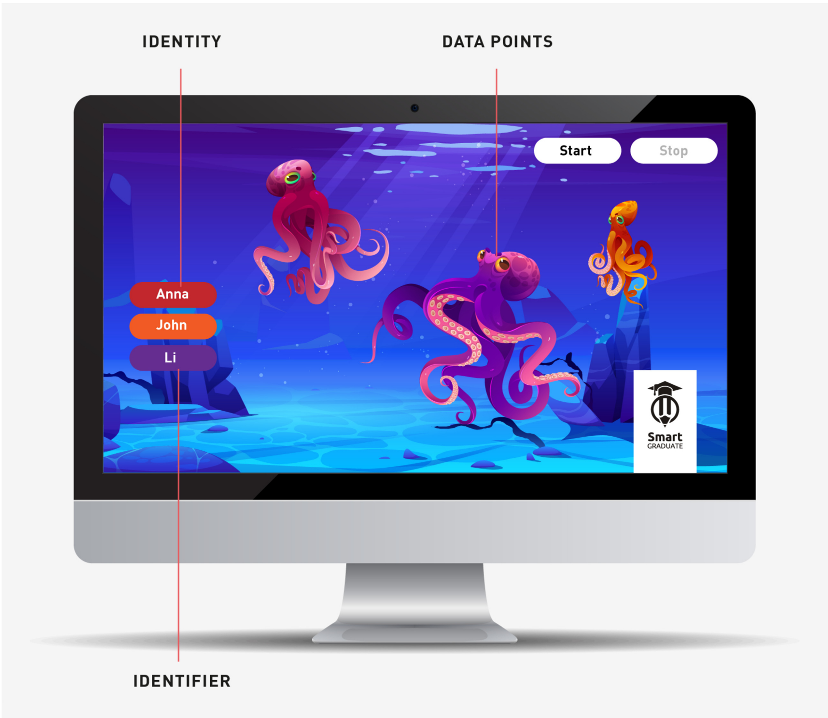

Fig. 11. One of our fictional design concepts used a metaphor to allow users to compare their learning performance with other students. © Upklyak and Enola99D.

on the screen; however, the additional three building blocks are not clearly recognizable. For example, in its current state, the design leaves unclear what precisely the user can compare themselves on, how this data has been measured (i.e., how exactly the learning performance is determined), and what the overall objective is (i.e., if the user should increase learning performance).

Another example that we conceptualized for a similar application area allows students to compare their study performance with fellow classmates (Figure 12). We chose for this example a table format resembling a leaderboard, which is a common operationalization of SCT. Applying the blueprint to the design revealed that multiple aspects were missing—for example, while two data points allow more than one comparison (score yesterday and score today), it remains unclear how this score was precisely determined. Furthermore, while the lack of a clear objective might not affect the users' opportunity to compare themselves with other users, it can be argued that this lack affects the intended behavior change effect—users can compare themselves, but there is no overall goal directing this. Similar to other designs, we also used color to indicate the average group value and data relating to the specific user of the system. When we reflected on the choice of color, we realized that our color choice of using red for the average group could also be interpreted as a warning indicator or a "bad" value. Such design choices are likely to affect the intended SC effect and should be carefully considered.

The second and third examples might seem like they have obvious shortcomings at first glance. However, when we looked at different examples mentioned in the work of Lemke and de Vries [49] and our scoped examples, we noticed multiple instances that lacked such elements. In some cases, such information was provided in the text, accompanying the testing instructions or the paper. Whatever the circumstances, for the purpose of abstracting knowledge and systematizing

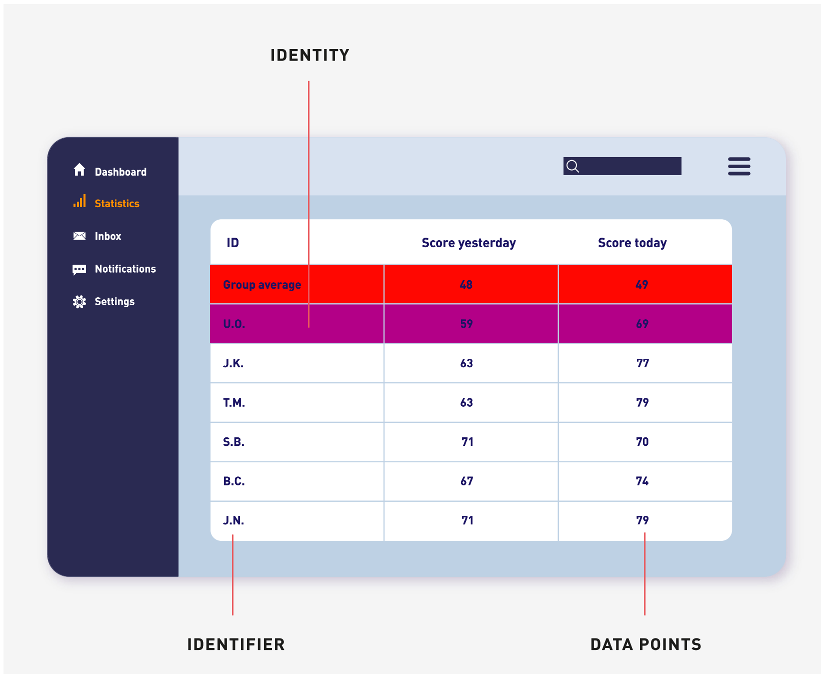

Fig. 12. We chose to use a leaderboard format to allow users to engage in social comparison for this scenario. As part of the design, group averages and individual scores are displayed. © Freepik.

the operationalization process, the lack of details on these things is detrimental. If the (functional) use of the design is not clear from the paper, nor from the design, it is hard to determine what certain elements actually do or do not contribute to the design.

Our blueprint's building blocks can help reflect on a specific design, but they are not necessarily a recipe for success in terms of that the BCD will evoke a comparison effect in the user. On the one hand, this effect depends on many contextual factors at play in this context (see Figure 3), and, on the other hand, also on the way the design is perceived by the user, including, for example, its aesthetic and interactive qualities.

Based on applying the blueprint as part of the development process to create different fictional BCDs and critically analyzing the scoped design examples, we observed multiple extensions that designs can incorporate—for instance, aesthetic as well as interactive elements. This is by no means an extensive overview of what design can add in this context but rather a starting point for future reflections and studies. We noted in the process multiple additions and extensions in the design examples that went beyond our blueprint. Three of these additions were related to the building blocks of our blueprint, including the inclusion of multiple *identifiers*, *identities*, and *comparison dimensions*. For example, designs included details relating to the comparison target, such as revealing a close spatial or personal relationship of the comparison target to the user (e.g., the target is a friend or a neighbor). Another extension related to comparison dimensions included allowing users to compare various factors as part of a design—for example, by presenting longitudinal data or presenting an average performance and the specific performance of the day next to each other.

We also noticed some additional design and interactive elements not directly related to our blueprint in the design examples we scoped. These included providing different persuasive aspects such as inviting messages as part of the system or evaluating the data of the user—for example, by displaying data as "your best score." Other elements included elements of self-evaluation by demonstrating how a specific behavior would influence one's future self (e.g., eating this amount of calories will contribute to weight gain). Furthermore, some concepts included interactive elements that allowed users to add personal data and see the data change over time. Visual aspects of a design can also facilitate the combination of multiple forms of data representation, such as displaying data points as numbers and charts next to each other. Symbols and colors can also be used as normative elements to emphasize a specific objective—for example, including smileys as part of a design or using colors with a particular connotation (e.g., green for good and red for bad).

6 DISCUSSION

Theory in the context of HCI seems to take on multiple roles, including descriptive, explanatory, predictive, prescriptive, and generative functions [9]. A recent review of the 25 best CHI papers shows that the use of theory is mentioned in half of the papers but often seems to play just a minor role in the design process [66]. This might be due to the papers' focus on designing novel technology and on exploring *if* a design works rather than *why* it works [47, 66]. Furthermore, the use of theory seems to be mainly focused on post hoc analyses and is primarily done *after* the design is completed [66]. This observation can be explained by several factors, such as that HCI research seeks inspiration rather than concrete guidance from other disciplines (e.g., psychology or social sciences), causing some confusion about what a theory and its applied nature is [41, 85] or the fact that designers often prefer to rely on their intuition rather than theory [72].

In this article we have explored the *operationalization process* of SCT in BCD. The creation of BCDs differs from most other design circumstances, which tend to be quite "underdetermined" and leave plenty of room for interpretation. Ideally, with theory-based designs such as BCDs, an inference loop is established where theory can help guide and ground the process of designing a prototype, and in turn, the evaluation of the design prototype—that is hopefully designed with reasonable fidelity—can help to evaluate and inform the theory. However, using a theory as part of a BCD seems to take place in a kind of metaphorical black box signified by a gap between applied design and theory. To this end, our first research question (RQ1: How is social comparison theory being applied in designing behavior change technologies?) aimed at finding out more about (the reporting of) the operationalization process of SCT. Based on a scoping review of the CHI and DIS literature we performed, we can say that explanations of the operationalization process of social comparison as part of HCI and design literature are very limited. This does not mean that a certain process for operationalization was not followed, only that the authors did not clearly describe it in said papers. We observed similar problems as others previously [4, 49] in regard to the lack of details explaining the steps taken to use the theory as part of the design. This makes it hard for this type of research to contribute beyond the exact scope of the study. We believe that a shared terminology to describe design decisions could facilitate the systematic grounding of design elements in a specific theory. This, in turn, would allow researchers to replicate or reproduce relevant components and open up the design space for the formation of more abstract knowledge.

As the scoping review points out, there are numerous dimensions along which BCDs can vary—for example, if the data is presented as a text or visual or if the prototype is made in a high or low-fidelity form. It would be an arduous task to systematize all these dimensions and evaluate their relative value in social comparison design terms. Therefore, we argued that we should first focus on what is essential from a functional perspective to incorporate a theory in a design, or in other words, what is simultaneously *necessary* and *sufficient*. These functional elements, which we refer to in this article as *building blocks*, allow us to create minimal functional design instances of a behavior change theory.

This leads to our second research question (RQ2: What are the minimal building blocks needed to describe a social comparison design?). Through an iterative process (explained in Section 4), in which we gained theoretical as well as practical knowledge on SCT and its operationalizations in the literature, we determined six essential building blocks. The six building blocks cover the areas of *who* is involved in comparison by the two building blocks of *identity* and *identifier*, the area of what *data* is used to allow comparison by the two building blocks of *data points* and *unit*, and finally the area of the overall *goal* the design seems to aim at by the two building blocks of *comparison dimension* and *objective*. We call this collection of building blocks a *blueprint* of social comparison for behavior change, which we see as a *lingua franca* for this specific behavior change theory to facilitate an explanation of design decisions and reflections on empirical findings. Moreover, building blocks can be exchanged or serve as inspiration between different blueprints, in a similar way that constructs between behavior change theories overlap. For example, the concept of self-efficacy is the basis of many behavior change theories and models and also has similar overlapping constructs such as perceived behavioral control and (internal) locus of control. A building block, a collection of building blocks, or even a full blueprint of self-efficacy could be transferable to theories with similar constructs. Potentially, the applicability and transferability of the concept of the blueprint are broad. If we take, for example, another straightforward and functional definition of a relatively complex and abstract theory, such as goal-setting theory, we can imagine some possible building blocks. For example, consider this definition of Michie et al. [56, p. 7] of goal-setting theory as a behavior change technique: "Set or agree on a goal defined in terms of the behavior to be achieved." One can imagine that a minimal building block for a blueprint of goal-setting theory is something like a "data-entry block," a place where the user can "set or agree on a goal." In practice, this could be as simple as a text box. Another building block that seems necessary to convey is that of a "goal," in terms of behavior that the user wants to achieve. Now, is this enough to sufficiently and necessarily cover goal-setting theory? Probably not, but we imagine one could come to a blueprint version by applying our process. Nevertheless, with theories and their "reduction" geared toward implementation, the balance between theoretical complexity and functional specificity should be carefully considered. We believe that the notion of a blueprint has the potential to be a new form of IMLK and a tool to facilitate the inference loop between behavior change theory and practice (see Figure 2).

6.1 Blueprint Characteristics

We see a blueprint as a kind of model of the minimal functional elements embedded in a design using a specific theory with reasonable theoretical fidelity (for the distinction between theory and model, we argue that a model is more of a "heuristic" representation of a theory [6]). The model, or blueprint, specifies which building blocks need to be minimally included if the goal is to represent the theory. The blueprint would require the designer to include certain functional elements that can be designed in many ways. For example, a building block for a theory could be that the system's actors need to be clearly distinguishable. However, how these actors (e.g., different users of the system) are portrayed is the designer's choice. Design considerations in this context can include the use of actual photos, names, or the development of avatars. Although the idea of functional building blocks required for the functional elements of a specific design instance might seem like a limitation to some in what is possible to create [34], it could, in turn, facilitate the theoretical fidelity with which this design hopefully represents the theory that would serve the inference process [66]. We see several important implications of this new form of IMLK to contribute to the HCI design research community.

First, blueprints are *prescriptive* and provide concrete guidance as part of the design process [9, p. 350] as they define essential functional elements named *building blocks* which are needed for a minimal functional design version of a particular theory definition. In this regard, they are similar to the notion of patterns [10, 24] and bridging concepts [22] that also provide concrete examples but focus on the implementation of functional elements already embedded in a design.

Second, blueprints can facilitate some *predictive power*. Behavioral theories are sometimes considered not sufficiently specific enough [61]. This lack of detail comes with additional inferences problems, such as type III errors (concluding a hypothesis is false when it was never tested), or conversely, the lack of detail in the taken approach combined with already a very 'general' theory will mean that neither will ever be falsifiable [41]. This is regrettable for the following reason: "A good theory should consist of constructs that are sufficiently specific so as to generate hypotheses. Such hypotheses should be testable, and, in principle at least, a good theory should be able to be rejected" [61, p. 425]. It should be noted that such a definition of a theory clearly differs from one used in social sciences. But even in this field where the terms *theory*, *theoretical*, and *theorizing* are constantly used, the essence and implications of a theory can differ significantly [1]. We see a blueprint as a kind of minimal functional implementation model of a theory with reasonable theoretical fidelity, specifying which building blocks need to be minimally included. In that way, blueprints could have two different functions. On the one hand, an initial blueprint could function as a kind of "falsifiable" hypothesis [61, 66] in terms of "this is what we think is minimally functionally needed for a theory-based design in this context.". On the other hand, after testing the blueprint and validating its functional aspects, the "established" blueprint could function more descriptively and constructively, specifying which elements have proven to be sufficient and necessary. In this way, blueprints themselves can also have some predictive power in that they can generate falsifiable hypotheses about the building blocks they argue are essential. This is an aspect that seems to be lacking in current tools and methods that work as IMLK. However, a blueprint is unlikely to have strong predictive power because it might be that a specific blueprint is not a good representation of theory or that the theory itself does not account for a lot of observed variance [41]. Blueprints can facilitate behavioral theories to have more predictive power as blueprints can function as falsifiable hypotheses. Of course, one could still always argue that the blueprint is not a good representation of the theory, and therefore never reject the theory, but only always the blueprint, similar to the current situation. However, we would argue that the blueprints would close the theory-practice gap a bit more because they reside closer to theories than prototypes, facilitate the discussion on representation in the design of certain constructs, and could function as a proxy for a falsifiable theory. In this context, the specific definition of a theory might influence the specifications of a blueprint. Moreover, some of the theories of behavior change, and theories, in general, are not falsifiable because their constructs are unspecific. In this case, a blueprint can offer at least a shared language of terms, conceptual elements, and interactions.

Third, blueprints are *constructive* or facilitate the constructive power of a theory. A theory is said to have constructive power, as introduced by Oulasvirta and Hornbæk [66], if it facilitates the construction of artifacts. For behavior change theories, the outcome wanted is usually a behavior change. Blueprints facilitate this by specifying the functional elements required for a design to have this impact. These required elements are a limitation not everyone thinks designers should have [34], but it hopefully facilitates the theoretical fidelity with which this design represents the theory [66]. Furthermore, perhaps counter-intuitively, we think that because the blueprint is solely focused on the functional specifications, it could also serve as *generative* inspiration. For example, due to the way a blueprint is formulated (e.g., a functional equation), the design space is potentially more open than it would be if you start from an existing design, with which you can suffer from design fixation. With a blueprint, you reduce the risk of design fixation, which likely occurs in the context of other IMLK where best practice examples are a core element. Additionally, the process of matching the blueprint to available designs (as we did in Section 5) can produce many extensions to the building blocks, which allows practitioners and researchers to add and combine elements and develop extensions that potentially increase the behavior change effect. This can lead to designs that have much more of a significant impact.

Fourth, blueprints are also *generic* and *situated* and lack prototype-specific aspects that can play a role and be a specific focus in the process such as the "look and feel," "implementation," and "role" [45], as well as design manifestation specific features such as material, resolution, and scope [51]. The blueprint instead outlines functional elements to secure the intended behavior change effect but leaves room for creative integration of the blueprint elements. Blueprints reside on an abstraction level above particular instances, which means they can take on many different design forms. Blueprints can be a practical tool to evaluate design prototypes and systematically assess and compare different prototypes, similar to the concept of generic design thinking approach [88]. The use of blueprint elements can highlight and evaluate similarities and unique features of social comparison design elements. The blueprint does not specify the "optimal" representation of theory but only what is necessary and sufficient. Additional building blocks could very well be specified to create more effective representations of theory and might therefore go beyond the *sufficient* aspect of blueprints creating more *refined blueprints*. *Situated blueprints* can be developed for differing circumstances. Iterations over blueprints would happen as old blueprints are rejected. And, after variations of blueprints would have come to pass, this could be taken as that the specific theory that has served as a basis should also be rejected, or updated, to get closer and closer to theories that are useful *and* testable.

6.2 Limitations

"Designers focus on the creation of artifacts through a process of disciplined imagination, because artifacts they make both reveal and become embodiments of possible futures" [90, p. 42]. From this point of view, the blueprint and its building blocks could be seen as a limitation in what is possible to create [34]. However, the focus on the functional aspects of a specific theory still leaves room for exploration and different design manifestations and aims to facilitate the theoretical fidelity with which this design hopefully represents the theory that would serve the inference process [66]. However, we acknowledge that a thorough description of complex design choices and implications is likely to be complicated by factors such as the complexity of a particular behavior change theory, user characteristics, and a specific socio-economic context [41, 47], as well as time and financial constraints and different research approaches. Furthermore, the exact influence of a behavior change theory and the blueprint might vary depending on the context of use and the user engaging with the BCD.

It needs to be considered that the implementation and effect of a specific behavior change theory can depend on several contextual factors, potentially increasing or decreasing the intended behavior change effect. For example, the experience and effect of a specific BCD can depend on the socio-economic context or user profile (e.g., being an introvert or extrovert, competitive or collaborative). As such, a blueprint cannot claim to be universal but can hope to be applicable to a certain range of situations or contexts. In addition, the quality and usefulness of a blueprint such as the one we outline in this work is dependent on the theory and its definition. Since there are many different behavioral theories that vary in terms of constructs and relationship [39], there are also many different blueprints. Potentially this might also be true for the same theory which might change depending on a specific context, or the theory itself might evolve and change over time based on empirical findings.

It could be argued that models could take on the role that we assign a blueprint. We would argue, however, that blueprints are geared toward implementation, and a model is for theoretical explanation. In this context, we also acknowledge that some theories might be easier to develop a blueprint for than others. For example, the Transtheoretical Model of Behavior Change proposes that behavior change takes place over different stages—usually five stages. The progress through these stages can also be based on a spiral pattern rather than a linear one. The model also describes 10 different processes of change that can work as a stimulant for an individual to process through the various stages [70]. Defining building blocks and a blueprint for such a multilayered theory will clearly be a challenge. However, it is our hope that specific building blocks or even complete blueprints of one specific theory could be transferred to the context of another behavior change theory if both of them share similar theoretical components.

Last, this article proposes a new tool—the blueprint—to facilitate the exchange between behavior change theory and practice, specifically developed in the context of SCT. We used the blueprint to develop different design concepts for SCT; however, more evaluations are needed not only for this specific blueprint of SCT but also for the idea of blueprints in general. We especially see a mix of novice and expert users contributing to the conversation around using such a tool for design purposes.

7 CONCLUSION

Ideally, with theory-based interventions, inference loops are established where theories can help guide and ground the design process, and in turn, evaluations of the design can help evaluate and inform the theories. However, using theories such as SCT as part of a BCD seems to take place in a kind of black box. Consequently, these designs cannot help in evaluating or informing the theories, making it hard for HCI research in this domain to contribute and evaluate that contribution beyond the exact scope of that study. To this end, this article focused on SCT and used (1) a scoping review of current operationalization practices of this theory, (2) an attempt at systematizing functional design for social comparison in terms of a new form of IMLK we call *blueprints*, and (3) a look at social comparison based design artifacts through the lens of our new blueprint. With this work, we aim to make a theoretical contribution to the field of BCD by introducing the concept of blueprints for SCT in particular and as a potential new form of IMLK for future studies.

In concluding this work, we return to our starting point, which illustrated the need to develop a systematic approach to justify and develop design solutions that effectively incorporate theory. Making the process clear allows justification of the level of rigor applied to the methods and rationale for decisions made [91]. We acknowledge the importance of the development of design prototypes as specific "instances" [44], or "ultimate particulars" [75]. Still, we also need tools that help bridge the area between theory and these particular instances to clarify and abstract research findings and overall contributions. As many studies use or adapt constructs from more than one theory, the operationalization process so far seems to rely on the designer's discretion [64]. Our argument is not that these design decisions are wrong but that there is a need to make decisions explicit, specifically in fields where one would like to do inferencing between theory and instance. As Oulasvirta and Hornbæk [66, p. 13] point out in the context of counterfactual thinking: "This does not mean rejecting the reality and value of designerly ways of knowing; [i]t means that in order to build a bridge to theory-formation and empirical research, design cannot treat theories as mere opinions and inspirations." We would like to believe that blueprints can facilitate this.

ACKNOWLEDGMENTS

We would like to thank Rick Schifferstein for providing feedback on an earlier version of this manuscript.

REFERENCES

[1] Gabriel Abend. 2008. The meaning of 'theory.' *Sociological Theory* 26, 2 (2008), 173–199.

[2] Christopher Alexander, Sara Ishikawa, Murray Silverstein, Max Jacobson, Ingrid Fiksdah-King, and Shlomo Angel. 1977. *A Pattern Language: Towns, Buildings, Construction.* Oxford University Press.
[3] Danielle Arigo, Megan M. Brown, Kristen Pasko, and Jerry Suls. 2020. Social comparison features in physical activity promotion apps: Scoping meta-review. *Journal of Medical Internet Research* 22, 3 (2020), e15642.
[4] Danielle Arigo, Jacqueline A. Mogle, Megan M. Brown, Kristen Pasko, Laura Travers, Logan Sweeder, and Joshua M. Smyth. 2020. Methods to assess social comparison processes within persons in daily life: A scoping review. *Frontiers in Psychology* 10 (2020), 2909.
[5] Danielle Arigo and Jerry M. Suls. 2018. Smartphone apps providing social comparison for health behavior change: A need for better tailoring to person and context. *Mhealth* 4 (2018), 46.
[6] L. Kay Bartholomew and Patricia Dolan Mullen. 2011. Five roles for using theory and evidence in the design and testing of behavior change interventions. *Journal of Public Health Dentistry* 71 (2011), S20–S33.
[7] Scott Bateman, Regan L. Mandryk, Carl Gutwin, Aaron Genest, David McDine, and Christopher Brooks. 2010. Useful junk? The effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 2573–2582.
[8] Jordan Beck and Hamid R. Ekbia. 2018. The theory-practice gap as generative metaphor. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* 1–11.
[9] Benjamin B. Bederson and Ben Shneiderman (Eds.). 2003. *The Craft of Information Visualization: Readings and Reflections.* Morgan Kaufmann.
[10] Jan O. Borchers. 2008. A pattern approach to interaction design. In *Cognition, Communication and Interaction.* Springer, 114–131.
[11] Marc Busch, Johann Schrammel, and Manfred Tscheligi. 2013. Personalized persuasive technology—Development and validation of scales for measuring persuadability. In *Proceedings of the International Conference on Persuasive Technology.* 33–38.
[12] Abraham P. Buunk and Frederick X. Gibbons. 2007. Social comparison: The end of a theory and the emergence of a field. *Organizational Behavior and Human Decision Processes* 102, 1 (2007), 3–21.
[13] Philip J. Cash. 2018. Developing theory-driven design research. *Design Studies* 56 (2018), 84–119.
[14] Philip J. Cash, Charlotte Gram Hartlev, and Christine Boysen Durazo. 2017. Behavioural design: A process for integrating behaviour change and design. *Design Studies* 48 (2017), 96–128.
[15] C. K. Bernard Choi and A. W. Anita Pak. 2006. Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness. *Clinical and Investigative Medicine* 29, 6 (2006), 351–364.
[16] Heather Cole-Lewis and Trace Kershaw. 2010. Text messaging as a tool for behavior change in disease prevention and management.*Epidemiologic Reviews* 32, 1 (April 2010), 56–69. https://doi.org/10.1093/epirev/mxq004
[17] Lucas Colusso, Cynthia L. Bennett, Gary Hsieh, and Sean A. Munson. 2017. Translational resources: Reducing the gap between academic research and HCI practice. In *Proceedings of the 2017 Conference on Designing Interactive Systems.* 957–968.
[18] Lucas Colusso, Tien Do, and Gary Hsieh. 2018. Behavior change design sprints. In *Proceedings of the 2018 Designing Interactive Systems Conference.* 791–803.
[19] Lucas Colusso, Gary Hsieh, and Sean A. Munson. 2016. Designing closeness to increase gamers' performance. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* 3020–3024.
[20] Lucas Colusso, Ridley Jones, Sean A. Munson, and Gary Hsieh. 2019. A translational science model for HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* 1–13.
[21] Katja Corcoran, Gayannee Kedia, Rifeta Illemann, and Helga Innerhofer. 2020. Affective consequences of social comparisons by women with breast cancer: An experiment. *Frontiers in Psychology* 11 (2020). 1234.
[22] Peter Dalsgaard and Christian Dindler. 2014. Between theory and practice: Bridging concepts in HCI research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 1635–1644.
[23] Rachel Davis, Rona Campbell, Zoe Hildon, Lorna Hobbs, and Susan Michie. 2015. Theories of behaviour and behaviour change across the social and behavioural sciences: A scoping review. *Health Psychology Review* 9, 3 (2015), 323–344.
[24] Andy Dearden and Janet Finlay. 2006. Pattern languages in HCI: A critical review. *Human–Computer Interaction* 21, 1 (2006), 49–102.
[25] Blake M. DiCosola III and Gina Neff. 2020. Using social comparisons to facilitate healthier choices in online grocery shopping contexts. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–8.
[26] Paul Dourish. 2006. Implications for design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 541–550.
[27] Steven Dow, T. Scott Saponas, Yang Li, and James A. Landay. 2006. External representations in ubiquitous computing design and the implications for design tools. InProceedings of the 6th Conference on Designing Interactive Systems (DIS '06). ACM, New York, NY, 241–250. https://doi.org/10.1145/1142405.1142443

[28] Thomas Erickson, Ming Li, Younghun Kim, Ajay Deshpande, Sambit Sahu, Tian Chao, Piyawadee Sukaviriya, and Milind Naphade. 2013. The Dubuque Electricity Portal: Evaluation of a city-scale residential electricity consumption feedback system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1203–1212.

[29] Thomas Erickson, Mark Podlaseck, Sambit Sahu, Jing D. Dai, Tian Chao, and Milind Naphade. 2012. The Dubuque Water Portal: Evaluation of the uptake, use and impact of residential water consumption feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 675–684.

[30] Daniel Fallman. 2003. Design-oriented human-computer interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 225–232.

[31] Leon Festinger. 1954. A theory of social comparison processes. *Human Relations* 7, 2 (1954), 117–140.

[32] Ken Friedman. 2003. Theory construction in design research: Criteria: Approaches, and methods. *Design Studies* 24, 6 (2003), 507–522.

[33] Bill Gaver and John Bowers. 2012. Annotated portfolios. *Interactions* 19, 4 (2012), 40–49.

[34] William Gaver. 2012. What should we expect from research through design? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 937–946.

[35] J. P. Gerber. 2020. Social comparison theory. In *Encyclopedia of Personality and Individual Differences,* Virgil Zeigler-Hill and Todd K. Shackelford (Eds.). Springer, 5004–5011.

[36] Frederick X. Gibbons and Bram P. Buunk. 1999. Individual differences in social comparison: Development of a scale of social comparison orientation. *Journal of Personality and Social Psychology* 76, 1 (1999), 129.

[37] Daniel T. Gilbert, R. Brian Giesler, and Kathryn A. Morris. 1995. When comparisons arise. *Journal of Personality and Social Psychology* 69, 2 (1995), 227.

[38] Phillip Gough. 2017. From the analytical to the artistic: A review of literature on information visualization. *Leonardo* 50, 1 (2017), 47–52.

[39] Joanna Hale, Janna Hastings, Robert West, Carmen E. Lefevre, Artur Direito, Lauren Connell Bohlen, Cristina Godinho, Niall Anderson, Silje Zink, Hilary Groarke, and Susan Michie. 2020. An ontology-based modelling system (OBMS) for representing behaviour change theories applied to 76 theories. *Wellcome Open Research* 5 (2020), 177.

[40] Paul Hekkert and Helmut Leder. 2008. Product aesthetics. *Product Experience* 2008 (2008), 259–285.

[41] Eric B. Hekler, Predrag Klasnja, Jon E. Froehlich, and Matthew P. Buman. 2013. Mind the theoretical gap: Interpreting, using, and developing behavioral theory in HCI research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3307–3316.

[42] Jennifer Hoewe. 2017. Manipulation check. In *The International Encyclopedia of Communication Research Methods*. Wiley, 1–5.

[43] Kristina Höök, Jeffrey Bardzell, Simon Bowen, Peter Dalsgaard, Stuart Reeves, and Annika Waern. 2015. Framing IxD knowledge. *Interactions* 22, 6 (2015), 32–36.

[44] Kristina Höök and Jonas Löwgren. 2012. Strong concepts: Intermediate-level knowledge in interaction design research. *ACM Transactions on Computer-Human Interaction* 19, 3 (2012), 1–18.

[45] Stephanie Houde and Charles Hill. 1997. What do prototypes prototype? In *Handbook of Human-Computer Interaction*. Elsevier, 367–381.

[46] Bernd Huber, Katharina Reinecke, and Krzysztof Z. Gajos. 2017. The effect of performance feedback on social media sharing at volunteer-based online experiment platforms. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 1882–1886.

[47] Predrag Klasnja, Sunny Consolvo, and Wanda Pratt. 2011. How to evaluate technologies for health behavior change in HCI research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3063–3072.

[48] Tuomas Lehto and Harri Oinas-Kukkonen. 2011. Persuasive features in web-based alcohol and smoking interventions: A systematic review of the literature. *Journal of Medical Internet Research* 13, 3 (2011), e1559.

[49] Mailin Lemke and Roelof A. J. de Vries. 2021. Operationalizing behavior change theory as part of persuasive technology: A scoping review on social comparison. *Frontiers in Computer Science* 3 (2021), 1–6.

[50] Gilly Leshed, Dan Cosley, Jeffrey T. Hancock, and Geri Gay. 2010. Visualizing language use in team conversations: Designing through theory, experiments, and iterations. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*. 4567–4582.

[51] Youn-Kyung Lim, Erik Stolterman, and Josh Tenenberg. 2008. The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas. *ACM Transactions on Computer-Human Interaction* 15, 2 (2008), 1–27.

[52] Jonas Löwgren. 2013. Annotated portfolios and other forms of intermediate-level knowledge. *Interactions* 20, 1 (2013), 30–34.

[53] Danwei Tran Luciani, Martin Lindvall, and Jonas Löwgren. 2018. Machine learning as a design material: A curated collection of exemplars for visual interaction. In *Proceedings of NordDesign 2018 (DS '91)*.

[54] Susan Michie and Marie Johnston. 2012. Theories and techniques of behaviour change: Developing a cumulative science of behaviour change. *Health Psychology Review* 6, 1 (2012), 1–6.

[55] Susan Michie, Marie Johnston, Jill Francis, Wendy Hardeman, and Martin Eccles. 2008. From theory to intervention: Mapping theoretically derived behavioural determinants to behaviour change techniques. *Applied Psychology* 57, 4 (2008), 660–680.

[56] Susan Michie, Michelle Richardson, Marie Johnston, Charles Abraham, Jill Francis, Wendy Hardeman, Martin P. Eccles, James Cane, and Caroline E. Wood. 2013. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine* 46, 1 (2013), 81–95.

[57] Andrew Vande Moere and Helen Purchase. 2011. On the role of design in information visualization. *Information Visualization* 10, 4 (2011), 356–371.

[58] Zachary Munn, Micah D. J. Peters, Cindy Stern, Catalin Tufanaru, Alexa McArthur, and Edoardo Aromataris. 2018. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology* 18, 1 (2018), 1–7.

[59] Kristina Niedderer, Stephen Clune, and Geke Ludden. 2017. *Design for Behaviour Change: Theories and Practices of Designing for Change.* Routledge.

[60] Donald A. Norman. 2010. The research-practice gap: The need for translational developers. *Interactions* 17, 4 (2010), 9–12.

[61] Jane Ogden. 2003. Some problems with social cognition models: A pragmatic and conceptual analysis.*Health Psychology* 22, 4 (2003), 424.

[62] Harri Oinas-Kukkonen and Marja Harjumaa. 2009. Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems* 24, 1 (2009), 28.

[63] Rita Orji. 2017. Why are persuasive strategies effective? Exploring the strengths and weaknesses of socially-oriented persuasive strategies. In *Proceedings of the International Conference on Persuasive Technology*. 253–266.

[64] Rita Orji and Karyn Moffatt. 2018. Persuasive technology for health and wellness: State-of-the-art and emerging trends. *Health Informatics Journal* 24, 1 (2018), 66–91.

[65] Rita Orji, Lennart E. Nacke, and Chrysanne Di Marco. 2017. Towards personality-driven persuasive health games and gamified systems. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 1015–1027.

[66] Antti Oulasvirta and Kasper Hornbæk. 2022. Counterfactual thinking: What theories do in design. *International Journal of Human–Computer Interaction* 38, 1 (2022), 78–92.

[67] Alicia O'Cathain, Liz Croot, Katie Sworn, Edward Duncan, Nikki Rousseau, Katrina Turner, Lucy Yardley, and Pat Hoddinott. 2019. Taxonomy of approaches to developing interventions to improve health: A systematic methods overview. *Pilot and Feasibility Studies* 5, 1 (2019), 1–27.

[68] Marianne Graves Petersen, Ole Sejer Iversen, Peter Gall Krogh, and Martin Ludvigsen. 2004. Aesthetic interaction: A pragmatist's aesthetics of interactive systems. In *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*. 269–276.

[69] Petromil Petkov, Felix Köbler, Marcus Foth, Richard Medland, and Helmut Krcmar. 2011. Engaging energy saving through motivation-specific social comparison. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. 1945–1950.

[70] James O. Prochaska, Carlo C. DiClemente, and John C. Norcross. 1997. In search of how people change: Applications to addictive behaviors.*American Psychologist* 47, 9 (1997), 1102–1114.

[71] Jonas Rehn and Kai Schuster. 2017. Clinic design as placebo-using design to promote healing and support treatments. *Behavioral Sciences* 7, 4 (2017), 77.

[72] Yvonne Rogers. 2004. New theoretical approaches for human-computer interaction. *Annual Review of Information Science and Technology* 38, 1 (2004), 87–143.

[73] Corina Sas, Steve Whittaker, Steven Dow, Jodi Forlizzi, and John Zimmerman. 2014. Generating implications for design through design research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1971–1980.

[74] Kim Sauvé, Saskia Bakker, and Steven Houben. 2020. Econundrum: Visualizing the climate impact of dietary choice through a shared data sculpture. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. 1287–1300.

[75] Erik Stolterman. 2008. The nature of design practice and implications for interaction design research. *International Journal of Design* 2, 1 (2008), 55–65.

[76] Erik Stolterman and Mikael Wiberg. 2010. Concept-driven interaction design research. *Human–Computer Interaction* 25, 2 (2010), 95–118.

[77] Jerry Suls, Rene Martin, and Ladd Wheeler. 2002. Social comparison: Why, with whom, and with what effect?*Current Directions in Psychological Science* 11, 5 (2002), 159–163.

[78] Kristian Torning and Harri Oinas-Kukkonen. 2009. Persuasive system design: State of the art and future directions. In *Proceedings of the 4th International Conference on Persuasive Technology*. 1–8.

[79] Andrea C. Tricco, Erin Lillie, Wasifa Zarin, Kelly K. O'Brien, Heather Colquhoun, Danielle Levac, David Moher, Micah D. J. Peters, Tanya Horsley, Laura Weeks, Susan Hempel, Elie A. Akl, Christine Chang, Jessie McGowan, Lesley Stewart, Lisa Hartling, Adrian Aldcroft, Michael G. Wilson, Chantelle Garritty, Simon Lewin, Christina M. Godfrey, Marilyn T. Macdonald, Etienne V. Langlois, Karla Soares-Weiswer, Jo Moriarty, Tammy Clifford, Ozge Tuncalp, and Sharon E. Straus. 2018. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine* 169, 7 (2018), 467–473.

[80] Nynke Tromp, Paul Hekkert, and Peter-Paul Verbeek. 2011. Design for socially responsible behavior: A classification of influence based on intended user experience. *Design Issues* 27, 3 (2011), 3–19.

[81] Selen Türkay, Daniel Seaton, and Andrew M Ang. 2018. Itero: A revision history analytics tool for exploring writing behavior and reflection. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems.* 1–6.

[82] Nina Valkanova, Sergi Jorda, Martin Tomitsch, and Andrew Vande Moere. 2013. Reveal-it! The impact of a social visualization projection on public awareness and discourse. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 3461–3470.

[83] Thomas van Arkel and Nynke Tromp. 2022. Designing appropriate things: An experiential perspective on the effectiveness of artefacts in contributing to behaviour change. In *DRS2022: Bilbao, D. Lockton, S. Lenzi, P. Hekkert, A. Oak, J. Sadaba, and P. Lloyd (Eds.).* Design Research Society, 1–19.

[84] Martijn Van Welie, Gerrit C. Van Der Veer, and Anton Eliëns. 2000. Patterns as tools for user interface design. In *Tools for Working with Guidelines.* Springer, 313–324.

[85] Raphael Velt, Steve Benford, and Stuart Reeves. 2017. A survey of the trajectories conceptual framework: Investigating theory use in HCI. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* 2091–2105.

[86] Thomas Webb, Judith Joseph, Lucy Yardley, and Susan Michie. 2010. Using the internet to promote health behavior change: A systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. *Journal of Medical Internet Research* 12, 1 (2010), e1376.

[87] Max S. Wenger, Jarad Bell, Peter McEvoy, Cherie Yamaguchi, and Auriana Shokrpour. 2014. Bloom: Fostering healthy and peaceful pregnancies with personal analytics. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems.* 245–250.

[88] Mikael Wiberg and Erik Stolterman. 2014. What makes a prototype novel? A knowledge contribution concern for interaction design research. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational.* 531–540.

[89] Jichen Zhu, Diane H. Dallal, Robert C. Gray, Jennifer Villareale, Santiago Ontañón, Evan M. Forman, and Danielle Arigo. 2021. Personalization paradox in behavior change apps: Lessons from a social comparison-based personalized app for physical activity. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–21.

[90] John Zimmerman and Jodi Forlizzi. 2008. The role of design artifacts in design theory construction. *Artifact: Journal of Design Practice* 2, 1 (2008), 41–45.

[91] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 493–502.

Received 11 May 2022; revised 12 May 2023; accepted 20 June 2023