How Instructional Data Physicalisation Fosters Reflection in Personal Informatics

Marit Bentvelzen
Utrecht University
Utrecht, the Netherlands
m.bentvelzen@uu.nl

Julia Dominiak
Lodz University of Technology
Łódź, Poland
julia.dominiak@dokt.p.lodz.pl

Jasmin Niess
University of St. Gallen
St. Gallen, Switzerland
jasmin.niess@unisg.ch

Frederique Henraat
Utrecht University
Utrecht, the Netherlands
f.c.henraat@students.uu.nl

Paweł W. Woźniak
Chalmers University of Technology
Gothenburg, Sweden
pawelw@chalmers.se

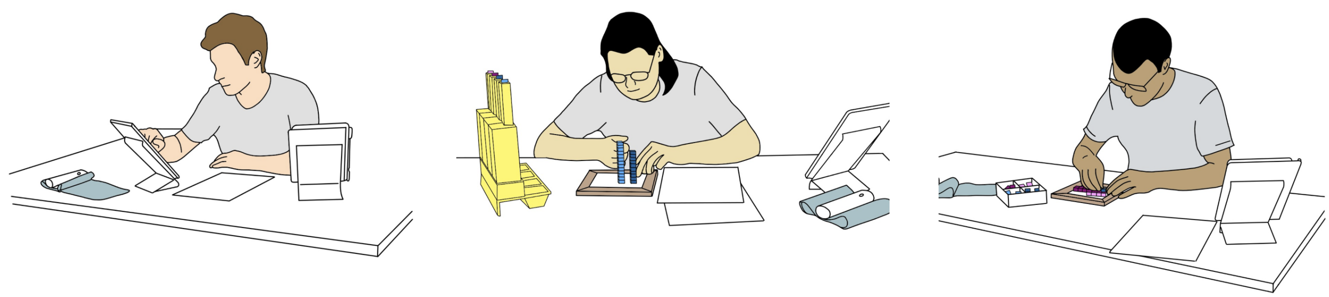

Figure 1: Three conditions were used in our study. They show the BASELINE condition in which a participant is interacting with their personal data using a mobile application (left), a participant building a data physicalisation guided by pre-counted bricks and instructions (middle), and a participant building a data physicalisation without instructions (right).

ABSTRACT

The ever-increasing number of devices quantifying our lives offers a perspective of high awareness of one's wellbeing, yet it remains a challenge for personal informatics (PI) to effectively support data-based reflection. Effective reflection is recognised as a key factor for PI technologies to foster wellbeing. Here, we investigate whether building tangible representations of health data can offer engaging and reflective experiences. We conducted a between-subjects study where $n = 60$ participants explored their immediate blood pressure data in relation to medical norms. They either used a standard mobile app, built a data representation from LEGO® bricks based on instructions, or completed a free-form brick build. We found that building with instructions fostered more comparison and using bricks fostered focused attention. The free-form condition required extra time to complete, and lacked usability. Our work shows that designing instructional physicalisation experiences for PI is a means of improving engagement and understanding of personal data.

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04.
https://doi.org/10.1145/3544548.3581198

CCS CONCEPTS

• **Human-centered computing** → **Interactive systems and tools**.

KEYWORDS

personal informatics, reflection

ACM Reference Format:
Marit Bentvelzen, Julia Dominiak, Jasmin Niess, Frederique Henraat, and Paweł W. Woźniak. 2023. How Instructional Data Physicalisation Fosters Reflection in Personal Informatics. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3544548.3581198

1 INTRODUCTION

Our lives are characterised by an increasing amount of data [13]. Ubiquitous personal trackers and measurement devices are equipped with novel sensors and collect increasingly accurate measurements [9]. Personal Informatics (PI), i.e. the practice of gathering data about our health, wellbeing and activity, has become a prominent feature of many everyday technologies. Consequently, as the volume and variety of personal data collected increases, it becomes a challenge for users to understand and manage their PI experience. Reflection [7] is regarded as a key design goal for systems if they are to use data collection for the wellbeing of the user. While research in

Human-Computer Interaction has explored systems which foster reflection, it is still unknown which design features benefit reflection and how to design effective reflection interfaces. To contribute to solutions for this challenge, this paper investigates how users can interpret data about themselves using tangible representations of data, i.e. physicalisation.

Data physicalisation is an established topic in HCI [25], yet its relationship with PI is yet to be fully explored. Past designs either build the physicalisation for the user or offer extensive freedom in constructing one. One class of systems in past research would automatically build physicalisations of personal data for users, not involving the user in the process of creating them. In this vein, Khot et al. [28] used 3D-printed tangibles to foster recollections of physical activity. Sauvé et al. [45] designed LOOP—a tangible situated display which visualised fitness tracker data. These studies showed that tangible representations of data provided an attractive alternative to screens for personal informatics. Another strain of work investigated cases where users were provided with materials for physicalisation and encouraged engaging in open-ended construction of physicalisations. Lee and Hong [30] used clay models to help college students express their daily emotional state. Thudt et al. [52] distributed craft material to participants and asked them to visualise a variety of personal data types. They found that the active construction of a physicalisation fostered reflection on data. This reflects the recognised property of tangible interfaces to being 'closely tied to the information they represent' [47]. While past research shows clear potential for physicalisation fostering reflection, the variety of results and levels of involvement with the artefacts show that there is a need to understand this design space more systematically. Thus, the potential of physcialisations fostering reflection in personal informatics is yet to be explored. To effectively design for personal informatics, recognising that dynamic, guided data exploration has been shown to offer positive personal informatics experiences [6, 14], we need to understand how much structure and instruction should be provided to the user in the process of creating a physicalisation. Further, there is a lack of comparative studies in HCI which investigate the influence of creating physicalisations (and how such physicalisations are created) on personal informatics experiences.

To bridge this gap, this paper investigates if building tangible representations of health data, as a common form of personal informatics [13], can offer engaging and reflective experiences. We endeavour to study if and how users can benefit from data physicalisation in the light of current theories in personal informatics. To this end, we conducted a between-subjects study where $n = 60$ participants explored their immediate blood pressure data in relation to medical norms. We compared three conditions: 1) a BASELINE condition in which participants used a standard mobile app for blood pressure measurements, 2) an INSTRUCTIONAL physicalisation condition in which participants received pre-counted LEGO® bricks and were guided in building a physical representation of their blood pressure measurement, and 3) a FREE-FORM physicalisation condition in which participants independently created a data physicalisation of their data using LEGO® bricks. We found that the INSTRUCTIONAL condition fostered reflection through comparison. The free-form condition engaged the participants for the longest time, yet lacked usability. Our work shows that self-built tangible representations can benefit PI systems if designed in a structured manner.

This paper contributes the following: (1) a between-subjects mixed-method study with $n = 60$ participants of how users explore blood pressure data using three different methods and (2) insights into how instructional physicalisation can offer benefits for fostering reflection through comparison in PI. In this work, we first review related literature in reflection, personal informatics, and data physicalisation. We then present the method and findings from our study, and conclude with a discussion and directions for future research.

2 RELATED WORK

In this section, we first provide an overview of the general understanding of reflection within HCI. We then review systems that aim to enhance reflection. Finally, we discuss the use of data physicalisation for encouraging data-driven reflection. We demonstrate how the area of inquiry where personal informatics and designing data physicalisation overlap requires further exploration.

2.1 Reflection in HCI

The HCI field has invested considerable effort in building an understanding of reflection. Reflection is seen as a multi-faceted concept [7] and previous work indicates a lack of conceptual agreement in the field regarding the definition of reflection [3, 7]. Even though there is not a universally agreed-upon definition, Schön's framing of reflection [3, 50] has been prevalent in HCI research [3, 7]. A systematic review by Baumer et al. [3] showed that 70% of HCI papers that explicitly defined reflection used Schön's notion of reflection-in-action or reflection-on-action. Reflection-in-action takes place during action, implying that a person is reflecting on their actions while performing them [46]. Usually, this requires unexpected events or unpredictable outcomes of actions [36]. Reflection-on-action is a post-hoc act. It is a reconstruction of an experience, based on our memories of it. This type of reflection allows users to draw conclusions based on an analysis of past events [15]. Furthermore, Schön emphasizes that reflection often does not occur automatically, but needs to be encouraged, as previously noted by Slovak [3, 50]. While Schön's framing of reflection offers a lens for reflection, it does not directly address how technology can support reflective processes [50], leading to a gap in our understanding of how reflection can be facilitated by technology. In this work, we contribute to HCI's understanding of how to foster reflection by studying how physicalisation can be employed to foster reflection on personal data. As there is only incidental evidence that data physicalisations may foster reflection, we explore ways in which physical data representations can be effectively used to encourage reflection through providing the user with guidance for data exploration as seen in many past examples of technologies for reflection [3, 7].

2.2 Reflection in Personal Informatics

Reflection was also studied in the context of personal informatics experiences. Personal informatics systems enable users to understand their health and wellbeing through automatically collected personal data. By reflecting on personal data, a user can notice patterns and trends, which can lead to more knowledge about oneself [13]. Personal informatics systems have also been critiqued in the past for not actively encouraging reflection [10, 29]. As noted by Baumer [3], these systems often carry an implicit assumption that by showing a user visualisations of their past data for the purpose of reflection, that reflection will occur. However, this conflicts with reflection theories that emphasize the importance of encouraging reflection, as it often does not occur automatically [50].

Several models that describe the user's journey in using personal informatics systems have been proposed. They aim to build a more in-depth understanding of the reflection process. Epstein et al. [14] proposed the *Lived Informatics Model of Personal Informatics*, which is an extension of Li's *Stage-Based Model of Personal Informatics Systems* [31]. The Lived Informatics Model consists of four stages: deciding to track, selecting tools, tracking and acting (which is considered to be an ongoing process of collection, integration, and reflection), and lapsing. While reflection is considered to be a prominent part of the lived experience of personal informatics, these models describe reflection from a meta-perspective.

To come to a deeper understanding of the reflection phase in the personal informatics experience, the *Lived Informatics Model of Personal Informatics* was later extended by Bentvelzen et al. [6]. Their work further dissected the reflection stage in the personal informatics process. Their *Technology-Mediated Reflection Model (TMRM)* describes user behaviours and practices in the reflection phase of a personal informatics journey. The model shows how users enter, exit and stay in the reflection phase. In the TMRM, reflection is considered to be a temporary state—a dynamic process in which a user constantly adapts their tracking experience to their evolving needs. The temporal and conceptual cycles are essential in this process. These cycles show how users' needs and perspectives evolve throughout their engagement with their tracker. The TMRM interprets reflection as a cyclic process with intermediate stages. This implies that systems that effectively support reflection help the user to stay within the reflection cycle. However, the iterative nature of this process implies that user needs undergo constant changes, which, in turn, requires flexibility from the tracking system. Everyday reflection requires everyday changes in perspective, indicating a need for dynamic data-driven reflection support. To this end, we inquire how tangible representations of health data, which is most often studied as an example of PI [13], can be used to allow users to explore their data from different perspectives. Thus, we investigate the design properties of systems which can contribute to the continuous reflection process as described in current models of personal informatics. Our research explores if and how data physicalisations can be effectively used in a PI context by studying personal data representations through the lens of PI and reflection theory, in order to stimulate the potential development of the design space of physicalisations for PI.

2.3 Facilitating reflection through data physicalisation

A wide variety of strategies have been used in HCI to foster reflection. A structured literature review by Bentvelzen et al. [7] lists these in a taxonomy of eleven design resources and 74 design patterns. One of the design resources that has been used in HCI artefacts and commercial mobile apps for reflection is *reframing*. Systems that use this aim to let the user see something in a new perspective, which, in turn, evokes reflection. Reframing is often implemented through the use of *data physicalisation* as a design pattern. Data physicalisation aims to help people explore, understand, and communicate data using physical data representations [12, 24–26, 33, 35]. Niess and Woźniak [38] noted that effective reflection on personal collected data required translating qualitative goals in quantitative measures. Data physicalisation shares a similar property where digital data points are represented through analogue physical artefacts, providing a viable alternative for fostering reflection in personal informatics. Further, work in personal informatics suggests that users need to invent new workflows for analysing data to sustain reflection in tracking [6]. Data physicalisations have been shown to offer alternatives to traditional analysis workflows [57], thus suggesting their key potential as personal informatics tools. However, we note that works that looked into processes of constructing physicalisation in detail, such as [42, 57], were conducted outside of the context of personal informatics. Consequently, it remains to be explored if and how the exploratory properties of physicalisation can be applied to exploring and reflecting on personal data.

Sauve et al. [44] differentiates between three types of physicalisations: (1) static composite physicalisations, (2) shape-changing interfaces, and (3) constructive visualisations. Static composite physicalisations consist of objects that represent a 'bucket' of data that cannot be reconfigured by the user. An example of such a physicalisation comes from Khot et al. [27] who created 3D printed artefacts called *SweatAtoms* as a means to physicalise the user's physical activity data (i.e. heart rate). Shape-changing interfaces, on the other hand, support interaction with *dynamic* composite physicalisations. For instance, Sauve et al. [41, 45] created LOOP, which provides an abstract visualisation of the user's activity data by changing its shape [44]. Finally, constructive visualisations support *the free reconfiguration of non-actuated token-based physical data representations* [17, 21, 22, 44, 52]. In such systems, the user can construct and reconfigure their data through physical building blocks (tokens). For instance, Huron et al. [21] explored the use of physical tokens as a data authoring tool for non-experts making sense of financial data. Related to this, Thudt et al. [52] conducted a qualitative study that explored the relationship between self-reflection and the construction of personal physicalisations. In their study, participants were familiarized with data physicalisation in a group workshop, and used a provided kit consisting of materials such as beads to construct their physicalisation at home over a period of 2–4 weeks. Their work showed that the flexibility and customisability enabled participants to make data collection and representation part of their everyday routines, and resulted in deep reflections for several participants. This implies that constructive physicalisation potentially facilitates data driven reflection by supporting users to engage with their data.

However, given the exploratory nature of previous work in this area, there is currently a lack of understanding regarding the relationship between the physical construction of data and reflection. Further, previous work regarding data physicalisations in a personal informatics context either focused on a tangible data representation in which a user could not influence or control the physicalisation process (e.g. LOOP [27, 45]), or on open ended constructive physicalisations in which users independently created a tangible representation from their data. A middle way, in which users are guided but are, at the same time, free to influence the physicalisation construction process, has, to the best of our knowledge, not been investigated so far. Consequently, we argue that there is a need for a more systematic approach. To this end we conducted a mixed-method between-subject study that explores whether building tangible representations can offer engaging and reflective experiences. Thus, we contribute to work on personal informatics by building a new understanding of the design properties of physical data representations in the context of reflecting on personal information.

3 METHOD

To investigate how data physicalisation can benefit data exploration in personal informatics, we conducted a between-subjects mixed-method controlled user study. In particular, we were interested in the consequences of introducing an instructional approach to building data physicalisations. In the study, we asked participants to explore their blood pressure data under three conditions: 1) a baseline condition using a mobile application (BASELINE), 2) an instructed physicalisation condition (INSTRUCTIVE), and 3) a free-form physicalisation (FREE-FORM). This study investigated how users interpreted and reflected on their personal data using tangible representations.

To allow comparison between the three conditions we made sure that participants were interacting with the same type of personal data. As such, we decided to use a metric that we could measure for each participant. In order to stay within the scope of the majority of work on personal informatics [13], we opted to use data relevant to health or wellbeing. This would allow us to later relate our findings to models of personal informatics experiences. We considered several metrics e.g., resting heart rate or heart rate variability, but these often need to be measured over a longer period of time to become useful and meaningful [48]. Therefore, we decided to use blood pressure, as this measurement offers a reliable value at a single point in time, and has the same ranges for each participant (e.g. regardless of age or sex). It is also a metric which is not featured in most commercial fitness trackers, which implies that the participants would be uniformly unaware of their current blood pressure values, reducing novelty bias. We chose a between-subject design as personal data insights can only be developed once, replicating past studies of data reflection, discovery or sensemaking, e.g., [23, 56]. In the remainder of this section, we present the details on the participants, conditions, measures, and analysis methods.

3.1 Participants

We used our social networks combined with snowball sampling to recruit $n = 60$ participants. We did not use exclusion criteria, as our study design was suitable for all possible participants. The participants were aged 18–46 years, $M = 23.6$, $SD = 4.89$. Thirty-four identified as male, twenty-four as female, one as non-binary, and one preferred not to answer. All participants were residents of the European Union and were interviewed in their native language or English. The participants were asked for consent for recording and informed that they could terminate the study at any time. Theywere compensated for their participation according to the average income of the respective countries. No participants reported cardiovascular health issues prior to participating in the study.

3.2 Conditions

Our study compared three conditions, each with a different mode of interacting with personal data representations:

Digital visualisation (BASELINE). In the BASELINE condition, participants reviewed their blood pressure data using a smartphone application which showed the systolic and diastolic measurements as a numeric value, visualized as a line graph (related) , as shown by Figure 2. Considering that the majority of commercial PI systems use smartphone applications as a way to present personal data to the user, we decided that it would be fitting to use this as the BASELINE condition. Here, we decided to choose an existing solution available on the consumer market in order to compare design alternatives to a strong baseline [19], which in turn enables comparisons relevant for future design. Further, studying practices around existing systems is a core practice in personal informatics [13]. Consequently, this condition enabled us to include an understanding of how users conceptualise their data in the context of the study using existing technologies. In the other two conditions, we asked participants to use tangible objects (i.e. LEGO® bricks) to build a physical representation of their blood pressure measurement. To understand the role of structure and instruction in building data physicalisations, we varied the level of instruction in these two conditions.

Instructive Physicalisation (INSTRUCTIVE). In the INSTRUCTIVE DATA PHYSICALISATION condition participants received pre-counted bricks from a custom-made brick dispenser, and were instructed to build a tower out of each brick colour. We decided to focus on bar-chart-like towers in the study as they are an example data representation type which is well understood in personal informatics research and widely used in commercial applications [20, 37]. Further, the activity of building towers (and the language associated with the construction thereof) is commonly used in serious play activities which employ LEGO® bricks [34]. Thus, tower building offered a rich example of an Instructive Physicalisation with documented prior use and limited complexity. We decided to deliver the bricks to the participants using a dispenser (as opposed to an experimenter counting and providing the bricks) in order to reduce a possible bias of this condition being perceived as more social, i.e. with increased involvement of the experimenter. This was a particular risk as many users desire social personal informatics experiences [14].

Free-form physicalisation (FREE-FORM). Conversely, in the FREE-FORM DATA PHYSICALISATION condition, participants were asked to build a data visualisation of their blood pressure measurement using LEGO® bricks in four colours. Eighty bricks, 20 per colour, were placed in a box in front of the participants. In contrast to INSTRUCTED DATA PHYSICALISATION, we provided no suggestions on the data representation nor the amount of bricks to be used.

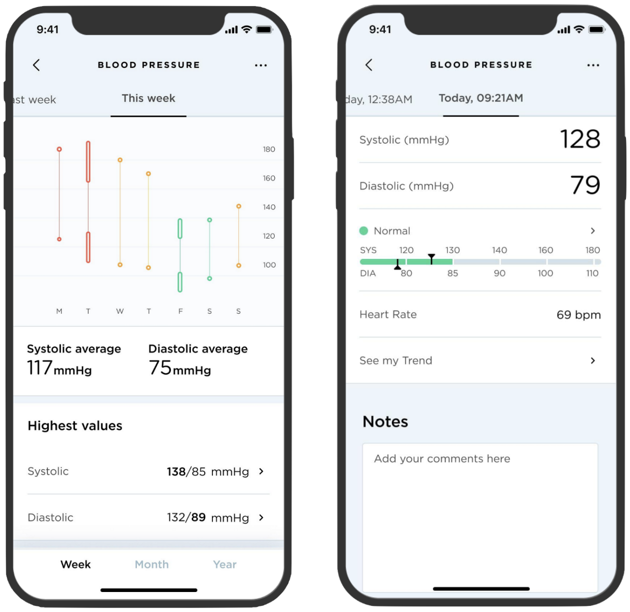

Figure 2: The interface of the Withings Health Mate smartphone application. Participants in the BASELINE condition explored their blood pressure measurement data using this interface. The application enables users to get an overview of their data on different time scales (i.e. day, week, month, year). By default, the app shows a weekly average (left), or as a single measurement (right). In our study participants performed a single measurement, and thus, the weekly overview also only showed one value (if accessed). Participants interacted primarily with the single measurement view (right).

3.3 Measures

In each condition we measured reflection, user engagement and the engagement time.

Reflection. We used the Technology-Supported Reflection Inventory (TSRI [5]) to compare the level of reflection induced between the three conditions. The TSRI offers overall reflection scores and three subscales which may identify the sources of reflection: insight, exploration and comparison. As suggested by Bentvelzen et al. [5], we measured trait reflection with the Self-Reflection and Insight (SRIS) scale [18] before using the TSRI to control for personality differences among participants. Past research implies that data physicalisation may have led to increased reflection [53].

User Engagement. Next to reflection, we measured engagement, using the short version of the User Engagement Scale (UES-SF) [39]. Considering that facilitated reflection [6] often requires encouragement [3, 16, 50] and therefore active involvement from the user, we wanted to evaluate if a user's perceived engagement differed between the three conditions. The UES-SF offers overall user engagement scores and four subscales: *Focused Attention* (FA), *Perceived Usability* (PU), *Aesthetic Appeal* (AE), and *Reward factor*, (RW).

Engagement time. Finally, we measured how long participants engaged with their blood pressure data (in seconds). Previous work by Baumer [4] noted that slowness can offer space for reflection [4], and time is considered to be a condition for reflection [16]. As such, the engagement time could possibly carry more information about the nature of the interaction and reflection in the study, which is why we decided to use this as a measure.

3.4 Apparatus

To conduct the study, we used: (1) a consumer-grade blood pressure measurement device, (2) a dispenser that would automatically provide the number of bricks needed to represent the blood pressure measurement and, (3) LEGO® bricks and base plate.

3.4.1 Withings BPM Connect. To measure the participant's blood pressure during the study, we used the *Withings BPM Connect*[1], which is an upper-arm monitor integrated in a cuff that inflates automatically with a press of a button. This type of monitor is easier to use compared to monitors that require the user to pump a bladder to inflate them manually, and therefore ensures that a participant can instantly use it without receiving extensive instructions. The device represents the current standard in consumer-grade blood pressure measurement [54]. Furthermore, this particular device is not intended to be used as a medical device in the doctor's office, but instead enables users to track their blood pressure measurements at home. As such, it is designed to be part of one's personal informatics device ecology, making it suitable for this study. Finally, this monitor is accompanied by a smartphone application that shows the systolic and diastolic measurements as a numeric value, and visualizes the measurement as a line graph.

3.4.2 Brick Dispenser. For the INSTRUCTIONAL condition we designed a LEGO® bricks dispenser. 3D models of the construction with the casing and trays were designed in Fusion 360 and then printed from PLA material on the Raise3D E2 printer. The main structure of the system consists of four segments (as shown in Figure 4): a chimney that can contain up to 20 bricks with slots for two servomotors; a cover to protect the motors; a base with rails for dispensers with a cover; and an elevation with trays for bricks. One dispenser was required for each color. For each single dispenser, there are two servos in the upper and lower position in the micro size (Redox S90). The servo located above is responsible for holding the entire column of elements in a chimney in place, leaving only one free brick under the horn. The last brick placed on the horn of the lower servo, acting as a trapdoor. When the bricks are delivered, the following algorithm is executed: holding the column, releasing the trapdoor with only one brick, closing the trapdoor, and releasing the column. Due to the necessity to hold up to 20 elements, we designed a dedicated wide horn to increase the surface that holds the column and allow it to be pressed down, and a flat side surface to ensure the reliability of releasing a single element.

The system operation is coordinated by an Arduino Uno Rev3 microcontroller, while the position of the servos is controlled by the Mini Maestro USB 18-channel servo driver connected via I2C to

[1] https://www.withings.com/eu/en/bpm-connect

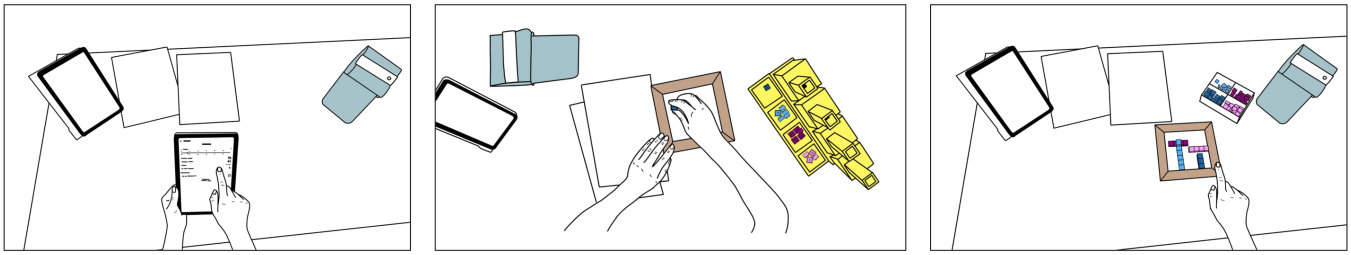

Figure 3: Three conditions used in our study show the BASELINE condition in which a participant is interacting with their personal data using a mobile application (left), a participant building a data physicalisation guided by pre-counted bricks and instructions in the INSTRUCTIONAL condition (middle), and a participant building a data physicalisation without instructions in the FREE-FORM condition (right).

Arduino. The results of the Withings BPM Connect were manually copied into the dispenser software by the experimenter.

3.4.3 LEGO® Bricks and Base Plate. We used LEGO® interlocking bricks as a tangible representation of the participant's blood pressure data. During the study we used 2x2 bricks in four colours: dark blue, blue, dark pink, and bright pink. This colour scheme ensures that the bricks are also distinguishable for users with Deuteranopia (red-green color blindness). We used a customised white LEGO® base plate, shaped into a square with side size of approx. $15cm$. In total we had 80 bricks available; 20 per colour. The reason for having 20 bricks is two-fold: first, it made it possible to represent almost all blood pressure measurements with 1 brick representing a value of $10mmHg$ (e.g. a blood pressure of 120/80 would then for instance be represented with 12 bricks of one colour and 8 bricks of another). Needing more than 20 bricks would mean that a participant would be in a *hypertensive crisis*, in which case the participant would be in a medical emergency and should not have participated in the study. Second, we restricted the number of bricks to 80 to control engagement time and avoid overwhelming the participants. The bricks were provided by either the dispenser, or in a wooden tray with four containers, one for each colour.

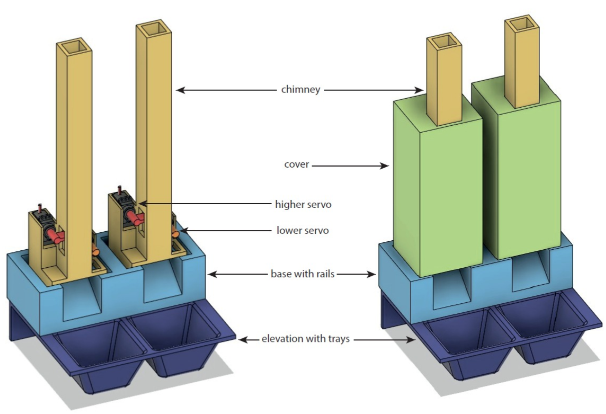

Figure 4: The individual components of the brick dispensers.

3.4.4 Tablets. We used a Samsung Galaxy Tab S6 tablet to display the survey that we built using the Qualtrics XM platform. This survey helped us to structure the session, and consisted of two consent forms, a video that introduced the blood pressure monitor, the demographics, and finally, the SRIS, TSRI and UES-SF scales. In the BASELINE condition we used an additional tablet (an iPad (2019)) to show the blood pressure measurement. The video and survey are available in the supplementary material.

3.4.5 Information sheet. We handed participants an information sheet consisting of a classification table with the values associated to optimal, normal, high normal and grade 1–3 hypertension blood pressure. Next to this classification table, the information sheet described high and low blood pressure. The information sheet was a print-out of the information offered in the Withings Health Mate application. By giving the sheet to all participants we ensured that the received information regarding blood pressure was the same in all three conditions. The sheet also included reference values which facilitated the interpretation of blood pressure values. Next to this, all participants received a sheet of paper and a pen to make notes about their data during the study.

3.5 Procedure

We randomly assigned participants to one of the three conditions. After introducing the purpose of the study, we asked participants to sign two consent forms, one for collecting data and recording the session, and one in which we made the participant aware that the study was not a medical examination, nor were we using a medical device. We then introduced the blood pressure monitor by showing a 30-second video that introduced the device and demonstrated how to take a measurement. After the participants watched the video, we helped them to attach the cuff of the blood pressure monitor and asked them to start the measurement by pressing the button. During the synchronisation of the monitor and the supporting app, we asked the participants to answer questions about demographic data, and to complete the SRIS scale. When the participant was finished, the interaction with their data started, which differed per condition.

Digital visualisation (BASELINE). In the BASELINE condition, we then gave participants an iPad tablet with the Withings Health Mate application opened, which showed participants their blood pressure measurement. We asked participants to review their data using the application.

Instructive physicalisation (INSTRUCTIVE). In the INSTRUCTIVE condition, we introduced the brick dispenser and explained that the dispenser would pre-count the number of bricks needed for the four colours (which represented the systolic and diastolic values for the optimal value and the actual measurement, based on the measurement and participant demographic information). In addition, we explained that each brick represented a value of $10mmHg$. We then asked participants to build a tower for each colour on the base plate, and started the dispensers.

Free-form physicalisation (FREE-FORM). In the FREE-FORM condition, we gave participants an A6-sized piece of paper with the systolic and diastolic values of both the optimal value, and their blood pressure measurement. We then explained that the task was to build a data visualisation of their blood pressure measurement on the base plate using the bricks in the wooden tray. We informed participants that there were 20 bricks available for each color and gave the tip that it might be easy to use a value of 10 for each brick, but that they were free to choose their preferred data representation. We then asked participants to build their visualisation.

Participants were instructed to use as much time as needed to understand their data. After finishing either viewing their data in the application, or building the data physicalisation, we asked participants to complete the remainder of the survey, which consisted of the TSRI and UES-SF scales. Next, we conducted a semi-structured interview, consisting of three questions: (1) *what insights did this blood pressure measurement offer you?*, (2) *what did you learn about your blood pressure data?*, and (3) *did the measurement lead to any self-reflective thoughts?*. Finally, we debriefed the participants and remunerated them for their participation.

4 RESULTS

Here, we first present our quantitative results. We then provide a detailed description of our qualitative findings based on the exit interviews and the analysis of the brick structures. The results are illustrated with excerpts from the interviews.

4.1 Quantitative results

We used analysis of variance (ANOVA) procedures to analyse the data collected in the survey. The full data set is available in the auxiliary material, along with the analysis. We checked normality assumptions for each test before performing the analysis. All p values reported were corrected with the Bonferroni-Holm method where appropriate.

4.1.1 *Reflection.*

The grand mean of TSRI scores was $M = 28.37$, $SD = 5.11$. We conducted a one-way ANCOVA to determine the effect of the different modes of interacting with personal data on the TSRI score, controlling for trait reflection. There was no significant effect for the full scale $F_{2,56} = 0.87$, $p = .43$. Additionally, we conducted the analysis for all the subscales of the TSRI, as shown in Table 1. Below, we report on the TSRI subscale for which we found a significant effect of condition, i.e. comparison.

Comparison. The grand mean of TSRI-Comparison was $M = 10.15$, $SD = 2.44$. As shown in Figure 5, the highest Comparison scores were recorded for the INSTRUCTED ($M = 36.08$, $SD = 4.27$) condition. The one-way ANCOVA showed a significant effect $F_{2,56} = 4.56$, $p < .05$. The covariate was not significant. Post-hoc tests using Tukey HSD showed that there were significant differences between the pair INSTRUCTED–BASELINE at $p < .05$, and between the pair INSTRUCTED–FREE-FORM at $p < .05$.

4.1.2 *User Engagement.*

The grand mean of UES-SF scores was $M = 45.45$, $SD = 6.27$. We conducted a one-way ANOVA to determine the effect of the different modes of interacting with personal data on the UES-SF score. There was no significant effect for the full scale $F_{2,57} = 0.21$, $p = .81$. Additionally, we conducted the analysis for all subscales of the UES-SF. Below, we report on the UES-SF subscales for which we found a significant effect of condition: focused attention (FA), and perceived usability (PU).

Focused Attention (FA). The grand mean of FA was $M = 9.45$, $SD = 2.47$. As shown in Figure 6, the highest FA scores were recorded for the FREE-FORM ($M = 10.40$, $SD = 1.57$) condition, followed by the INSTRUCTED ($M = 9.90$, $SD = 2.65$) condition. The one-way ANOVA showed a significant effect of condition $F_{2,57} = 5.87$, $p < .01$. Post-hoc tests using Tukey HSD showed that there were significant differences between the pair FREE-FORM–BASELINE at $p < .01$, and between the pair INSTRUCTED–BASELINE at $p < .05$.

Perceived Usability (PU). The grand mean of PU was $M = 12.55$, $SD = 2.35$. As shown in Figure 6, the highest PU scores were recorded for the BASELINE ($M = 13.40$, $SD = 2.30$) condition, followed by the INSTRUCTED ($M = 12.65$, $SD = 1.60$) condition. The one-way ANOVA showed a significant effect of condition $F_{2,57} = 3.19$, $p < .05$. Post-hoc tests using Tukey HSD showed that there was significant difference between the pair BASELINE–FREE-FORM at $p < .05$.

4.1.3 *Engagement time.*

The grand mean of the engagement time (measured in seconds) was $M = 314.38$, $SD = 208.34$. A one-way ANOVA showed a significant effect of condition $F_{2,57} = 4.56$, $p < .05$. Post-hoc tests using Tukey HSD showed a significant difference between the pair BASELINE–FREE-FORM at $p < .05$.

4.2 Qualitative results

We analysed two sets of qualitative data obtained in the study. First, we report on the results of exit interviews conducted with participants. Then, we discuss the brick structures built during the study, obtained from photos and videos.

4.2.1 *Exit Interviews.*

Interview data was transcribed verbatim for analysis, based on a total of 307 minutes of recording. We adopted the pragmatic approach to qualitative analysis as proposed by Blandford et al. [8]. Two coders open-coded the data set using the Atlas.ti software package. We then held iterative discussion sessions to merge codes and subsequently constructed three themes which described the users' experience of exploring their blood pressure data: (i) understanding a complex metric, (ii) facilitating reflection, and (iii) construction process. In the following, we report our findings and illustrate them with excerpts from the transcripts, annotated

Table 1: The mean value and standard deviation for the TSRI, UES and the engagement time (measured in seconds). In addition, the results from the respective ANOVAs (ANCOVAs for the TSRI and its subscales) are presented. † and ‡ show significantly different pairs, calculated using Tukey HSD.

	Total		Baseline		Instructed		Free-form		ANOVA
	M	SD	M	SD	M	SD	M	SD	
TSRI	28.37	5.11	28.10	5.11	29.55	4.75	27.45	5.47	$F_{2,56} = 0.87, p = .43$
TSRI-Insight	6.77	2.43	6.55	2.31	6.40	2.33	7.35	2.66	$F_{2,56} = 0.87, p = .43$
TSRI-Exploration	11.45	1.97	11.90	1.94	11.75	1.83	10.70	2.00	$F_{2,56} = 2.27, p = .11$
TSRI-Comparison	10.15	2.44	9.65†	2.48	11.40†‡	1.85	9.40‡	2.54	$F_{2,56} = 4.56, p < .05$
UES	45.45	6.27	44.70	6.28	45.75	6.71	45.90	6.04	$F_{2,57} = 0.21, p = .81$
UES-FA	9.45	2.47	8.05†‡	2.48	9.90†	2.65	10.40‡	1.57	$F_{2,57} = 5.87, p < .01$
UES-PU	12.55	2.35	13.40†	2.30	12.65	1.60	11.60†	2.74	$F_{2,57} = 3.19, p < .05$
UES-AE	11.92	2.20	11.85	2.46	11.90	2.13	12.00	2.13	$F_{2,57} = 0.02, p = .98$
UES-RW	11.53	1.97	11.40	1.79	11.30	2.32	11.90	1.80	$F_{2,57} = 0.52, p = .60$
Engagement time (s)	314.38	208.34	233.40†	110.59	292.30	104.11	417.45†	305.16	$F_{2,57} = 4.56, p < .05$

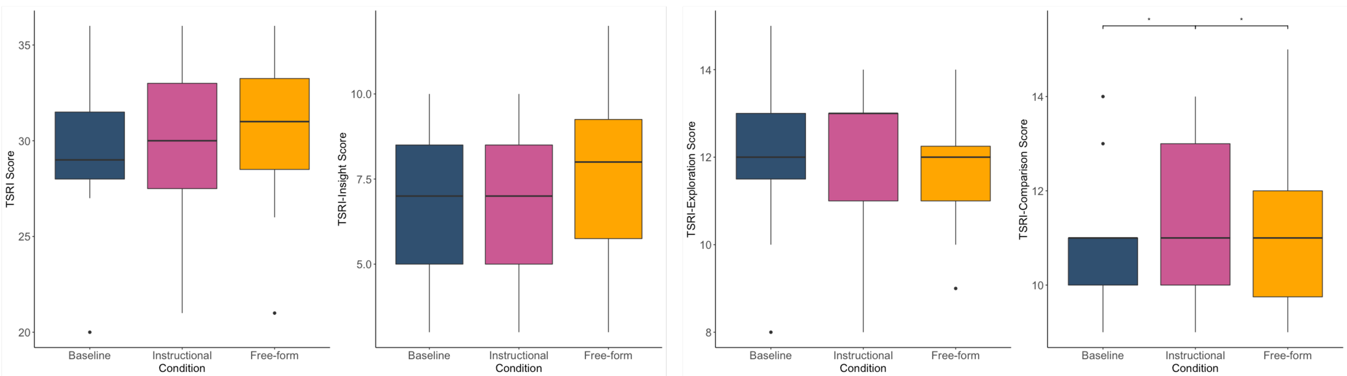

Figure 5: The distribution of TSRI scores and its subscales across conditions: Insight, Exploration, and Comparison. Significant pairs are marked with annotations. A single asterisk denotes a significance of ($p < .05$).

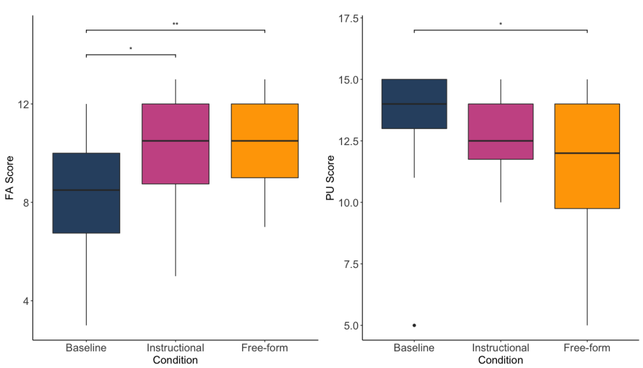

Figure 6: The distribution of scores of two UES subscales across conditions: Focused Attention (FA), and Perceived Usability (PU). Significant pairs are marked with annotations. A single asterisk denotes a significance of ($p < .05$); double asterisks represent a significance of ($p < .01$)

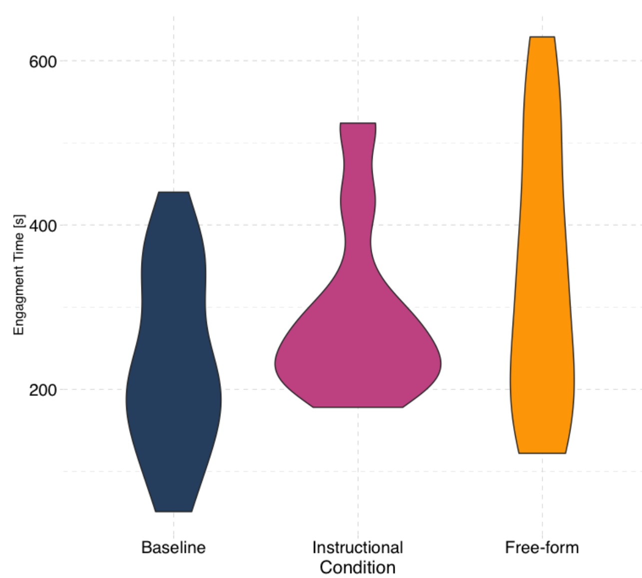

Figure 7: The engagement time in seconds per condition.

with participant numbers and the condition in which the participant conducted the task.

Understanding a complex metric. Many participants in the BASELINE condition found it challenging to grasp the meaning of their blood pressure measurement. Some expected a single value for the metric, and mentioned the complexity of understanding the notion of systolic and diastolic blood pressure. This resulted in the participants often expressing an explicit need for the application to provide additional information for interpreting the numbers, beyond stating the desired values. One participant reflected on how the measurement did not match their expectations:

> *I miss some sort of frame of reference, because I've never really thought about blood pressure before. I would need to understand a bit more how blood pressure works before I can grasp the meaning of these numbers. I see these two, systolic and diastolic, values here; and I imagined that there would be just one value for blood pressure and that I could either be above or below it. (P10, BASELINE)*

Further, participants mentioned that, while they did understand the values and the result of their measurement, they were wondering how their systolic and diastolic values could be related to their body awareness and health. Participants actively tried to tell a story which could help them to explain their measurements. They would connect their lifestyle choices and health status to the values provided by the blood pressure meter. P9, prompted by the measurement, listed choices which may result in increased blood pressure and reflected on the ways to improve health:

> *I don't have an immediate idea what affects my blood pressure. But I assume that lack of sleep is not very optimal, drinking alcohol is not very optimal, maybe exercising a little more. Yes, in terms of insight; maybe I should start living a little healthier, to get my blood pressure down. (P9, BASELINE)*

Participants also discussed the circumstances of the measurement. As many were not familiar with a blood pressure measurement device, they wondered about the role of the act of measuring pressure on the measurement obtained. This also prompted temporal reflections, discussing when during a day a blood pressure measurement should be taken. One participant discussed the contextual factors which, according to him, would affect blood pressure at a given time and indicated that they would like the blood pressure monitor to include these factors:

> *I would like to now understand how the circumstances of taking the blood pressure measurement affects the result. It would really help me if the device would also know about the environment; like how my day was until now, what is the time of day, have I been working before, did I just wake up, all these kinds of things. I would think these things would play a role. (P16, BASELINE)*

Overall, our results show that participants recognised blood pressure as important for assessing one's health. Yet, they also sought more information about blood pressure with the goal of contextualising the measurement within their perception of a healthy lifestyle.

Facilitating reflection. Across all conditions, being confronted with a current blood pressure measurement was an opportunity for the participants to dedicate time to thinking about their health data. Participants appreciated that the tasks explicitly required them to focus on their bodies in a dedicated interactive experience. P42 commented on how the need to construct the data representations necessitated focus and calmness:

> *During this work [building with the bricks] I can stop and try to think about my measurements [. . .] the biggest advantage of the system is finding the time, stop and slow down to look at the measurement (P42, INSTRUCTIONAL)*

Building data representations from bricks facilitated comparison and implicitly required interpretation of data while the representation was constructed. Participants found that the physical representation of data highlighted differences and offered an engaging analysis process. One participant reflected on how the visual and haptic experience of the bricks contributed to noticing differences between data points:

> *Yeah, I actually really liked to build this, it made it easy to compare the values. Especially in comparison to regular graphs or just a sheet with quantities. This makes it much more fun and maybe it also helps in making a lifestyle decision. Because you can clearly see and feel [with the bricks] that the measurement is higher than the optimum. (P48, INSTRUCTIONAL)*

Using bricks to represent blood pressure was appreciated by participants who preferred not to perform operations with numbers. Participants could rapidly assess their measurements with reference to desired values. P60 commented on how the bricks were glanceable and offered a better perception of the relations between values:

> *Usually these [measurements] are [represented with] numbers. And I always immediately forget what the values were. And with this, especially because I put it in such a way [points to bricks], it gave me quite a good insight of 'ok, this is my measurement and this is the optimal'. And this brings it into focus. So it's easier to understand [my blood pressure measurement] visually. I can see the ratio [between the values] better. (P60, INSTRUCTIONAL)*

As the task (in the FREE FORM and INSTRUCTIONAL) required the participants to conceptualise a representation of their data, using the bricks fostered an engaging experience. Participants appreciated the possibility to directly manipulate data and the involvement throughout building their data representation. P32 remarked that she felt more conscious of their data while building the representation:

> *It is funny, being so physically involved with it [my data], it does make me reflect on it a bit more consciously. I was thinking more consciously while building, even though at the beginning I was thinking: 'Oh help how am I going to do this?'. (P32, FREE FORM)*

We observed that using bricks necessitated reflection in participants who undertook the task of building a representation of their blood pressure data which reflected their interpretation of the blood pressure values provided by the Withings BPM. On the other hand, participants who chose to build an abstract representation rarely spoke of engagement during the task. Some participants found the task in the FREE FORM condition too complex:

> *So I can build anything I want? [Silence.] (P39, FREE FORM)*

This particular participant was waiting, reading, doubting, reading again for more than 8 minutes before starting to build something with the bricks. In total it took this participant 22 minutes and 27 seconds to finish the task, while the average engagement time for the other participants in this condition was 6 minutes and 8 seconds.

Construction process. Here, we describe the participants' perception of the choices they made to build their brick representation. Many participants in the FREE FORM condition commented that they had to make a number of design choices before building their data representation. These choices included both the visual design of their data representation as well as how unit bricks would relate (or not) to the numerical measurements provided by the Withings BPM. P36 provided examples of the questions to be answered when building a data representation:

> *I think during the build I was mainly concerned with, which dimensions am I going to use? Am I going to make it [my visualisation] flat, or in height, how am I going to visualize units of say 5? (P36, FREE FORM)*

For some participants, the need to decide on a principle to be used to relate bricks to numerical values was a challenging task. This resulted in focusing solely on the visualisation task and drew attention away from possibilities for reflection. Participants felt disconnected from their data, focusing on the representation:

> *I think I rather thought about how to represent numbers with bricks, instead of relating those numbers to my blood pressure. It disconnected me [. . .] I didn't really interpret, I just tried to visualize [my data]. So, I think it disconnected me from the meaning behind the two numbers. (P40, FREE-FORM)*

In contrast, in the INSTRUCTIONAL condition, where we provided a suggestion on how to build a data representation for blood pressure, users focused on the relative differences between the four numbers to be represented. This often took the form of highlighting differences with a dedicated brick colour or purposely locating brick towers next to each other. Understanding differences was seen as key to data interpretation:

> *This is a much clearer way to compare my measurement, it is easier to grasp. I also think building this will make it easier to remember in the future. I can imagine that I still remember oh yes it was all fine. I will not remember the numeric value of the optimal and my personal measurement. Now I just think oh yeah the bricks are the same height, so that's good. I'm not really concerned with precise numbers anymore. It gets a little bit more relative. (P50, INSTRUCTIONAL)*

The participants adopted a variety of strategies to represent their blood pressure data, depending on how well they felt they could understand the data and what aspects of the results provided by Withings BPM were important to them. This resulted in a number of different physicalisation designs. Next, we study the different final physicalisations built by the participants.

4.2.2 *Final builds and assembly.*

We took pictures of all the data physicalisations built in the FREE FORM and INSTRUCTIONAL conditions. Further, we recorded the process of constructing each physicalisation on video (total length: 314 minutes). Two researchers then open-coded the pictures and videos, focusing on the type of data representation deemed to be the final artefact as well as the process of constructing the data representations. We then built classes of data representations and labels for the ways in which the users assembled the representation in an integrative discussion, merging and grouping codes.

Final data representations. Participants created a variety of data representations, as illustrated in Figure 8. In the INSTRUCTIONAL condition participants were instructed to build four towers, e.g. a tower for each colour, yet, the types of tower representations were diverse. Most participants created linear (uniform) towers, either in a row, in a square, or in parallel pairs of towers (i.e. a pair for the systolic values and vice versa). Seven participants created what we called 'non-linear towers', in which colours were combined, bricks staggered or both. The FREE-FORM condition resulted in a variety of other data representations. Two participants created a *spatial design* based on their measurement and the optimal values; one in a 2D version, the other in a 3D representation. One participant created a scatterplot of the bright pink and dark pink bricks. This participant was not using the numeric values of her measurement or the optimal blood pressure values, but instead built a metaphorical representation of what she imagined her blood looked in her veins:

> *When I think of blood pressure, I think about a blood vessel. And in this blood vessel there are white blood cells [points to the bright pink bricks], and uh, yeah I think it must get really crowded in my blood vessel as my blood pressure goes up. And a little less crowded if my blood pressure would be a little lower. (P37, FREE FORM)*

Two other participants built walls of bricks, while one participant created an *unattached* build, where the bricks were loosely stacked on top of each other, in combination with two small stacks (consisting of 4 blue and 6 bright pink bricks) that were placed perpendicularly on top of the unattached bricks.

Assembly strategies. Participants used several strategies to assemble their data representations. In the INSTRUCTIONAL condition, which made use of the dispenser to present the pre-counted bricks, some participants decided to wait until the dispensers were finished, and counted the bricks it produced before starting. Others started right away with building. During the analysis, we noticed that participants either used a *tower-by-tower* or a *brick-by-brick* strategy for assembling their representation. In the *tower-by-tower* condition participants first finished building one complete tower that represented one value of either their measurement or the norm, before moving on to the next colour. In some of these cases a participant first built the towers on the table before putting the towers on the base plate. In the *brick-by-brick* strategy participants switched

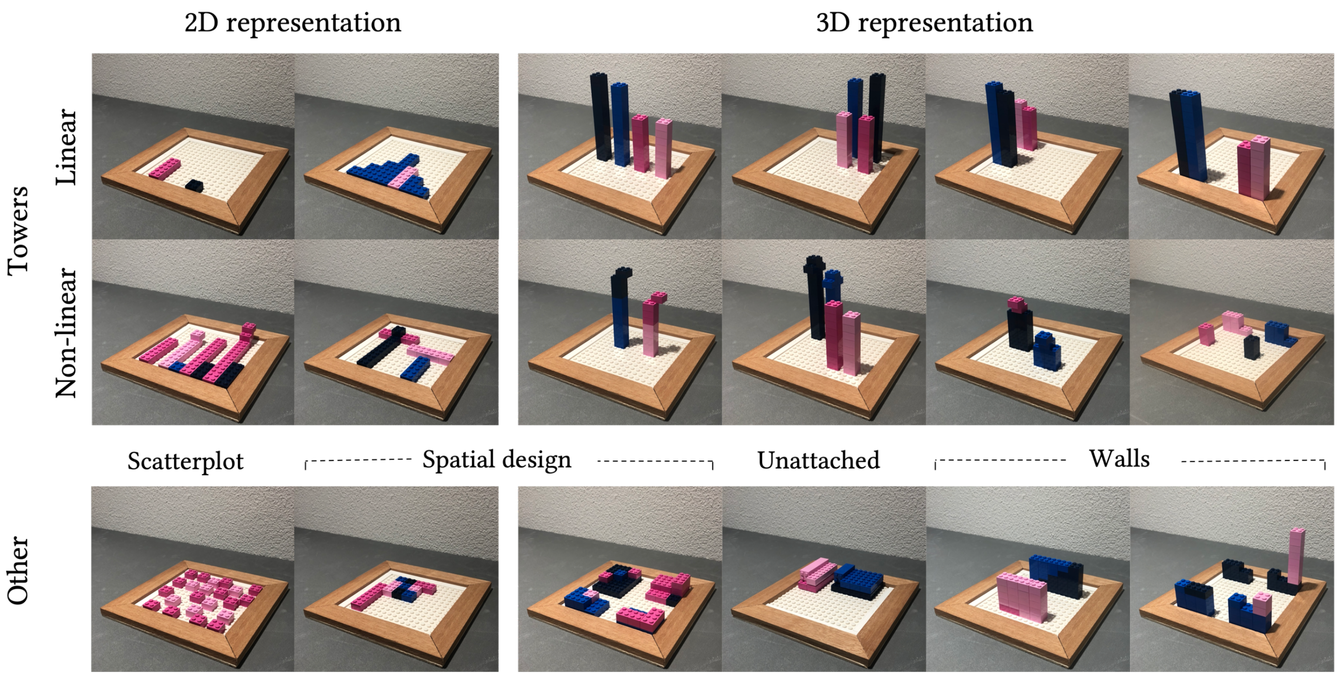

Figure 8: The final data representations that participants created during the session. The pictures are representative examples of participant creations coded under the respective codes in the FREE FORM and INSTRUCTIONAL CONDITIONS. Note that participants were not instructed to build 2D or 3D at any time.

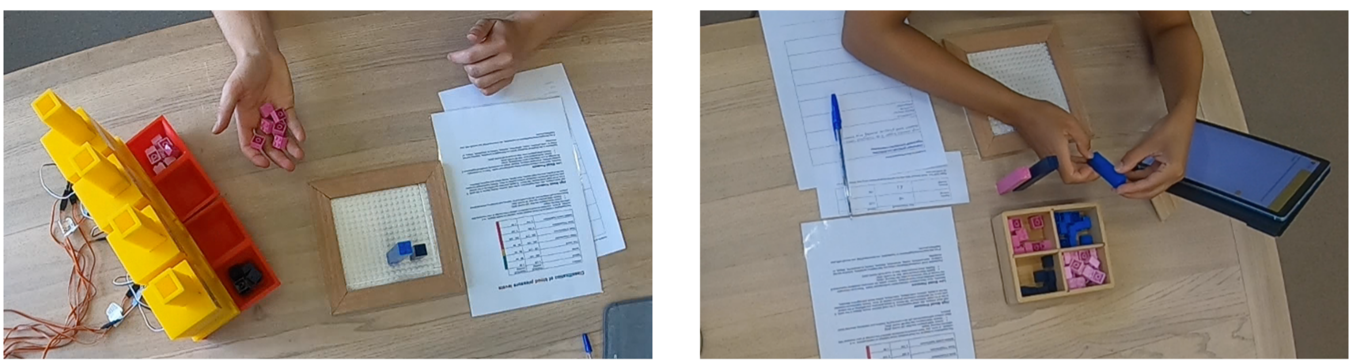

Figure 9: Participants used several strategies for assembling their data physicalisation. Several participants held groups of bricks in their hand to weigh the amount of bricks (as shown on the left), others assembled one tower at a time and placed it on the table before placing them on the baseplate (as shown on the right).

back and forth between the four different colours while constructing their physicalisation. Many participants also held groups of bricks in hand or compared two different groups of bricks in hand, weighing the difference in two palms, as shown in Figure 9.

5 DISCUSSION

Here, we summarise the findings from our study and reflect on the differences between the modes of interacting with personal data representations to understand how they impact engagement and reflection experience.

Our results show that physicalisation of data does not automatically lead to higher levels of reflection. The full TSRI did not show significant differences between the three conditions. However, one of the subscales of the TSRI, i.e. comparison, did show a significant difference; the INSTRUCTIONAL condition outperformed both the BASELINE and the FREE-FORM condition. During the interviews, participants in the INSTRUCTIONAL condition mentioned that it was easier to grasp the relative difference between their measurement and the norm when compared to numeric values, cf. *facilitating reflection*. On the other hand, both the quantitative and qualitative data show that this was not the case for participants in the

FREE-FORM condition. Interestingly, without the dispensers and the instructions participants did not perceive any support from the bricks for comparing their data. Our data shows that participants in the FREE-FORM condition struggled with the cognitively challenging task of mapping the values to bricks and constructing these in such a way that enabled comparison. This struggle was also reflected by their *engagement time*, spending much more time on the task than those in the other two conditions.

Furthermore, there was also no significant difference in perceived user engagement between conditions. There was however a significant difference between the BASELINE condition and the conditions in which participants interacted with bricks for *focused attention*, a subscale of the UES-SF. The bricks allowed participants to slow down and take the time to engage with their data, and therefore facilitated focused attention. In contrast, there was a significant difference in *perceived usability*. Participants considered the BASELINE to be significantly more usable when compared to the FREE-FORM condition. This might have been caused by familiarity bias, as participants were much more used to interacting with a smartphone app than a physicalisation. Yet, the significant difference in engagement time could also form an explanation, as the time it took participants to construct a physicalisation possibly influenced their perception of the usability of the interaction mode.

5.1 Designing instructive physicalisation experiences

Our results indicate that designing physicalisation for personal data reflection involves balancing between offering freedom on the one hand and providing structure and guidance on the other. We note that the need for balancing structure and freedom is similar to the design dilemmas involved in creating technologies for learning, particularly tangible tabletops [11]. The data representations created by the participants, assembled on a table, benefit from 'socio-constructivist flavour' inherent in table-based and tangible interfaces [11, 47]. Constructing a data physicalisation is a cognitively challenging endeavour [25] and only certain forms of a physicalisation allow for effective comparison. The towers that participants built in the INSTRUCTIONAL condition allowed for a visually straightforward way of comparing the four values. Several participants specifically mentioned that the height difference between towers enabled them to easily compare their measurement to the norm, see FACILITATING REFLECTION. In contrast to this, the participants who created a scatterplot or spatial designs as their final data representation (see Figure 8) struggled to compare their measurement to the norm in their creations. As such, the INSTRUCTIONAL condition can be beneficial, as it allows participants to efficiently build a data physicalisation without having to consider the number of bricks needed or having to envision the way in which they can construct their data representation. This however, comes at a cost of expressive freedom, which was also recognized by Thudt et al. [53]. It remains an open question how we can design for an instructive physicalisation experience that offers guidance while also allowing for expressive freedom. Our work shows that there is a spectrum of the level of user control in physicalisation for personal informatics. *Consequently, future personal informatics systems that use physicalisation should use the degree of control over building the physicalisation as a key design choice.*

Our qualitative inquiry shows that users were often focused on *understanding a complex metric*. Thus, the complexity that was studied here was relating the physicalisation to a real-life metric. This is in contrast with studies which investigate how users make sense of complex physical models [42]. While past work in data physicalisation explored models much more complex than the ones in our study, our participants reported thinking intensively despite the low complexity of the data (only four data points featured in our study, which is low compared to the majority of works on physicalisation [43]). This suggests that, in a personal informatics context, the understanding of the concepts underlying the data, and not the numeric complexity of the data per se, is key to how the physicalisation is constructed. Similarly to Thudt et al.'s [53], we also observed that the physicalisation was perceived as a personal representation of self (see *facilitating reflection*). This implies that future designers of data physicalisations for personal informatics will face challenges similar to digital representations, such as the threat of negatively affecting users [37].

5.2 Physicalisation as part of a PI ecology

As opposed to regular data visualisations in mobile applications, constructive data physicalisations require time to obtain even an initial overview of the data. Constructing a physicalisation requires the user to invest time and actively engage with their data. This implies that it may be less suited for everyday use. One of the strengths of personal informatics systems such as fitness trackers is that they automate the tracking of metrics that were previously hidden to a user [32] (e.g. heart rate variability). Manually reconstructing all data in a tangible form would be time-wise impossible and defy the purpose of personal informatics systems, which was also reflected in the lower perceived usability scores for the FREE-FORM condition. At the same time, our study indicates that constructive data physicalisation can benefit the comparison of data and offer users more focused attention on their personal data. Thus, data physicalisation can be a useful tool in the personal informatics ecology, where it should not be seen as a replacement for existing systems, but an addition. In our CONSTRUCTION PROCESS we saw how the act of building a physicalisation prompted users to ask questions about the nature of the data. This implies that physicalisations can be useful in personal informatics processes when a user desires to get a new viewpoint on their data, requiring a dedicated moment for reflection. Past research by Epstein et al. [13] showed that users periodically revise their choice of metrics and tracking practices to assure continued reflection. Our results suggest that physicalisation may facilitate aligning tracking practices with desired outcomes of personal informatics. On the other hand, the fact that an instructional physicalisation for PI does require extra engagement time implies that it will be less useful for users without an established personal tracking routine.

5.3 Ways forward

Based on our results, we highlight possible ways forward for using data physicalisation in a personal informatics context. First, the study looked at the process of data physicalisation in a controlled environment, with a single metric. A challenge for data physicalisation in the real world is that users of personal informatics systems will track a wider variety of metrics that they try to understand and reflect upon. The challenge for HCI is to guide this process in a way that avoids distraction by the possibilities of designing physicalisations. There is a need for developing adaptive guidance of such experiences, especially considering the challenges of designing for facilitated reflection [6]. *The Mediated Reflection Model* (TMRM) highlights two specific challenges of personal informatics systems, i.e., preventing conceptual and temporal mismatches. Data physicalisation offers users the freedom to choose a granularity level (temporal cycle, cf. [6]) of data that is relevant to them (conceptual cycle), yet without guidance about the mapping of data to a tangible form. Thus, identifying a fitting visualisation can form an obstacle to reflection. This issue can be mitigated by designing instructional experiences of data physicalisation in PI.

Second, our work shows that participants appreciated that the interaction with the bricks allowed them to slow down and dedicate time to the activity. We hypothesise that constructing their personal data allowed them to mindfully reflect. As previously mentioned by Rapp and Tirassa [40], personal informatics systems can have a utilitarian perspective on personal data and self-knowledge, in which the focus on self-reflection is meant more as a means of goal pursuit rather than as an end in itself. Our results indicate that the physical interaction with one's data can create a mindful moment that helps people self-regulate their attention [55], enabling a focused reflective interaction with their data, which is in line with previous studies on the use of bullet journals for self-tracking [2, 51]. Data physicalisation offers a possibility to explore personal informatics systems that see self-reflection less as a performance-driven means to an end, but, instead, a way to foster reflection as an activity in itself.

Further, we note that this experiment was designed to ignore social aspects in understanding one's tracking experience. We asked participants to use their thinking and the system to understand their blood pressure data. While the design of our experiment stemmed from the need for a focused, systematic study, future work should investigate the intersection of personal informatics and data physicalisation in a social context. Our work shows that the bricks facilitated comparison even in a socially isolated setting. Social comparison is a recognised means for some users to build meaningful personal informatics experiences [14, 38]. Thus, our results show that there most likely is an exciting design space for social personal informatics data exploration using physicalisations, which goes beyond situated artefacts, e.g. [45]. This, in turn, results in a key challenge for theory development in HCI to reconcile the insights from models of personal informatics [14], understanding reflection in PI [6] and concepts in physicalisation [43]. Our study provides a starting point for a broader inquiry in this direction by showing that effective comparison and instruction are key design elements for PI experiences with physicalisations.

Finally, a consequence of using either data physicalisations with bricks, or self-tracking in bullet journals, is that the data becomes analogue. While this physical aspect of manually moving bricks or drawing in a bullet journal benefits reflection, it complicates the use of such tools in a personal informatics ecology as it creates two separate data streams. There is a need to combine manual and digital self-tracking in order for data physicalisation to become a useful tool in a complex PI experience, which was also recognized by Abtahi et al. [1]. In the case of interaction with bricks, tracking and digitally preserving personal brick-based data representations is a promising avenue for future research.

5.4 Limitations and Future work

We recognise that our study is subject to certain limitations. First, we remark that our choice of using blood pressure data as the subject for reflection in the study has its limitations. This metric was used in all three conditions and was therefore consistent, yet blood pressure is a complex concept and several participants mentioned that the metric was quite abstract to them. We chose this metric because it is less prevalent in commercial personal informatics systems when compared to other metrics such as steps, heart rate or stress levels. None of our participants used a blood pressure monitor at home, preventing the possibility that participants would be reflecting on the same data twice. While there are benefits to using the same metric for all participants (in our study, this allowed for comparisons between participant groups), it could mean that not all participants had a *conceptual match* [6] with this metric. This type of data may be less interesting to some participants than others, which could prevent facilitated reflection. On the other hand, having participants choose a metric themselves would limit our ability to compare between conditions. Consequently, future studies should investigate if the use of less abstract metrics, as well as self-selected data, affects the perceived engagement and reflection of participants. We also note that we chose a specific (commercial) baseline for the experiment and we compared it to two artefacts designed specifically for this study. Studies of this nature can be biased by the design qualities of the artefacts under study [49]. For example, we could have used different colours, which may have offered more contrast but render the representation inaccessible for colour-blind participants. An alternative approach would also be to focus on representations more related to the visualisation used by Withings. In our work, we reduced the complexity of the data and its representations to reduce these biases as much as possible.

Moreover, it remains a question if the use of data physicalisations could also benefit users over a longer period of time. In our study participants used a snapshot of data, and on average engaged with bricks for a little under 6 minutes. We do not know if longer use, or building data physicalisations more often, could support self-reflection or conversely become repetitive. Third, another limitation could be the role of preexisting knowledge. There were participants that were more familiar with the meaning of blood pressure. Some had less knowledge about the meaning of the concept and were trying to come to a deeper understanding of the difference between their *systolic* and *diastolic* values. We did not offer participants any additional explanation for their physiological data as this would be considered medical advice, which would, in turn, render the study ethically inadmissible. Here, we note that our user sample only included participants who did not regularly measure their blood pressure. Our insights are likely limited to users reflecting on blood pressure as discovery and future studies should investigate the potential benefits of physicalisation in long-term health monitoring. In such cases, legacy bias is likely to play a stronger role.

6 CONCLUSION

In this paper, we reported on a study of three ways to explore personal wellbeing data. In a user study, participants engaged with one of three representations of their measured blood pressure data: a standard mobile app, an instructional brick building experience, and building a free-form physicalisation from bricks. We found that the instructional experience facilitated comparison effectively. Further, using bricks required more focused attention from the participants. Our study shows that the degree of control in building a physicalisation is a key design dimension for designing physicalisations for personal informatics. We show the potential for instructional physicalisation experiences to offer dedicated moments for reflection which can improve the personal informatics experience. We hope that our work broadens the scope of inquiry within data representation for personal informatics and inspires future studies into engaging users in personal data journeys with tangible representations.

REFERENCES

[1] Parastoo Abtahi, Victoria Ding, Anna C Yang, Tommy Bruzzese, Alyssa B Romanos, Elizabeth L Murnane, Sean Follmer, and James A Landay. 2020. Understanding physical practices and the role of technology in manual self-tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–24.

[2] Amid Ayobi, Tobias Sonne, Paul Marshall, and Anna L. Cox. 2018. Flexible and Mindful Self-Tracking. (2018), 1–14. https://doi.org/10.1145/3173574.3173602

[3] Eric P.S. Baumer, Vera Khovanskaya, Mark Matthews, Lindsay Reynolds, Victoria Schwanda Sosik, and Geri Gay. 2014. Reviewing reflection: On the use of reflection in interactive system design. *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, DIS* (2014), 93–102. https://doi.org/10.1145/2598510.2598598

[4] Eric P S Baumer. 2015. Reflective Informatics: Conceptual Dimensions for Designing Technologies of Reflection. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, New York, NY, USA, 585–594. https://doi.org/10.1145/2702123.2702234

[5] Marit Bentvelzen, Jasmin Niess, Mikołaj P. Woźniak, and Paweł W. Woźniak. 2021. The Development and Validation of the Technology-Supported Reflection Inventory. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 366, 8 pages. https://doi.org/10.1145/3411764.3445673

[6] Marit Bentvelzen, Jasmin Niess, and Pawel W. Wozniak. 2021. The technology-mediated reflection model: Barriers and assistance in data-driven reflection. In *Conference on Human Factors in Computing Systems - Proceedings (CHI '21)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445505

[7] Marit Bentvelzen, Paweł W. Woźniak, Pia S.F. Herbes, Evropi Stefanidi, and Jasmin Niess. 2022. Revisiting Reflection in HCI: Four Design Resources for Technologies That Support Reflection. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 1, Article 2 (mar 2022), 27 pages. https://doi.org/10.1145/3517233

[8] Ann Blandford, Dominic Furniss, and Stephann Makri. 2016. Qualitative HCI Research: Going Behind the Scenes. *Synthesis Lectures on Human-Centered Informatics* 9 (2016), 1–115. https://doi.org/10.2200/S00706ED1V01Y201602HCI034

[9] Robin Singh Chandel, Sudeepti Sharma, Swapandeep Kaur, Sehijpal Singh, and Raman Kumar. 2022. Smart watches: A review of evolution in bio-medical sector. *Materials Today: Proceedings* 50 (Jan. 2022), 1053–1066. https://doi.org/10.1016/j.matpr.2021.07.460

[10] Eun Kyoung Choe, Bongshin Lee, Haining Zhu, and Nathalie Henry Riche. 2017. Understanding self-reflection: How people reflect on personal data through visual data exploration. In *ACM International Conference Proceeding Series (PervasiveHealth '17)*. Association for Computing Machinery, New York, NY, USA, 173–182. https://doi.org/10.1145/3154862.3154881

[11] Pierre Dillenbourg and Michael Evans. 2011. Interactive tabletops in education. *International Journal of Computer-Supported Collaborative Learning* 6 (2011), 491–514.

[12] Pierre Dragicevic, Yvonne Jansen, Andrew Vande Moere, Pierre Dragicevic, Yvonne Jansen, Andrew Vande, Moere Data, Physicalization Jean, Pierre Dragicevic, Yvonne Jansen, and Andrew Vande Moere. 2019. Data Physicalization To cite this version : HAL Id : hal-02113248. (2019).

[13] Daniel A Epstein, Clara Caldeira, Mayara Costa Figueiredo, Xi Lu, Lucas M Silva, Lucretia Williams, Jong Ho Lee, Qingyang Li, Simran Ahuja, Qiuer Chen, Payam Dowlatyari, Craig Hilby, Sazeda Sultana, Elizabeth V Eikey, and Yunan Chen. 2020. Mapping and Taking Stock of the Personal Informatics Literature. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4 (12 2020). https://doi.org/10.1145/3432231

[14] Daniel A. Epstein, An Ping, James Fogarty, and Sean A. Munson. 2015. A lived informatics model of personal informatics. *UbiComp 2015 - Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2015), 731–742. https://doi.org/10.1145/2750858.2804250

[15] Sebastiao Ferreira. [n. d.]. *Reflecting in and on Action.* Technical Report.

[16] Rowanne Fleck and Geraldine Fitzpatrick. 2010. Reflecting on reflection: Framing a design landscape. *ACM International Conference Proceeding Series* (2010), 216–223. https://doi.org/10.1145/1952222.1952269

[17] Pauline Gourlet and Thierry Dassé. 2017. Cairn: A Tangible Apparatus for Situated Data Collection, Visualization and Analysis. In *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17)*. Association for Computing Machinery, New York, NY, USA, 247–258. https://doi.org/10.1145/3064663.3064794

[18] Anthony M Grant, John Franklin, and Peter Langford. 2002. The Self-Reflection and Insight Scale: A new measure of private self-consciousness. *Social Behavior and Personality* 30, 8 (2002), 821–836. https://doi.org/10.2224/sbp.2002.30.8.821

[19] Kasper Hornbæk. 2013. Some whys and hows of experiments in human–computer interaction. *Foundations and Trends in Human–Computer Interaction* 5, 4 (2013), 299–373.

[20] Dandan Huang, Melanie Tory, Bon Adriel Aseniero, Lyn Bartram, Scott Bateman, Sheelagh Carpendale, Anthony Tang, and Robert Woodbury. 2014. Personal visualization and personal visual analytics. *IEEE Transactions on Visualization and Computer Graphics* 21, 3 (2014), 420–433.

[21] Samuel Huron, Sheelagh Carpendale, Alice Thudt, Anthony Tang, and Michael Mauerer. 2014. Constructive Visualization. In *Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14)*. Association for Computing Machinery, New York, NY, USA, 433–442. https://doi.org/10.1145/2598510.2598566

[22] Samuel Huron, Yvonne Jansen, and Sheelagh Carpendale. 2014. Constructing Visual Representations: Investigating the Use of Tangible Tokens. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 2102–2111. https://doi.org/10.1109/TVCG.2014.2346292

[23] Ellen Isaacs, Artie Konrad, Alan Walendowski, Thomas Lennig, Victoria Hollis, and Steve Whittaker. 2013. Echoes from the past: How technology mediated reflection improves well-being. *Conference on Human Factors in Computing Systems - Proceedings* (2013), 1071–1080. https://doi.org/10.1145/2470654.2466137

[24] Yvonne Jansen, Pierre Dragicevic, Petra Isenberg, Jason Alexander, Abhijit Karnik, Johan Kildal, Sriram Subramanian, and Kasper Hornb\aek. 2015. Opportunities and Challenges for Data Physicalization. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3227–3236. https://doi.org/10.1145/2702123.2702180

[25] Yvonne Jansen, Pierre Dragicevic, Petra Isenberg, Jason Alexander, Abhijit Karnik, Johan Kildal, Sriram Subramanian, and Kasper Hornbæk. 2015. Opportunities and Challenges for Data Physicalization. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3227–3236. https://doi.org/10.1145/2702123.2702180

[26] Rohit Ashok Khot, Larissa Hjorth, and Florian Mueller. 2020. Shelfie: A Framework for Designing Material Representations of Physical Activity Data. *ACM Trans. Comput.-Hum. Interact.* 27, 3 (5 2020). https://doi.org/10.1145/3379539

[27] Rohit Ashok Khot, Larissa Hjorth, and Florian 'Floyd' Mueller. 2014. Understanding physical activity through 3D printed material artifacts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3835–3844. https://doi.org/10.1145/2556288.2557144

[28] Rohit Ashok Khot, Jeewon Lee, Larissa Hjorth, and Florian 'Floyd' Mueller. 2014. SweatAtoms: Understanding Physical Activity Through Material Artifacts. *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (2014), 173–174. https://doi.org/10.1145/2559206.2579479

[29] Rafal Kocielnik, Lillian Xiao, Daniel Avrahami, and Gary Hsieh. 2018. Reflection Companion: A Conversational System for Engaging Users in Reflection on Physical Activity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 2 (7 2018). https://doi.org/10.1145/3214273

[30] Kwangyoung Lee and Hwajung Hong. 2017. Designing for Self-Tracking of Emotion and Experience with Tangible Modality. In *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17)*. Association for Computing Machinery, New York, NY, USA, 465–475. https://doi.org/10.1145/3064663.3064697

[31] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A Stage-Based Model of Personal Informatics Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, New York, NY, USA, 557–566. https://doi.org/10.1145/1753326.1753409

[32] Deborah Lupton. 2016. You are your data: Self-tracking practices and concepts of data. In *Lifelogging*. Springer, 61–79.

[33] Deborah Lupton. 2017. Feeling your data: Touch and making sense of personal digital data. *New Media \& Society* 19, 10 (2017), 1599–1614. https://doi.org/10.

1177/1461444817717515
[34] Natalia Martin-Cruz, Adrian Martin-Gutierrez, and Miguel Rojo-Revenga. 2021. A LEGO® Serious Play activity to help teamwork skills development amongst business students. *International Journal of Research & Method in Education* (2021), 1–16.
[35] Andrew Vande Moere. 2008. Beyond the Tyranny of the Pixel: Exploring the Physicality of Information Visualization. In *Proceedings of the 2008 12th International Conference Information Visualisation (IV '08)*. IEEE Computer Society, USA, 469–474. https://doi.org/10.1109/IV.2008.84
[36] Jennifer Moon. 1999. *Reflection in learning & professional development.* Routledge-Falmer, New York, NY, USA. 229 pages. https://doi.org/10.4324/9780203822296
[37] Jasmin Niess, Kristina Knaving, Alina Kolb, and Paweł W. Woźniak. 2020. Exploring Fitness Tracker Visualisations to Avoid Rumination. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (Oldenburg, Germany) *(MobileHCI '20)*. Association for Computing Machinery, New York, NY, USA, Article 6, 11 pages. https://doi.org/10.1145/3379503.3405662
[38] Jasmin Niess and Paweł W. Woźniak. 2018. Supporting meaningful personal fitness: The tracker goal Evolution Model. *Conference on Human Factors in Computing Systems - Proceedings* 2018-April (2018), 1–12. https://doi.org/10.1145/3173574.3173745
[39] Heather L. O'Brien, Paul Cairns, and Mark Hall. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human Computer Studies* 112, December 2017 (2018), 28–39. https://doi.org/10.1016/j.ijhcs.2018.01.004
[40] Amon Rapp and Maurizio Tirassa. 2017. Know thyself: a theory of the self for personal informatics. *Human–Computer Interaction* 32, 5-6 (2017), 335–380.
[41] Kim Sauvé, Steven Houben, Nicolai Marquardt, Saskia Bakker, Bart Hengeveld, Sarah Gallacher, and Yvonne Rogers. 2017. LOOP: A Physical Artifact to Facilitate Seamless Interaction with Personal Data in Everyday Life. In *Proceedings of the 2017 ACM Conference Companion Publication on Designing Interactive Systems (DIS '17 Companion)*. Association for Computing Machinery, New York, NY, USA, 285–288. https://doi.org/10.1145/3064857.3079175
[42] Kim Sauvé, Argenis Ramirez Gomez, and Steven Houben. 2022. Put a Label On It! Approaches for Constructing and Contextualizing Bar Chart Physicalizations. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 82, 15 pages. https://doi.org/10.1145/3491102.3501952
[43] Kim Sauvé, Miriam Sturdee, and Steven Houben. 2022. Physecology: A Conceptual Framework to Describe Data Physicalizations in their Real-World Context. *ACM Transactions on Computer-Human Interaction* 29, 3 (2022), 1–33.
[44] Kim Sauvé, David Verweij, Jason Alexander, and Steven Houben. 2021. Reconfiguration Strategies with Composite Data Physicalizations. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445746
[45] Kim Sauvé, Saskia Bakker, Nicolai Marquardt, and Steven Houben. 2020. LOOP: Exploring Physicalization of Activity Tracking Data. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (NordiCHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3419249.3420109
[46] Donald A. Schön. 1983. *The Reflective Practitioner.* https://doi.org/10.1542/peds.2005-0209
[47] Orit Shaer, Eva Hornecker, et al. 2010. Tangible user interfaces: past, present, and future directions. *Foundations and Trends in Human–Computer Interaction* 3, 1–2 (2010), 4–137.
[48] Fred Shaffer, Zachary M Meehan, and Christopher L Zerr. 2020. A critical review of ultra-short-term heart rate variability norms research. *Frontiers in neuroscience* 14 (2020), 594880.
[49] Herbert A Simon. 1988. The science of design: Creating the artificial. *Design Issues* (1988), 67–82.
[50] Petr Slovak, Chris Frauenberger, and Geraldine Fitzpatrick. 2017. Reflective practicum: A framework of sensitising concepts to design for transformative reflection. *Conference on Human Factors in Computing Systems - Proceedings* 2017-May (2017), 2696–2707. https://doi.org/10.1145/3025453.3025516
[51] Jakob Tholander and Maria Normark. 2020. Crafting Personal Information - Resistance, Imperfection, and Self-Creation in Bullet Journaling. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376410
[52] Alice Thudt, Uta Hinrichs, Samuel Huron, and Sheelagh Carpendale. 2018. Self-reflection and personal physicalization construction. *Conference on Human Factors in Computing Systems - Proceedings* 2018-April (2018), 1–13. https://doi.org/10.1145/3173574.3173728
[53] Alice Thudt, Uta Hinrichs, Samuel Huron, and Sheelagh Carpendale. 2018. Self-Reflection and Personal Physicalization Construction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173728
[54] Jirar Topouchian, Parounak Zelveian, Zoya Hakobyan, Heghine Gharibyan, and Roland Asmar. 2022. Accuracy of the Withings BPM Connect Device for Self-Blood Pressure Measurements in General Population–Validation According to the Association for the Advancement of Medical Instrumentation/European Society of Hypertension/International Organization for Standardization Universal Standard. *Vascular Health and Risk Management* 18 (2022), 191.
[55] David R Vago and David A Silbersweig. 2012. Self-awareness, self-regulation, and self-transcendence (S-ART): a framework for understanding the neurobiological mechanisms of mindfulness. *Frontiers in human neuroscience* 6 (2012), 296.
[56] Paweł Wozniak, Nitesh Goyal, Przemysław Kucharski, Lars Lischke, Sven Mayer, and Morten Fjeld. 2016. RAMPARTS: Supporting Sensemaking with Spatially-Aware Mobile Interactions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2447–2460. https://doi.org/10.1145/2858036.2858491
[57] Tiffany Wun, Jennifer Payne, Samuel Huron, and Sheelagh Carpendale. 2016. Comparing bar chart authoring with Microsoft Excel and tangible tiles. In *Computer Graphics Forum*, Vol. 35. Wiley Online Library, 111–120.