



Technology which Makes You Think: The Reflection, Rumination and Thought in Technology Scale

Downloaded from: <https://research.chalmers.se>, 2024-07-05 22:51 UTC

Citation for the original published paper (version of record):

Loerakker, M., Niess, J., Woźniak, P. (2024). Technology which Makes You Think: The Reflection, Rumination and Thought in Technology Scale. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8(2). <http://dx.doi.org/10.1145/3659615>

N.B. When citing this work, cite the original published paper.



Technology which Makes You Think: The Reflection, Rumination and Thought in Technology Scale

MEAGAN B. LOERAKKER, Chalmers University of Technology, Sweden

JASMIN NIESS, University of Oslo, Norway

PAWEŁ W. WOŹNIAK, Chalmers University of Technology, Sweden and TU Wien, Austria

Reflection is widely regarded as a key design goal for technologies for well-being. Yet, recent research shows that technologies for reflection may have negative consequences, in the form of rumination, i.e. negative thought cycles. Understanding how technologies support thinking about oneself, which can take the form of rumination and reflection, is key for future well-being technologies. To address this research gap, we developed the *Reflection, Rumination and Thought in Technology* (R2T2) scale. Contrary to past research, R2T2 addresses ways of self-focused thinking beyond reflection. This scale can quantify how a technology supports self-focused thinking and the rumination and reflection aspects of that thinking. We developed the scale through a systematic scale development process. We then evaluated the scale's test-retest reliability along with its concurrent and discriminant validity. R2T2 enables designers and researchers to compare technologies which embrace self-focused thinking and its facets as a design goal.

CCS Concepts: • **HCI design and evaluation methods;**

Additional Key Words and Phrases: Reflection; rumination; self-focused thinking; scale; reflection, rumination and thought in technology scale

ACM Reference Format:

Meagan B. Loerakker, Jasmin Niess, and Paweł W. Woźniak. 2024. Technology which Makes You Think: The Reflection, Rumination and Thought in Technology Scale. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 2, Article 59 (June 2024), 24 pages. <https://doi.org/10.1145/3659615>

1 INTRODUCTION

As technologies become an increasingly prevalent part of our lives, they affect our thinking more and more. In Human-Computer Interaction (HCI), reflection is seen as a key way for technologies to positively affect our thinking, prompting users to analyse information about themselves and their world [6]. Reflection has been identified as a desired quality across application domains, most prominently in Personal Informatics (PI), education, and sustainability [11]. Yet, reflection is not the only way in which technologies can provoke thinking about oneself. Recent work emphasises the risk of technologies triggering negative thought cycles—rumination [31, 78]. Thus, it is a challenge for HCI to understand the ways in which technologies make users think about themselves and how to design to support (or avoid supporting) self-focused thought.

In HCI and beyond, there is an active debate regarding the definitions of reflection, rumination and other related concepts. Some have argued that reflection and rumination lead to opposite consequences (e.g. [75]), whereas others suggest that their concepts are aspects of a more general phenomenon—*self-focused motivation* [95]. Consequently, it remains unclear to what extent we can truly analyse and measure how technologies make users

Authors' Contact Information: [Meagan B. Loerakker](mailto:meagan@chalmers.se), Chalmers University of Technology, Gothenburg, Sweden, meagan@chalmers.se; [Jasmin Niess](mailto:jasminni@uio.no), University of Oslo, Oslo, Norway, jasminni@uio.no; [Paweł W. Woźniak](mailto:pawel.wozniak@tuwien.ac.at), Chalmers University of Technology, Gothenburg, Sweden, pawel.wozniak@chalmers.se and TU Wien, Vienna, Austria, pawel.wozniak@tuwien.ac.at.



This work is licensed under a Creative Commons Attribution International 4.0 License.

© 2024 Copyright held by the owner/author(s).

ACM 2474-9567/2024/6-ART59

<https://doi.org/10.1145/3659615>

Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., Vol. 8, No. 2, Article 59. Publication date: June 2024.

think about themselves. Although some standardised scales used to measure reflection and rumination have been developed in the broader social sciences and used in HCI realm (e.g. the Rumination-Reflection Questionnaire [105], Self-Reflection and Insight Scale [44] and Groningen Reflection Ability Scale [4]), the majority of these scales have been developed to measure a person’s natural predisposition—also called ‘trait’—towards these cognitive processes. In HCI, the Technology-Supported Reflection Inventory [10] was designed specifically to measure reflection, yet it can only be applied to systems designed with a focus on data. Consequently, in order to develop systems which effectively support users in self-focused thinking, there is a need for a measurement instrument which would include reflection, rumination and other facets of how technologies provoke thinking about oneself.

To this end, this paper contributes the development and validation of the *Reflection, Rumination and Thought in Technology* (R2T2) scale, a scale that evaluates to what extent a technical artefact fosters reflection, rumination and self-focused thinking. This paper is structured as follows. First, we explore previous work on the theories of reflection, rumination and other self-focused attention, which form the theoretical underpinnings of R2T2. We then report on the development and validation of the scale. Finally, we discuss the implications of our scale design and possible applications for R2T2.

2 RELATED WORK

In this section, we summarise the prevailing perspectives on reflection and rumination in HCI. Next, we explore past research in HCI and related fields that attempted to design for and measure reflection in technology-related scenarios. We conclude by discussing viewpoints on self-focused thinking, also known as ‘self-focused attention’ in the field of Psychology.

2.1 Reflection and Rumination in Human-Computer Interaction

Reflection and rumination are two cognitive processes that have been researched in a variety of fields, including Psychology, Cognitive Science and Human-Computer Interaction. Here, we discuss reflection and rumination from a multitude of perspectives to inform our adherence to a specific definition for each of the concepts.

The Dual-Process (DP) model, which makes a distinction between Systems 1 and 2, referring to fast and intuitive decision-making, and slow and deliberate decision-making, respectively [37, 38, 51, 52], is a well-known topic both in the mainstream and in the academic world. Cognitive reflection has been thought to be related to people’s propensity to use their System 2 [14, 43, 51, 52, 98]. Reflection and rumination are argued to be akin to one another [105], given that they both refer to cognitive processes entailing ‘thinking about the self’ [76]. Reflection can be a result of curiosity (about oneself) [78]. Hence, it is commonly associated with positive feelings [105]. In the context of personal informatics, Parker [83] defined reflection as a process in which a person’s past, present and future behaviours regarding their health and well-being are considered. Another framework developed by Fleck and Fitzpatrick [41] describes reflection as a process consisting of several levels: R0 (description), R1 (reflective description), R2 (dialogic reflection), R3 (transformative reflection) and R4 (critical reflection).

Contrary to reflection, rumination is considered to be a negative cognitive process, as it can involve a focus on loss and failures [76, 97]. In fact, work by Niess et al. [78] found that negative thought cycles—which is typically used as one of the definitions of *rumination*—can be one of the negative consequences of reflection, especially in the context of technologies that are aimed towards reinforcing reflection. From a Cognitive Science standpoint, prior work suggests that self-focused rumination in response to dysphoric states (e.g. depressed moods, anxiety, or anger), has a multitude of negative consequences, including problem-solving deficits, impaired attention and concentration, and negatively biased thinking [49, 65–68, 80, 84, 88]. Consequently, one’s thoughts and attention are likely to be drawn to negative memories when persistent focus is being drawn to feelings and life circumstances [109], which is likely due to the fact that a person’s mood influences how retrievable similarly framed memories are (e.g. [12, 15, 16, 42, 58]).

While studying reflection has been a recurring theme in HCI, the research community lacks consensus on a clear definition of this concept [6, 8]. For example, Anderson [2] summarises the definition of reflection as a ‘conscious, experientially informed thought, at times involving aspects of evaluation, criticality, and problem-solving, and leading to insight, increased awareness, and/or new understanding’ (p. 480). Whereas reflection is frequently described as a positive cognitive process, rumination is—on the other hand—commonly considered a similar process, but negative. In particular, rumination is described as a thought process that results in negative feelings and even negative actions [97]. Consequently, rumination may lead to goal discrepancy and failure [70]. Reflection and rumination are sometimes also described in terms of ‘directionality’. For example, Eikey et al. [31] argue that reflection has a purpose to the thinking process, whereas rumination does not due to its circular nature. Additionally, Morin [76] provides a classification scheme for so-called ‘self-related terms’, which includes processes like self-reflection and self-rumination. Morin [76] defines rumination as a neurotic self-attentiveness through persistent and negative self-focus, which is typically awakened if a person perceives threats to the self, or experiences loss or injustice.

Due to this lack of consensus, various conceptual and theoretical accounts co-exist within HCI [96]. Most HCI papers use Schön’s [93] concept of reflection-in-action and reflection-on-action [11]. In parallel, past work in HCI applied alternative definitions by Boud [86], Dewey [6, 24], Mezirow [73, 81, 110], or Moon [41], all of which consider reflection as a retrospective process, akin to reflection-on-action. The diversity of approaches illustrates the HCI community’s quest to conceptualise reflection to inform the design of technologies for aiding users’ reflection. In this paper, we explore how HCI can quantify reflection and related processes, contributing a broader understanding of how technology supports reflection and related forms of thinking.

2.2 Self-Focused Attention and Thinking

Similarly to the previous section, we discuss the construct of self-focused attention and self-focused thinking from a cognitive, psychological and HCI standpoint to inform our scale development process.

The construct of self-focused attention originates from social-psychological theory and self-evaluation research [30]. Later, Carver [18] defined this construct as attention that is self-directed, i.e., towards one’s attitudes or previous events, which can take on the form of ‘enhanced awareness of one’s present or past physical behavior, that is, a heightened cognizance of what one is doing or what one is like’ (p. 1255). Notably, the literature has reported on the connection between self-focused attention and negative affect, like clinical disorders, including depression [61, 77, 84], anxiety [20], social anxiety [17], test anxiety [91, 92, 112, 113], and even alcohol abuse [48].

Recent work in HCI noted that the range of thinking about oneself that technology can provoke goes beyond reflection. For example, Eikey et al. [31] illustrated how applications for dieting can trigger negative thought cycles, i.e. rumination. Along similar lines, Niess et al. [78] explored visualisations for tracker goals and found that visualising goal failure can lead to rumination. Consequently, there is a need for HCI to consider the relationship between rumination and reflection. Silvia et al. [95] argue that although reflection and rumination differ in their motivation behind self-focused attention, their similarity is rooted in the person’s heightened inwardly-focused attention. Alternatively, reflection and rumination have also been described as private subconsciousnesses [105]. Trapnell and Campbell [105] argued that it is crucial that the assessment of these self-focused cognitive processes is separate due to the differing reasons and motivations for people to start self-focusing. Although reflection and rumination are self-focused processes, there is a lack of agreement on whether or not they are ‘attention’- or ‘motivation’-focused. For example, the notion that reflection and rumination measure motivation rather than attention is supported by Teasdale and Green [103]. Notably, it has been found that neuroticism is a highly predictive factor for rumination, which is commonly described as an underlying motivation layer to rumination [103]. Rumination is sometimes also described as a ‘tendency’ related to self-focus, and as a result, a predisposition for mental health problems like depression and anxiety [80]. At times, neuroticism is a

predictor for such negative self-focus outcomes [40]. Self-focus is sometimes also referred to as self-consciousness or self-awareness [40]. However, measuring rumination through neuroticism is unnecessarily difficult, hence posing issues with rumination's incremental validity [95].

Self-focus is commonly described as an aversive mechanism, possibly magnifying negative emotional states, but researchers have started to argue that foci on either positive or negative aspects of the self are separate types of processes (e.g. [74]). In fact, some adaptive and beneficial outcomes to self-focused attention can be conceptualised into decision-making and self-regulation processes, hence resulting in increased self-knowledge and psychological adjustment [69, 105]. Trapnell and Campbell [105] distinguish this contradiction through the Rumination-Reflection Questionnaire (RRQ), which defines reflection and rumination as two subtypes of 'private self-consciousness'. Prior work (e.g. [19, 45]) has also tried to quantify self-focus through concepts like 'insight' [44], 'satisfaction' [29], 'self-consciousness' [39] and 'well-being' [89]. Despite the attempt to conceptualise self-focus through those concepts, there is no consensus in the field on how to approach it in applied settings, like HCI and PI. Generally, it remains unclear what definite or alternate factors contribute to self-focus.

2.3 Supporting Reflection and Rumination with Technology

Studying reflection in a technological context is a recurring topic in HCI research, particularly PI. In the Ubicomp community, designing for reflection is often seen as a design goal. In 2020, Epstein et al. [35] conducted a literature review of PI literature. One of the main findings was that HCI literature, including the Ubicomp community, focuses primarily on gaining an understanding of reflection, including how technologies for well-being and health can be designed to support reflection. At the same time, the authors note that the field lacks artefact contributions. Notably, Li et al. [63] argue that Ubicomp technologies, including PI tools, should be designed around the users' self-reflection needs. Through interviews, the authors found that users tend to reflect on their data based on aspects like *Status*, *History*, *Goals*, *Discrepancies*, *Context* and *Factors*. Reflection features prominently in well-being applications. For instance, Kay et al.'s [54] *Lullaby* was an application that combined a touch screen device and sensors to allow a user to reflect on what factors may cause potential disruptions in their sleep. Another example of reflection-supporting work in the Ubicomp community is by Chang et al. [21], who designed *Lunch Line*, a system consisting of a public display and a personal web application to promote healthier eating habits in the workplace. Consolvo et al. [26] designed *UbiFit*, a system employing positive reinforcement to offer opportunities for self-reflection. Each of these works approach the study of reflection in a unique manner, highlighting a need for a standardised measure that can be applied across a wide variety of technologies and implemented in different stages of development.

In Li et al.'s [62] *Stage-Based Model of Personal Informatics Systems*, reflection is seen as a stage that comes after the *preparation*, *data collection* and *integration* phases. Following this reflection stage, users proceed to take action, which marks the final stage in Li's model. Epstein et al. [36] later expanded upon this model with the introduction of the *Lived Informatics Model of Personal Informatics*. While these two models offer a broad overview of the role of reflection in PI, their description of the reflection stage offers limited detail. As pointed out by Baumer et al. [8], Li et al.'s and Epstein et al.'s models assume that reflection is naturally triggered when users access 'prepared, combined and transformed' data. However, this contradicts reflection theories, which emphasise that reflection does not always happen automatically but requires encouragement [96], such as sometimes offered by fitness tracking technologies in the context of achieving fitness goals [79].

Simultaneously, there is a strong body of work that considered designing technological tools for reflection [3, 5, 7, 50]. *MoodMap* is a computer application designed for reflection at the workplace, encouraging users to track their mood and compare it with colleagues [86]. *Trackly* [5] is designed for individual use, aiding patients with Multiple Sclerosis to manually track their symptoms and reflect their data through visualisations. While

Trackly focuses on individualistic reflection, *MoodMap* emphasises the role of social interaction in encouraging reflection. Valkanova et al. [107] in ‘Reveal-it!’ presented an interactive public display for visualising energy consumption data. Users voluntarily input their data, which is then displayed as sunburst charts. Their installation fosters reflection by encouraging comparison to others and creating neighbourhood-wide statistics.

These examples indicate specific qualities of interactive systems that contribute to their efficiency in supporting reflection. While these systems address various application areas, fostering reflection is their overarching quality. Future designs can benefit from comparing prototypes at various design stages to identify the most optimal alternatives for their context of use. Yet, current practice lacks effective methods for evaluating reflection-supporting systems, particularly if these technologies are not centred on data. In our work, we aim to address this gap by developing a validated evaluation method for such systems—a scale.

In summary, past research in HCI suggests that rumination may arise whenever a technology supports reflection. Given that there is no consensus in Psychology on the mechanisms behind these phenomena, there is a need to operationalise reflection and rumination. This can enable a better understanding of how to design technologies that may evoke rumination and/or reflection. In this work, we provide a scale that allows for informed design with reflection and rumination in mind. To account for the varying views on what kind of thinking technology can foster, our scale development process includes reflection, rumination and other kinds of self-focused thinking as candidate factors.

2.4 Measuring Reflection and Rumination

To create a useful reflection measure for HCI, we first explore existing reflection measures in other fields. Various scales have been developed to gauge a person’s reflective capacity. The two most prominent scales are the Self-Reflection and Insight Scale (SRIS) [44] and the Groningen Reflection Ability Scale (GRAS) [4]. The SRIS was developed to help researchers probe socio-cognitive and meta-cognitive processes of individual change [44]. The scale consists of three factors: *engagement in self-reflection*, *need for reflection* and *insight*. Analogously, the GRAS also examines reflection on three dimensions: self-reflection, empathetic reflection, and reflective communication. Self-reflection involves introspectively framing one’s feelings, thoughts, beliefs, and norms. Empathetic reflection extends self-reflection into the social sphere, where one considers others’ perspectives, while reflective communication encompasses both self-reflection and empathetic reflection behaviourally [4]. In 2021, Bentvelzen et al. [10] developed a Technology-Supported Reflection Inventory (TSRI) that provided an assessment tool for data-driven reflection in a technological context. Their scale examines reflection in dimensions of *Insight* – describing the extent the interactive technology offers insight to users, *Exploration* – evaluating ease of use and enjoyment of data exploring, and *Comparison* – measuring the social dimension of comparing oneself to others, providing a holistic overview of qualities of reflection-supporting artefacts.

Drawing from insights in Psychology, both SRIS and GRAS probe a person’s reflective capacity, linking the tendency to experience reflection to one’s personality traits. However, our review of interactive solutions indicates that interactive technologies can enhance user’s reflective capacity. Thus, related work implies certain qualities that determine the effectiveness of interactive technologies in supporting reflection. Nevertheless, these scales were primarily developed with participants from medical and psychological backgrounds, making them most suitable for use in those contexts. Therefore, there is an emerging need for measures that would allow for the evaluation of technology-supported reflection from a system-centric viewpoint. To inform future designs of technologies for reflection, it is essential to develop technology-oriented tools to assess the qualities of interactive artefacts that aim to trigger or foster reflection. A systematically developed, technology-focused instrument, validated using rigorous techniques like confirmatory factor analysis and test-retest procedures, holds the potential to enable more in-depth and appropriate examination of reflection within HCI systems, especially within the domain of personal informatics.

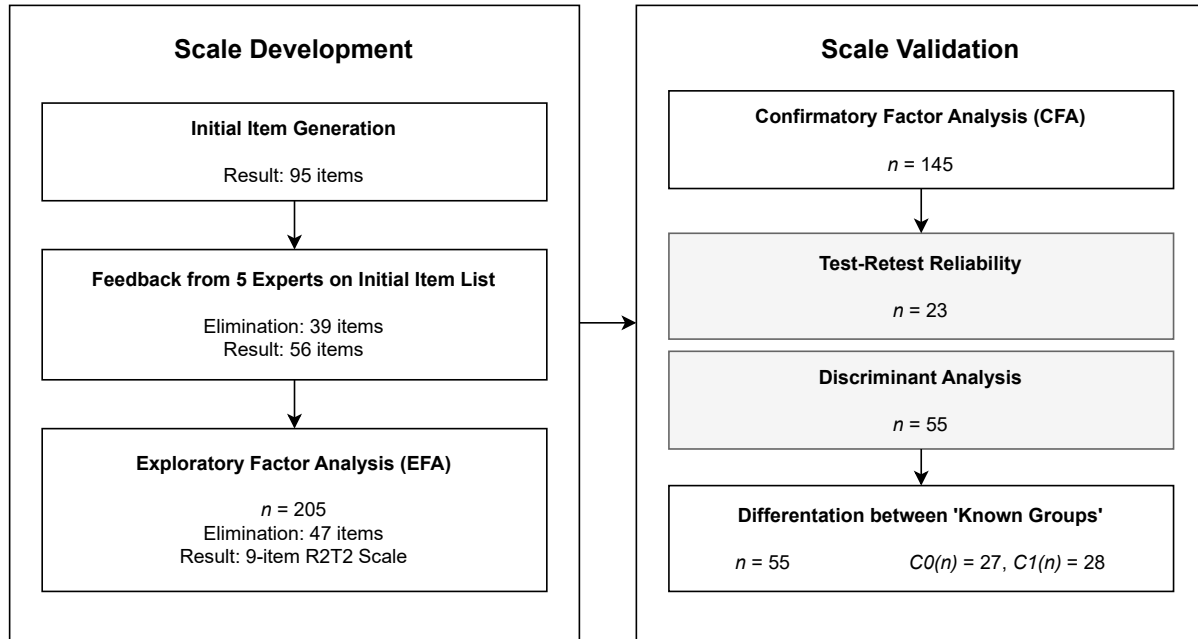


Fig. 1. A detailed overview of the development and validation process for the R2T2 scale. The Discriminant Analysis and Test-Retest Reliability phases are coloured in grey to signify that they were conducted in parallel.

3 METHOD

Our review of related work shows that the concepts of self-focused thinking, reflection, and rumination in interactive artefacts are important to HCI work yet require more systematic understanding. To this end, we decided to develop a structured questionnaire. A validated questionnaire would provide designers and researchers a tool to gauge systems in terms of their capacity for self-focused thinking. Moreover, by selecting appropriate scale items, we can deepen our insights into how users perceive and understand self-focused thinking in these systems.

We employed a systematic approach to craft our scale, drawing inspiration from the methodologies summarised by Boateng et al. [13]. Recent research in HCI contributed a number of scales, which also influenced our method choices. We used a varimax rotation for factor analysis as performed by Mejia and Yarosh [72] and Bentvelzen et al. [10]. Similarly to Bentvelzen et al. [10] and Woźniak et al. [114], we used expert feedback to refine the initial item list by allowing experts to freely comment on the items, and we used Confirmatory Factor Analysis, Differentiation between Known Groups, Discriminant Validity and Test-Retest Reliability as evaluation methods to test the scale's validity and reliability. Figure 1 presents a detailed overview of the scale formation and validation processes. The data collected from both the development and validation stages, the full demographic data of the participants, and the R script to run the analyses are available in the auxiliary material.

The ethics procedures necessary for this study, including participant recruitment, were conducted in compliance with the regulations of the local jurisdiction associated with the first author's affiliation.

4 SCALE FORMATION

In this section, we describe the development of the R2T2 scale, starting with the theoretical background and conceptual considerations to initialise a first list of potential items for the scale and ending with a final list of items for validation.

4.1 Initial Item Generation

In order to generate items to develop a standardised scale, we reviewed literature on reflection, rumination and other related concepts. To generate the items, we used core literature on reflection, as reviewed above, as starting points. Through the exploration of related work, we identified several theories, scales and concepts related to reflection, rumination and self-focused thinking. In particular, we reviewed Trapnell and Campbell's [105] paper on the development of the Rumination-Reflection Scale (RRQ), a commonly used standardised scale in the field of HCI and PI in order to measure a participant's trait reflection and rumination, with the aim to inspect the rationale behind constructing the scale. Similarly, the Self-Reflection and Insight Scale (SRIS) developed by Grant et al. [44] was used both as an inspiration for the scale and as a core paper for our search on reflection theory. The last core paper with the aim to conceptualise reflection and rumination is by Eikey et al. [31].

We further used literature about reflection, focusing on Schön [93], Fleck and Fitzpatrick [41], Leary et al. [59], Epstein et al. [36] and Li et al. [62]. We also considered related concepts like self-compassion (e.g. [90]) to inspire the initial generation of items.

The first author generated the items based on related work from a variety of research fields, including HCI, personal informatics and Psychology. Next, the second and last author, two experienced researchers in HCI and personal informatics with different academic backgrounds, evaluated all the items and provided extensive feedback on them. We iteratively discussed how to narrow the focus of the scale so that it would specifically address interactions with technology. Based on these discussions we defined criteria for items to be considered. This also ensured that we did not consider concepts addressed by existing scales. We ensured that the items did not pertain to personality, habits, how individuals perceive technology, behavioural changes, or connections to other people. Instead, they focused on the immediate use of technology and how it aids thinking processes. It was imperative that they did not specify any particular types of technologies or their applications, like personal data. If any item had two or more subsentences, we split them to maintain clarity. Lastly, we prioritised the use of accessible language to cater to a wide range of individuals. After two feedback rounds, the items were approved by the second and last author. This left us with an initial list of 95 items.

4.2 Expert Reviews

Per Boateng et al.'s [13] guidelines for the item reduction process, we asked HCI and Health Technology experts to provide feedback on this initial list of items. In particular, we recruited research experts who had experience and broad knowledge of technologies that facilitate reflection. As such, these experts were able to judge the relevancy of the items for measuring reflection and related cognitive processes. Furthermore, these experts were able to identify phrases that may be hard to understand for users with varying degrees of lived experience with reflection-enhancing technologies, hence aiding in the process of removing problematic items from the list. The experts were aged 34-58 ($M = 40.40$, $SD = 10.14$), whereby $n = 3$ identified as male, $n = 1$ as female, and $n = 1$ as non-binary. The sample included participants from a variety of academic positions and occupations, including an Assistant Professor, an Associate Professor, a Research Institute Director, and a Research Group Lead. Experts were asked to provide their opinions on all items in an online spreadsheet. Given the conceptual flux in HCI with regard to reflection and rumination, we decided not to take a quantitative approach but rather to gather qualitative feedback. After collecting the feedback from all experts, we adopted a restrictive approach where we eliminated items if they were considered problematic by at least one expert. Additionally, we altered items

accordingly if an expert found the item relevant but provided a suggestion on how to improve it. These revisions resulted in the elimination of 39 items, resulting in a list of 56 items for the next step of the process.

4.3 Exploratory Factor Analysis and Item Reduction

After the expert reviews, we designed an online survey on the Qualtrics XM platform in order to conduct an exploratory factor analysis. The survey asked participants to assess an artefact on the 56 items. The artefact to be assessed was a dashboard of the MyFitnessPal application. With the aim of facilitating reflection and rumination, we included positive, neutral and negative stimuli in the dashboard. These stimuli are grounded in Tversky and Kahneman's [106] so-called 'Framing Effects' by making negative or positive cues more accessible. A person's perception of technology is partially based on their prior experiences, and although framing is recognised as a cognitive bias [106], the inclusion of positive, negative and neutral stimuli offers a range of responses to the artefact. For example, prior work has also shown that positive [22] and negative framing [55] can support reflection. Though, notably, framing has some positive effects, including the cultivation of self-efficacy [22]. For the positive framing, we used the light-based framing approach by incorporating a variety of colours (e.g. [60]). For the negative framing, we avoided using strong colours for ethical purposes. Instead, the non-fulfilment of the goal is displayed at the top of the image with relatively large fonts. Framing is argued to be one of the manners in which we contextualise information, as shown by Epstein et al. [34] and Bentley et al. [9]. Another example is a study by Loerakker et al. [64], who found that framing techniques—in the form of highlighting successes and failures in one's data—significantly influence cognitive processes like self-reflection, rumination, and even self-compassion. For the neutral framing, the cues are made less accessible by putting emphasis on an ordinary pie chart.

Here, we note that there is no consensus in the literature on how many inputs are to be presented to participants during a scale development questionnaire. Ultimately, the goal of the procedure is 'that the scale used generates sufficient variance among respondents for subsequent statistical analysis' [46, p. 972]. Further, the scale developed here is what Torgerson [104] would call a response scale, i.e., a scale where score variation is attributed to both items and subjects [28]; in our case, the participants' character traits will lead to them perceiving the interactive systems differently and past literature has shown that approaches to technology-supported reflection are varied across the population [11]. Past approaches in developing scales that rate a particular artefact included piloting two mobile applications and using one application in the questionnaire for a mobile application scale [99]. An alternative approach was not controlling what was rated in a course rating scale [111]. We note, that the scale development process used here relies on reliability *within* the candidate scale items. This, in turn, that our analysis is reliable as long as the questionnaire results in generating enough variance. We aim to achieve this by using a dashboard which will elicit a range of responses among diverse participants.

Further, the MyFitnessPal app was used as we expect most participants to be familiar with the principles of diet, calorie and health tracking as presented in these types of apps. Further, we chose MyFitnessPal as it was the most used application in the 'Health and Fitness' category across Apple and Google stores at the time of the study.¹ Our choice was further based on previous work by Eikey et al. [31], which showed that diet data could elicit a range of different facets of self-focused thinking. Figure 2 displays the stimuli used. This allowed us to engage a broad audience, including those unfamiliar with personal informatics, in reflection and rumination. The dashboard can induce varied user interpretations, which, would effectively support the process of item reduction. We note that the role of the dashboard in our work is to stimulate responses effectively in order to measure how items correlate and the exact assessment (score) of MyFitnessPal by the participants is irrelevant in terms of item reduction. We collected all data using Qualtrics XM.

¹Determined using <https://www.data.ai/>.

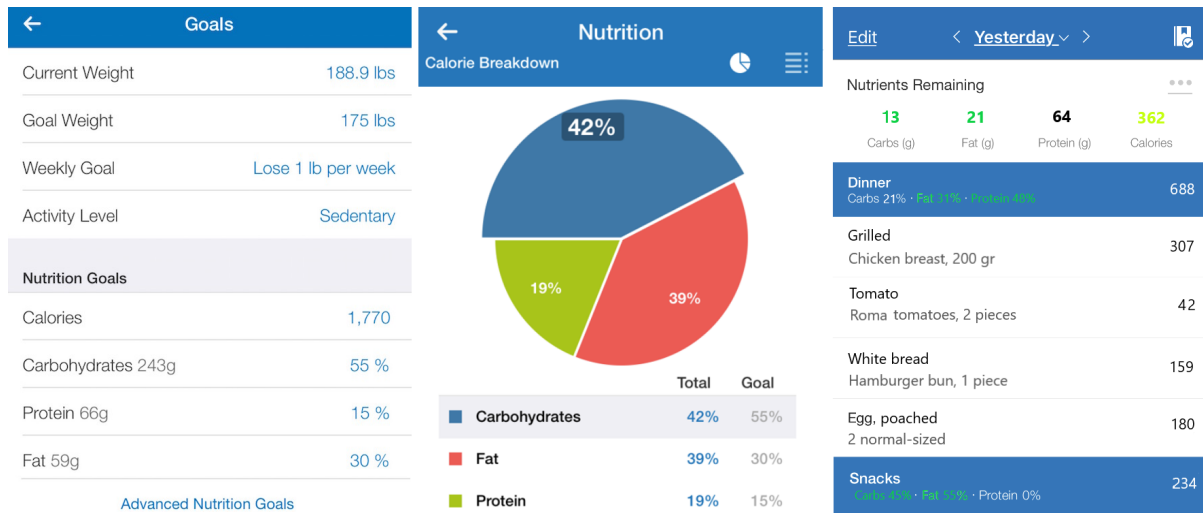


Fig. 2. The MyFitnessPal dashboard used for participants to rate using the R2T2 scale's initial item set. The three pictures in the dashboard each represent a negatively, neutrally and positively framed message. The left picture is negative, as this shows that the user has not reached their weight goal yet. The centre picture is framed neutrally, as the user has reached their macronutrient goal for fats and proteins but not for carbs. The right picture has a positive framing, as the green colours suggest that the user is doing well in getting enough nutrients.

4.3.1 Participants. In total, we recruited $n = 205$ participants, in line with sample size suggestions by Comrey [25], aged from 19 to 64 ($M = 30.49$, $SD = 9.82$) through the online recruitment platform Prolific.² Participants were paid at a rate of 6 GBP per hour, the recommended Prolific rate at the time of the study. In the sample, $n = 90$ participants identified as female, $n = 110$ identified as male, $n = 4$ as non-binary, and 1 participant preferred not to say. Due to the specifics of Prolific, all participants resided in the United Kingdom or the United States.

4.3.2 Results. We followed an iterative item reduction process as described by Boateng et al. [13], using exploratory factor analysis (EFA) with varimax rotation, similar to Mejia and Yarosh [72] and Bentvelzen et al. [10]. Visual inspection of Scree plots indicated that the solution had three or four factors. Yet, a four-factor solution only had one item loading on the fourth factor. Thus, we proceeded with item reduction for a three-factor scale. We removed low-loading items and items which loaded on multiple factors. In doing so, we computed theoretical Cronbach's alphas to optimise reliability. Alpha values were satisfactory ($\alpha > 0.70$ [101]) for the three factors and the full scale. Also, we computed the Tucker Lewis Index for the theoretical correlational three-factor model, $TLI = 0.97$, and the Root Mean Square Error of Approximation of $RMSEA = 0.05$, which conformed to the minimum required parameters [13]. The process resulted in a three-factor scale with three items per factor. In an integrative discussion with the authors, we analysed the items and compared our results to the theoretical background of the scale. We found that the factors corresponded to the concepts of reflection (REF), rumination (RUM), and self-focused thinking (THK). We decided to adopt these factor names for the scale.

5 THE R2T2 SCALE

These final set of dimensions describe the qualities of any technology for a variety of cognitive processes. Table 1 presents an overview of all the items.

²<https://app.prolific.co/>

5.1 Reflection

This dimension represents the degree to which the technology stimulates reflection in the user. In other words, this dimension describes the extent to which a person's positive self-focus is prompted. The items from this dimension are inspired by a variety of works on, i.e., reflection. Item REF1 is inspired by Anderson [2] with their quote: 'reflection is conscious, experientially informed thought, at times involving aspects of evaluation, criticality, and problem-solving, and leading to insight, increased awareness, and/or new understanding' (p. 480). Item REF2 is inspired by Slovák et al.'s [96] conceptual framework of the Social-Emotional Learning (SEL) experience, wherein reflection-in/on-action can take place when a person is given the ability for the 'safe exploration of alternative actions' (p. 2701). Inspiration for item REF3 was taken from Epstein et al.'s [36] Stage-Based Model, which outlines the process of learning from one's behaviour and making changes in one's life: 'We define this practice of *tracking and acting* as the ongoing process of *collecting, integrating, and reflecting*' (p. 735).

5.2 Rumination

The second dimension describes to what extent the technology fosters rumination when a user is exploring and/or interacting with it. Similar to the reflection dimension, the rumination dimension is inspired by works from a wide range of authors. In specific, RUM1 was constructed based on concepts from work by Smith and Alloy [97], whom argue in their work that unhappiness (e.g. stress, sadness) can initiate a self-regulatory cycle, which is also called the 'ruminative cycle' in the field of Psychology. RUM2 was constructed from the following quote: 'rumination often results in stagnation, where individuals think, rethink, and rehearse something without making any progress about or understanding of what that something means for the big picture' [31, p. 604]. The last item from the RUM subscale, RUM3, is a reconstruction of an item from Trapnell and Campbell's [105] Rumination-Reflection Questionnaire: 'I tend to "ruminate" or dwell over things that happen to me for a really long time afterward.'

5.3 Self-Focused Thinking

The last dimension focuses on the measurement of the degree to which a user 'thinks' about the self when using a technology. The items are based on prior work on the SRIS scale [44]. In specific, THK1 is a reconstruction of item 'It is important to me to try to understand what my feelings mean', THK2 is a reconstruction of item 'I usually know why I feel the way I do', and THK3 is a reconstruction of item 'It is important to me to be able to understand how my thoughts arise.' We note that the three items address how technology supports a general understanding of oneself and directs attention to one's feelings and state. Thus, we consider these items to be in line with the concept of self-focused thinking [95].

6 VALIDATING THE R2T2 SCALE

In order to validate our R2T2 scale, we used several standard validation methods, namely Confirmatory Factor Analysis (CFA), Discriminant Validity, Differentiation between 'Known Groups' and testing for Test-Retest Reliability [13]. We first built and validated a bifactor model for the scale, which we later validated with CFA. Having confirmed the validity of the model, we performed the other tests.

6.1 Confirmatory Factor Analysis

We conducted an interactive CFA to find the best model which would fit the factor structure of R2T2. To that end, we first recruited additional participants who completed the R2T2 survey with the nine items only. We note that CFA is not always used in scale development in HCI [72, 100], yet, according to Boateng et al. [13], it is necessary for scale quality. Here we adopt an approach common in the behavioural sciences, where we examine the factor

Table 1. An overview of the R2T2 scale items categorised in the three factors ‘reflection’, ‘rumination’ and ‘self-focused thinking’. For each of these factors, we report their Cronbach’s alpha, and for each item of these dimensions, a factor loading is provided.

<i>The Reflection, Rumination and Thought in Technology (R2T2) scale, $\alpha = 0.73$</i>	
Factor and items	Factor Loading
Reflection, $\alpha = 0.73$	
REF1: Using this technology made me conscious of my behaviours	0.79
REF2: This technology helps me to be able to reflect more easily on my actions	0.62
REF3: This technology supports reflecting on my behaviours as an ongoing activity	0.65
Rumination, $\alpha = 0.75$	
RUM1: This technology can put me in a negative thought cycle	0.73
RUM2: This technology makes me more likely to ruminate about a past situation	0.68
RUM3: Using this technology can make me ruminate or dwell over things that happened for a long time afterward	0.72
Self-Focused Thinking, $\alpha = 0.87$, General factor	
THK1: This technology makes me feel that it is important to me to understand what my feelings mean	0.84
THK2: This technology supports me in usually knowing why I feel the way I do	0.73
THK3: This technology makes me feel that it is important to me to be able to understand how my thoughts arise	0.90

Table 2. Comparison of three theoretical models for R2T2. The bifactor model offered the best performance.

Model	$\Delta\chi^2$	df	p	RMSEA	TLI	CFI	SRMR
Unidimensional				0.23	0.38	0.54	0.18
Three-factor correlational	469.72	27	< .001	0.05	0.91	0.92	0.05
Bifactor	23.89	24	< .001	0.01	0.99	0.99	0.02

structure of a scale by comparing multiple models which fit the theoretical background of the scale. A prior HCI-related example of such a procedure was Vintilă et al. [108]’s work on the smartphone addiction scale.

6.1.1 Participants. For the CFA, we recruited a total of $n = 145$ participants in addition through Prolific. The additional participants were remunerated similarly to those who partook in the EFA. For the sample of $n = 145$ participants, $n = 81$ identified as male, $n = 63$ as female, and 1 participant as non-binary. The participants were aged from 18 to 63, $M = 27.86$, $SD = 8.16$.

6.1.2 Modelling. We computed three models which could constitute the underlying structure of R2T2. The first model represents the naive approach—a unidimensional scale. Second, we built a correlational model with the correlation between all factors, which is the solution most commonly used in HCI scales [72, 114]. Third, we also computed a bifactor model where self-focused thinking is a general factor. This model reflects the theory proposed by Silvia et al. [95] where rumination and reflection are facets of a more general phenomenon. We use chi-square tests to compare pairs of models. We adopted target model parameters from Boateng et al. [13]: $RMSEA \leq 0.06$, $TLI \geq 0.95$, $CFI \geq 0.95$, $SRMR \leq 0.08$. The bifactor model was the only one which conformed to the target criteria, as seen in Table 2. Figure 3 shows the final model with standardised weights.

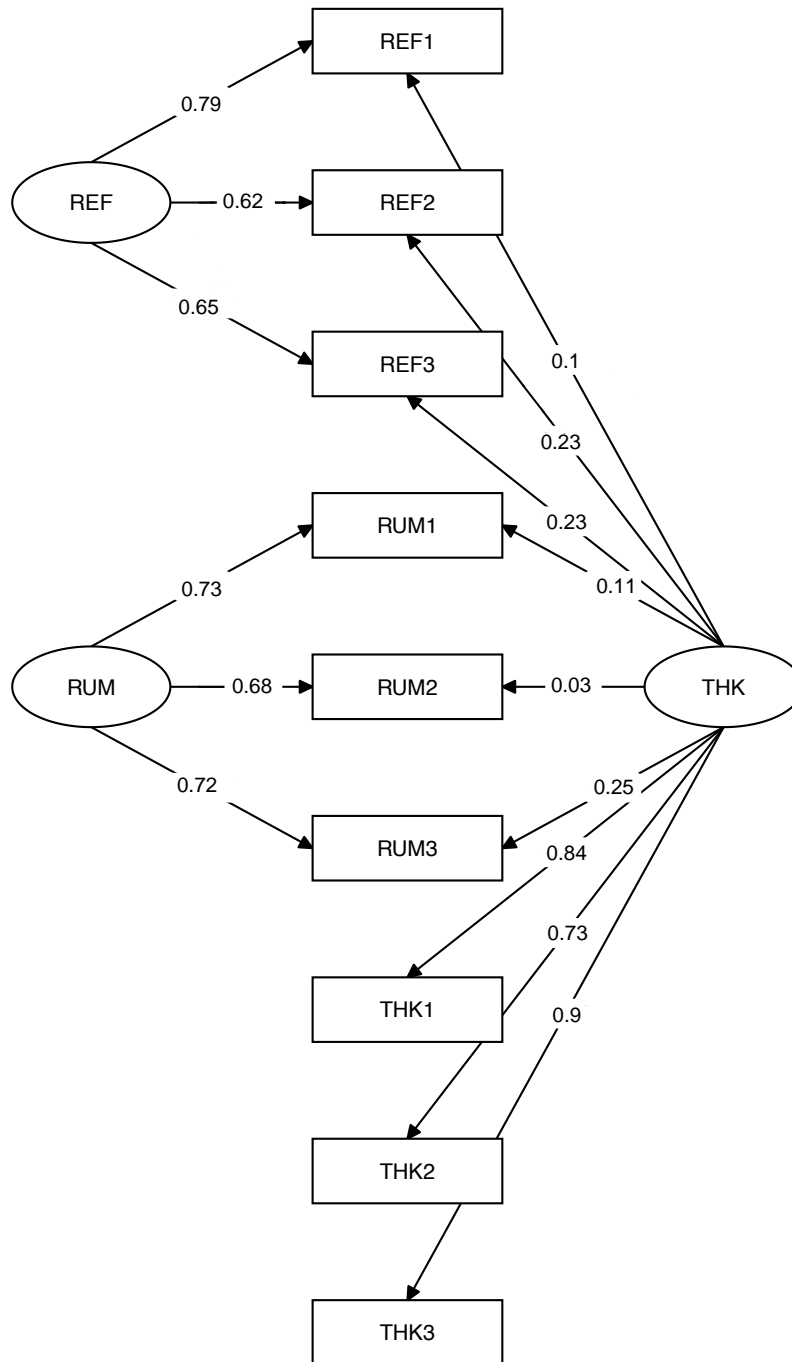


Fig. 3. The bifactor model underlying R2T2. Three factors, reflection (REF), rumination (RUM) and self-focused thinking (THK), describe how technologies can foster thinking about oneself.

6.2 Discriminant Validity

Discriminant validity describes to what degree the scale in question is novel rather than a reflection of another construct or scale [23]. To analyse the scale's Discriminant Validity, we conducted an online survey.

6.2.1 Participants. For the discriminant validity analysis, we recruited $n = 55$ participants through Prolific, which is a similar sample size to previous work (e.g. [10, 114]). Participants were compensated with 6 GBP per hour, as per Prolific's recommendation. The participants were aged 19-68, $M = 28.24$, $SD = 9.33$. In this sample, $n = 19$ identified as female, $n = 35$ as male, and $n = 1$ as non-binary.

6.2.2 Scales for Comparison. We compared our scale to scales that measure similar constructs and/or inspired our development of R2T2. We used the TSRI [10], which is a scale to measure the extent in which a person reflects on their personal data. We also compared R2T2 to the RRQ [105] as it similarly features reflection and rumination as factors. Further, we also collected responses to the short form of the User Engagement Scale (UES-SF) [82] in order to ensure that we can differentiate between self-focused thinking and simply grabbing attention. Finally, we compared R2T2 to the Goal Commitment Scale [47] to ensure that our scale was separate from the complexities of goal setting as it is often intertwined with reflection [79].

6.2.3 Results. To calculate the degree to which R2T2 and the scales used for comparison are correlated, we conducted Spearman correlations for all comparison pairs, in line with Boateng et al. [13]. Table 3 presents a full overview of the correlations between the comparisons.

The results on the R2T2 scale's reflection dimension indicate that the R2T2 is conceptually different from the other scales. There is a moderate correlation between R2T2-REF and the TSRI and two of its subscales. This can be attributed to the fact that the TSRI also measures reflection. Yet, the TSRI is only tailored to data-driven reflection. The larger scope of R2T2 and low correlation with its other factors allow us to conclude that our scale offers adequate discriminant validity compared to the TSRI. We also note weak-to-moderate correlations with scales that measure activities which were found to accompany reflection, especially those measured on the UES scale. This includes focused attention (UEF-FA) [82], reward (UES-RW) [82] and goal completion (GC) [47]. Other research has also suggested that users expect, i.e., a rewarding outcome or experience after having used—or even reflected on—a technology, like an app (e.g. [53]). There was also a moderate correlation with trait rumination (RRQ-RUM) which shows that character traits may influence the R2T2 score. In sum, most of the correlations are low to moderate [85], indicating that R2T2 measures other concepts. The scale does have some overlapping properties with the TSRI, yet it is broader in scope.

6.3 Differentiation between 'Known Groups'

Differentiation between 'known groups'—also known as a comparison between known groups—examines the extent to which a newly developed scale can identify significant differences between binary items, or 'groups', where their difference is found in the variable the scale aims to measure [23]. We evaluated the R2T2 using this method to verify its possibility to differentiate between technologies that do and do not evoke self-focused thinking.

6.3.1 Survey. To conduct this analysis, we created two 30-second videos representing the conditions. For the Control Condition (SPOTIFY), we created an interaction video with the Spotify app interface on the mobile phone, see Figure 4. This type of interaction was chosen as there are no clear stimuli for self-focused thinking. For the Alternative Condition (FITBIT), we created a slide show of Fitbit's smartphone app interface. The slide show mainly focused on stress-related metrics, including stress management scores, and smiley visualisations to indicate a person's stress level and heart rate. Due to the stress-related design and data elements, we hypothesised that this condition would excite more self-focused thinking, including reflection and rumination. Our assumption

Table 3. Correlation matrix for the R2T2 scale and its subscales and other scales from the research field, including the User-Engagement Scale (UES) and its subscales, the Goal Commitment Scale (GS), the Rumination-Reflection Questionnaire (RRQ), and the Technology-Supported Reflection Inventory (TSRI) and its subscales. Asterisks denote p -values: * $p < .05$, ** $p < .01$, *** $p < .001$.

	R2T2	R2T2-REF	R2T2-RUM	R2T2-THK
R2T2-REF	0.50***			
R2T2-RUM	0.75***	0.19		
R2T2-THK	0.76***	0.23	0.29	
UES-SF	0.28	0.28	-0.09	0.40**
UES-FA	0.40**	0.03	0.39**	0.29
UES-PU	0.15	-0.09	0.31	0.05
UES-AE	0.20	0.27	-0.11	0.29
UES-RW	0.22	0.34*	-0.14	0.37**
GC	-0.02	0.37**	-0.19	0.01
RRQ-REF	-0.02	0.08	0.01	0.02
RRQ-RUM	0.19	0.22	0.35**	-0.09
TSRI	0.34*	0.50***	-0.03	0.41**
TSRI-I	0.40**	0.60***	0.01	0.41**
TSRI-E	0.41**	0.62***	0.18	0.29
TSRI-C	0.12	0.11	-0.16	0.32*

was based on a large body of research on fitness tracking and reflection/rumination [36, 78]. In a between-subject online survey in Qualtrics XM, the participants were shown one or the other video. We then asked them to complete the questionnaire with R2T2's items.

6.3.2 Participants. We recruited $n = 55$ participants through Prolific. Similar to previous steps in the process, participants were compensated with 6 GBP per hour. For this comparison analysis, a between-subject study design was used, whereby $n = 24$ participants identified as female and $n = 31$ participants as male, aged 20-54 ($M = 28.24$, $SD = 7.87$).

6.3.3 Results. We conducted a Mann-Whitney U test to investigate if there were significant differences for the R2T2 scores between the two conditions. For the Reflection subscale, we found that $U(54) = 235$, $p < .05$. The result for the Rumination subscale was $U(54) = 313.5$, $p = .28$. For the full R2T2 Scale, the result was $U(54) = 243$, $p < .05$. In summary, we found a significant difference between the two conditions for the full R2T2 scale, including its reflection subscale, but not for its rumination subscale. This shows that R2T2 was effective in detecting the difference between a system prototype that evokes no self-focused thinking and a system tailored for that purpose. The scale effectively showed which aspect of self-focused thinking was relevant in the comparison.

6.4 Test-Retest Reliability

At last, we evaluated R2T2's reliability over time, i.e. temporal stability. The appropriate time interval between two measurement instances for assessing test-retest reliability in scales related to interactive technology is still debated. Marx et al. [71] identified no notable discrepancy between intervals of two days and two weeks. In our research, we employed an interval spanning seven to ten days. We asked participants to rate the MyFitnessPal dashboard (see Figure 2) using R2T2's items.

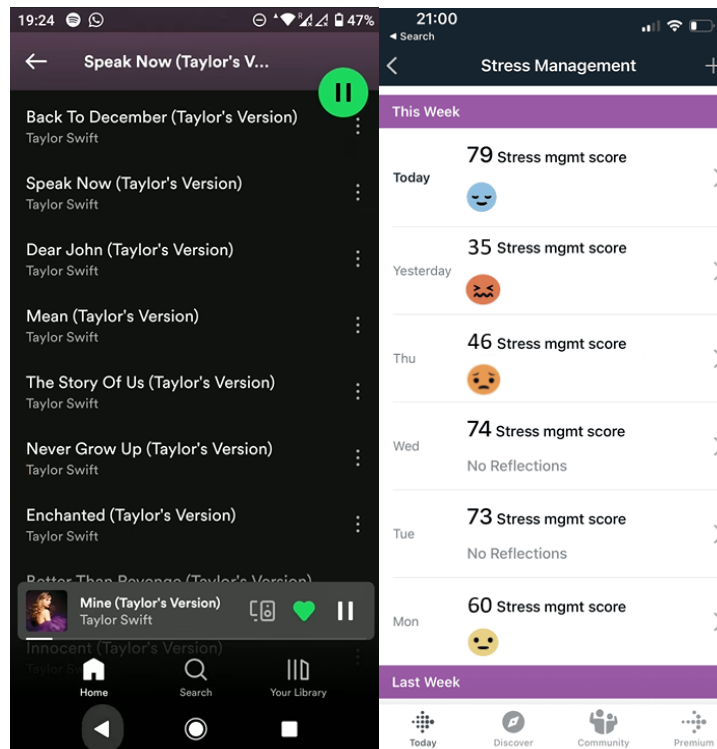


Fig. 4. The control (SPOTIFY) and alternative (FITBIT) conditions used for the Differentiation between ‘Known Groups’ validity, with the left picture representing SPOTIFY and the right picture FITBIT. SPOTIFY is a screenshot of the 30-second video of a Spotify interaction, and FITBIT is a screenshot of the 30-second video of an interaction with the Fitbit app.

6.4.1 Participants. For the test-retest reliability, we recruited $n = 23$ participants, which is a sample size in line with previous scale development work (e.g. [10, 44, 114]). The participants were aged between 19 and 32 ($M = 26.17$, $SD = 3.34$), and $n = 15$ participants identified as female, and $n = 8$ as male. Participants were recruited through convenience and snowball sampling. The sample used for the test-retest reliability analysis only includes participants who completed both questionnaires.

6.4.2 Results. We calculated the intra-class correlation coefficient for fixed raters per Boateng et al.’s [13] suggestion. We used a single rating, absolute-agreement, 2-way mixed-effects model, in line with recommendations by Koo and Li [57]. We obtained intra-class correlations (Fleiss’ kappa) of $\kappa = 0.86$ for the scale (and, per definition, for THK), indicating good to excellent reliability [57]. Subscale results were: $\kappa = 0.71$ for REF, $\kappa = 0.85$ for RUM. These results show that R2T2 maintains temporal stability.

7 DISCUSSION

Here, we first share practical remarks about using the scale and then discuss our results in the context of designing systems that support self-focused thinking.

7.1 Scoring

R2T2 is scored on a five-point Likert scale, whereby (1) represents Strongly Disagree and (5) represents Strongly Agree. Unlike other scales, R2T2 has no reversed items. For the reflection subscale, the sum of the scores represents the extent to which any kind of technology incites reflection. Similarly, the sum of the rumination subscale indicates that the technology stimulates rumination. The third factor, self-focused thinking, constitutes the overall degree to which the person is thinking inward. Self-focused thinking can constitute thoughts about one's actions, thoughts, feelings, and more. Hence, the lowest possible score on the scale is 9, and the highest is 45. In short, R2T2 is scored as follows:

$$R2T2 = R2T2-REF + R2T2-RUM + THK1 + THK2 + THK3$$

$$R2T2-REF = REF1 + REF2 + REF3$$

$$R2T2-RUM = RUM1 + RUM2 + RUM3$$

We adopt this scoring scheme as weighted scoring offers limited benefits according to [13]. Our factor structure suggests that R2T2-REF and R2T2-RUM can be used when reflection is a specifically targeted design goal of a system or rumination is a key concern. We recommend that the researcher remember that there is no THK factor in R2T2, and self-focused thinking can only be assessed using all nine items. This scoring structure reflects a key theoretical principle underlying R2T2: While reflection and rumination can be considered separate phenomena, there is an underlying process of self-focused thinking which is not separable from the two concepts.

7.2 Utility of the R2T2 Scale

Based on our validation, the R2T2 scale will be effective in empirical studies, particularly in comparing two or more systems or system versions. Such a capability makes the R2T2 scale an asset for researchers and designers aiming for rapid prototyping of systems that emphasise self-focused thinking. Since R2T2 is a relatively short scale, it can be easily employed in earlier stages of development for a wide range of technologies and artefacts. While the scale's absolute values do not have a clear interpretation at this stage, there is a potential for these values to become interpretable if the scale is adopted and used within the wider research community.

R2T2 can be used to evaluate systems that utilise reflection in a variety of manners. R2T2 can be used to evaluate to what extent reflection helped in decision-making processes. An example of such a system was developed by Rogers et al. [87], wherein the goal of their installation was to promote taking the stairs over the elevator. In this particular study, participants reported being unaware of the system's influence on their behaviour. R2T2 allows for periodic measurement of thinking about oneself. Furthermore, R2T2 can evaluate systems that employ 'reflection-in-action' in particular, meaning that reflection invoked during a particular interaction with a system can be captured [93]. R2T2's short form allows for reflection measurement when interacting with systems like *Lullaby* [54] that provide real-time feedback on one's habits, potentially through using R2T2 for experience sampling.

In particular, R2T2 supports comparison across various technologies on the broader construct of 'thinking about oneself', and allows for getting a general notion of whether or not a prototype triggers thinking about oneself. It also provides information on whether that thinking is directed towards reflection or rumination. This is particularly relevant as more personal informatics tools are being developed to support reflection as a means towards intentions like increased self-efficacy [1], engagement with the tool [102], and even improved goal-setting towards more feasible future activities [56]. However, considering design for reflection as a means can result in increased observations of 'discouraging' habits (e.g. [56]), potentially resulting in rumination, R2T2 can provide designers and researchers with the means to estimate the technology's influence on a person's overall self-focused thinking, including rumination and reflection. R2T2 enables evaluating interactive systems in PI and beyond in a more holistic way than scales which focus solely on reflection. Compared to other scales, R2T2

provides more fine-grained insight into what parts of the design of a system are effective or could be improved through explicitly identifying three factors in the assessment: reflection, rumination, and self-focused thinking. When comparing systems, R2T2 can help identify specific design elements that contribute to rumination or reflection.

Designed with a technology-agnostic approach, the R2T2 scale can be utilised across an array of contexts and modalities, spanning from mobile applications to tangible interfaces. Thanks to its three-factor structure, the scale not only allows for a broad evaluation of a system but also offers a granular lens to identify specific design qualities of system prototypes. By mapping features to the respective dimensions of reflection, rumination, and self-focused thinking, the R2T2 scale provides an actionable framework for designers to refine and optimise future systems for different facets of self-focused thinking. We encourage researchers to employ R2T2 in a wide range of settings, including studies and workshops, to test R2T2's full capabilities.

7.3 R2T2 and Systems that Evoke Self-Focused Thinking

The development and validation of the R2T2 scale in this paper lead to certain insights for the design and study of systems that evoke thought, especially in areas like personal informatics [35]. As research and practice are likely to continue designing systems that mediate self-awareness and personal growth, the insights from the R2T2 scale provide a nuanced understanding and a tool for more informed, holistic, and user-centric design decisions.

The findings from the R2T2 scale underscore the notion that reflection and rumination should not be singularly pursued or viewed as isolated design objectives. Historically, the design emphasis may have leaned towards cultivating reflection [6] in user experiences, especially with systems meant to help users understand themselves better. However, this study suggests that focusing solely on these objectives might inadvertently overlook other potential impacts or consequences, such as the emergence of rumination or the more general need for self-focused thinking. The structure of R2T2 suggests that a system that is not necessarily perceived as evoking reflection can still offer tangible benefits to the users' thinking processes.

The intertwined relationship between reflection, rumination, and self-focused thinking highlighted by the R2T2 scale is pivotal. This interconnectedness is especially salient when we consider the overarching goal of personal discovery in many systems, e.g. [94]. It is important for future systems to view these phenomena not as isolated entities but as intricately linked processes. This perspective offers a holistic approach to research on systems designed to stimulate introspection and self-awareness, ensuring that users derive meaningful insights without adverse effects.

Lastly, while the quest for fostering reflection in system designs is laudable, our findings indicate that this endeavour is not without risks. Specifically, there is an inherent risk of fostering rumination, which might not always be beneficial for users whenever designing for reflection. The R2T2 scale emerges as a valuable tool in this regard, equipping designers and researchers with a mechanism to gauge and potentially mitigate the unintended consequences of their systems. Balancing reflection with the avoidance of excessive rumination is a nuanced challenge, and the R2T2 scale can assist in navigating this intricate landscape. These findings are in line with studies which earlier recognised the risk of rumination in personal informatics [31, 78].

7.4 Limitations

Our study possesses certain limitations that warrant acknowledgement. First, the participants enlisted for scale development were sourced from Prolific. Such a method of recruitment can introduce a limitation in the scope of the study, particularly concerning geographical diversity and the diversity of specific work conditions. This sampling might not be wholly representative of broader populations or different workplace settings. However, we note that many scales in HCI and beyond are now developed in online studies and later proven to be ecologically valid [13].

Second, we note that the theoretical foundation within the broader social sciences surrounding reflection, rumination, and self-focused thinking is still evolving. Consequently, the R2T2 scale is a pragmatic operationalisation aimed to help HCI researchers and designers include the concepts in their work. Yet, the R2T2 may not fully encapsulate a comprehensive theory regarding these constructs. As research in this domain progresses, it is possible that newer models or theories might provide more nuanced insights or different perspectives, which would then call for redesigning R2T2.

Third, to evaluate R2T2, we used positive, neutral and negative design elements in the MyFitnessPal dashboard. Research suggests that the interpretation of colour can differ on an individual basis, depending on personal preference [27], the context [32] and learned (cultural) associations [33]. Furthermore, prior work found that a person's performance on cognitive tasks can be affected by colour [33]. We remark that scale development requires a questionnaire and if the scale is to be used for assessing artefacts, it needs an artefact to be assessed. We recognise that a different choice of what was presented along the questionnaire may have led to an alternative item choice. However, as the parameters of our final model are satisfactory, we can conclude that the chosen item set holds validity.

Similar to scales like the Technology-Supported Reflection Inventory (TSRI) [10] and the Perceived Creepiness of Technology Scale (PCTS) [114], R2T2 captures a person's judgement—in terms of reflection, rumination and self-focused thinking—of an artefact. In other words, R2T2 measures a general stimulus provided by an external source. On the other hand, scales like the Self-Reflection and Insight Scale (SRIS) [44] and the Rumination-Reflection Questionnaire (RRQ) [105] aim to capture a person's (personality) traits, which are more stable characteristics over time. Considering external stimuli affect cognitive processes like reflection and rumination differently on an individual basis, a general limitation of R2T2 and similar scales is that they measure a person's response to an artefact which is partly determined by the properties of the artefact and partly by the user's character traits (Torgerson [104] calls that a 'response scale'). This inherent limitation of building scales for assessing artefacts can be addressed by using R2T2 in studies which employ within-subject design. Alternatively, trait scale scores can be used as covariates. In the case R2T2, future work should consider using RRQ as a covariate.

A further limitation arises from our expert pool's composition. Comprising three men, one woman, and one non-binary individual, all experienced in research or technology related to reflection, the expert had an influence R2T2 scale's development. Acknowledging this, a broader or differently specialised expert group might have offered divergent perspectives, potentially altering the item composition. Therefore, the current scale may reflect our expert group's specific biases and viewpoints. We hope that, as the scale is used by the research community, it will be verified and potentially refined to better control for such biases.

Lastly, although surveys are a conventional method for scale development and provide quantifiable metrics, it remains uncertain whether the R2T2 scale would be applicable or maintain its validity during longer interactions with artefacts that prompt self-directed thinking. Future research should consider in-depth engagement and experiential studies to test the scale's robustness in different scenarios.

8 CONCLUSION

In this paper, we introduce the *Reflection, Rumination and Thought in Technology* (R2T2) scale, a scale that measures the extent of reflection, rumination and self-focused thinking evoked by a technology. Through related work, we generated a list of items based on a variety of backgrounds, including HCI and Psychology. Based on expert reviews and an exploratory factor analysis, we were able to reduce the list of items to a total of nine items, representing the final R2T2 scale. The nine-item scale consists of the following factors: reflection, rumination, and the general factor of self-focused thinking. Based on the items, we build a bifactor model that accurately describes the scale structure.

We demonstrated the scale's validity with its discriminant validity, its capacity to differentiate between known groups and its consistency over time through test-retest reliability. The scale allows studying a variety of technologies which are designed to either facilitate reflection or diminish rumination. The scale can be applied swiftly to compare or assess technological artefacts. We hope that R2T2 will benefit HCI research and allow for more informed and targeted design of technologies that evoke thinking.

ACKNOWLEDGMENTS

This work was supported by the Swedish Research Council, award number 2022-03196. Paweł W. Woźniak is supported by an endowment from TU Wien.

REFERENCES

- [1] Deemah Alqahtani, Caroline Jay, and Markel Vigo. 2020. The effect of goal moderation on the achievement and satisfaction of physical activity goals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4, Article 116 (12 2020), 18 pages. <https://doi.org/10.1145/3432209>
- [2] Jason Anderson. 2020. Reflection. *ELT Journal* 74, 4 (2020), 480–483. <https://doi.org/10.1093/elt/ccaa039>
- [3] Riku Arakawa and Hiromu Yakura. 2020. INWARD: A computer-supported tool for video-reflection improves efficiency and effectiveness in executive coaching. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI 2020*). Article 574, 13 pages. <https://doi.org/10.1145/3313831.3376703>
- [4] Leo C Aukes, Jelle Geertsma, Janke Cohen-Schotanus, Rein P Zwierstra, and Joris PJ Slaets. 2007. The development of a scale to measure personal reflection in medical practice and education. *Medical Teacher* 29, 2-3 (2007), 177–182. <https://doi.org/10.1080/01421590701299272>
- [5] Amid Ayobi, Paul Marshall, and Anna L. Cox. 2020. Trackly: A customisable and pictorial self-tracking app to support agency in multiple sclerosis self-care. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI 2020*). Association for Computing Machinery, New York, NY, USA, Article 680, 15 pages. <https://doi.org/10.1145/3313831.3376809>
- [6] Eric PS 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* (Seoul, Republic of Korea) (*CHI 2015*). Association for Computing Machinery, New York, NY, USA, 585–594. <https://doi.org/10.1145/2702123.2702234>
- [7] Eric PS Baumer, Sherri Jean Katz, Jill E Freeman, Phil Adams, Amy L Gonzales, John Pollak, Daniela Retelny, Jeff Niederdeppe, Christine M Olson, and Geri K Gay. 2012. Prescriptive persuasion and open-ended social awareness: Expanding the design space of mobile health. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (Seattle, Washington, USA) (*CSCW 2012*). Association for Computing Machinery, New York, NY, USA, 475–484. <https://doi.org/10.1145/2145204.2145279>
- [8] Eric PS 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. In *Proceedings of the 2014 Conference on Designing Interactive Systems* (Vancouver, BC, Canada) (*DIS 2014*). Association for Computing Machinery, New York, NY, USA, 93–102. <https://doi.org/10.1145/2598510.2598598>
- [9] Frank Bentley, Konrad Tollmar, Peter Stephenson, Laura Levy, Brian Jones, Scott Robertson, Ed Price, Richard Catrambone, and Jeff Wilson. 2013. Health Mashups: Presenting statistical patterns between wellbeing data and context in natural language to promote behavior change. *ACM Transactions on Computer-Human Interaction (TOCHI)* 20, 5, Article 30 (11 2013), 27 pages. <https://doi.org/10.1145/2503823>
- [10] 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 2021*). Association for Computing Machinery, New York, NY, USA, Article 366, 8 pages. <https://doi.org/10.1145/3411764.3445673>
- [11] Marit Bentvelzen, Paweł W Woźniak, Pia SF Herbes, Evropi Stefanidi, and Jasmin Niess. 2022. Revisiting reflection in HCI: Four design resources for technologies that support reflection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1, Article 2 (3 2022), 27 pages. <https://doi.org/10.1145/3517233>
- [12] Paul H Blaney. 1986. Affect and memory. *Psychological Bulletin* 89 (1986), 678–682.
- [13] Godfred O Boateng, Torsten B Neilands, Edward A Frongillo, Hugo R Melgar-Quiñonez, and Sera L Young. 2018. Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health* 6, Article 149 (6 2018). <https://doi.org/10.3389/fpubh.2018.00149>
- [14] Ulf Böckenholt. 2012. The Cognitive-Miser Response model: Testing for intuitive and deliberate reasoning. *Psychometrika* 77, 2 (4 2012), 388–399. <https://doi.org/10.1007/s11336-012-9251-y>
- [15] Gordon H Bower. 1981. Mood and memory. *American Psychologist* 36, 2 (2 1981), 129–148. <https://doi.org/10.1037/0003-066X.36.2.129>

- [16] Gordon H Bower. 1991. *The psychology of learning and motivation: Advances in research and theory*. Vol. 27. San Diego, CA: Academic Press, Inc.
- [17] Arnold H Buss. 1980. *Self-consciousness and social anxiety*. San Francisco: W.H. Freeman.
- [18] Charles S Carver. 1979. A cybernetic model of self-attention processes. *Journal of Personality and Social Psychology* 37, 8 (8 1979), 1251–1281. <https://doi.org/10.1037/0022-3514.37.8.1251>
- [19] Charles S Carver and Michael F Scheier. 1978. Self-focusing effects of dispositional self-consciousness, mirror presence, and audience presence. *Journal of Personality and Social Psychology* 36, 3 (1978), 324–332. <https://doi.org/10.1037/0022-3514.36.3.324>
- [20] Charles S Carver and Michael F Scheier. 1986. Functional and dysfunctional responses to anxiety: The interaction between expectancies and self-focused attention. In *R. Schwarzer (Ed.), Self-related Cognitions in Anxiety and Motivation*. Hillsdale, NJ: Erlbaum, 111–141.
- [21] Kerry Shih-Ping Chang, Catalina M Danis, and Robert G Farrell. 2014. Lunch line: Using public displays and mobile devices to encourage healthy eating in an organization. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (Seattle, Washington) (UbiComp 2014)*. Association for Computing Machinery, New York, NY, USA, 823–834. <https://doi.org/10.1145/2632048.2636086>
- [22] Eun Kyoung Choe, Bongshin Lee, Sean Munson, Wanda Pratt, and Julie A Kientz. 2013. Persuasive performance feedback: The effect of framing on self-efficacy. In *AMIA Annual Symposium Proceedings*, Vol. 2013. American Medical Informatics Association, 825–833.
- [23] Gilbert A Churchill, Jr. 1979. A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research* 16, 1 (2 1979), 64–73. <https://doi.org/10.1177/002224377901600110>
- [24] Andrew John Clayphan, Roberto Martinez-Maldonado, and Judy Kay. 2017. A student-facing dashboard for supporting sensemaking about the brainstorm process at a multi-surface space. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction (Brisbane, Queensland, Australia) (OzCHI 2017)*. Association for Computing Machinery, New York, NY, USA, 49–58. <https://doi.org/10.1145/3152771.3152777>
- [25] Andrew L Comrey. 1988. Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology* 56, 5 (1988), 754–761. <https://doi.org/10.1037/0022-006X.56.5.754>
- [26] Sunny Consolvo, Predrag Klasnja, David W McDonald, Daniel Avrahami, Jon Froehlich, Louis LeGrand, Ryan Libby, Keith Mosher, and James A Landay. 2008. Flowers or a robot army? Encouraging awareness & activity with personal, mobile displays. In *Proceedings of the 10th International Conference on Ubiquitous Computing (Seoul, Korea) (UbiComp 2008)*. Association for Computing Machinery, New York, NY, USA, 54–63. <https://doi.org/10.1145/1409635.1409644>
- [27] Roy D'Andrade and M Egan. 1974. The colors of emotion. *American Ethnologist* 1 (2 1974), 49–63. <https://doi.org/10.1525/ae.1974.1.1.02a00030>
- [28] René V Dawis. 1987. Scale construction. *Journal of Counseling Psychology* 34, 4 (1987), 481–489. <https://doi.org/10.1037/0022-0167.34.4.481>
- [29] Ed Diener, Robert A Emmons, Randy J Larsen, and Sharon Griffin. 1985. The Satisfaction With Life Scale. *Journal of Personality Assessment* 49, 1 (1985), 71–75. https://doi.org/10.1207/s15327752jpa4901_13
- [30] Shelley Duval and Robert A Wicklund. 1972. *A theory of objective self-awareness*. New York: Academic Press.
- [31] Elizabeth Victoria Eikey, Clara Marques Caldeira, Mayara Costa Figueiredo, Yunan Chen, Jessica L Borelli, Melissa Mazmanian, and Kai Zheng. 2021. Beyond self-reflection: Introducing the concept of rumination in personal informatics. *Personal and Ubiquitous Computing* 25, 3 (2021), 601–616. <https://doi.org/10.1007/s00779-021-01573-w>
- [32] Andrew J Elliot and Markus A Maier. 2012. Color-in-context theory. In *P. Devine & A. Plant (Eds.), Advances in experimental social psychology*. Vol. 45. Elsevier, 61–125. <https://doi.org/10.1016/B978-0-12-394286-9.00002-0>
- [33] Andrew J Elliot and Markus A Maier. 2014. Color psychology: Effects of perceiving color on psychological functioning in humans. *Annual Review of Psychology* 65 (2014), 95–120. <https://doi.org/10.1146/annurev-psych-010213-115035>
- [34] Daniel A Epstein, Daniel Avrahami, and Jacob T Biehl. 2016. Taking 5: Work-breaks, productivity, and opportunities for personal informatics for knowledge workers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (San Jose, CA, USA) (CHI 2016)*. 673–684. <https://doi.org/10.1145/2858036.2858066>
- [35] 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, Sazedra Sultana, Elizabeth V Eikey, and Yunan Chen. 2020. Mapping and taking stock of the personal informatics literature. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4, Article 126 (12 2020), 38 pages. <https://doi.org/10.1145/3432231>
- [36] Daniel A Epstein, An Ping, James Fogarty, and Sean A Munson. 2015. A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (Osaka, Japan) (UbiComp 2015)*. Association for Computing Machinery, New York, NY, USA, 731–742. <https://doi.org/10.1145/2750858.2804250>
- [37] Jonathan St BT Evans and Keith E Stanovich. 2013. Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science* 8, 3 (2013), 223–241. <https://doi.org/10.1177/1745691612460685>
- [38] Jonathan St BT Evans and Peter C Wason. 1976. Rationalization in a reasoning task. *British Journal of Psychology* 67, 4 (1976), 479–486. <https://doi.org/10.1111/j.2044-8295.1976.tb01536.x>

- [39] Allan Fenigstein, Michael F Scheier, and Arnold H Buss. 1975. Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology* 43, 4 (1975), 522–527. <https://doi.org/10.1037/h0076760>
- [40] Adam K Fetterman and Michael D Robinson. 2012. Interpersonal cognitive self-focus as a function of neuroticism: Basal tendencies and priming effects. *Personality and Individual Differences* 52, 4 (2012), 527–531. <https://doi.org/10.1016/j.paid.2011.11.022>
- [41] Rowanne Fleck and Geraldine Fitzpatrick. 2010. Reflecting on reflection: Framing a design landscape. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction (Brisbane, Australia) (OzCHI 2010)*. Association for Computing Machinery, New York, NY, USA, 216–223. <https://doi.org/10.1145/1952222.1952269>
- [42] Joseph P Forgas. 1991. *Emotion and social judgments (1st edition)*. Oxford, UK: Pergamon Press. <https://doi.org/10.4324/9781003058731>
- [43] Shane Frederick. 2005. Cognitive reflection and decision making. *Journal of Economic Perspectives* 19, 4 (2005), 25–42. <https://doi.org/10.1257/089533005775196732>
- [44] 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>
- [45] Rick Harrington and Donald A Loffredo. 2011. Insight, rumination, and self-reflection as predictors of well-being. *The Journal of Psychology* 145, 1 (2011), 39–57. <https://doi.org/10.1080/00223980.2010.528072>
- [46] Timothy R Hinkin. 1995. A review of scale development practices in the study of organizations. *Journal of Management* 21, 5 (1995), 967–988. <https://doi.org/10.1177/014920639502100509>
- [47] John R Hollenbeck, Charles R Williams, and Howard J Klein. 1989. An empirical examination of the antecedents of commitment to difficult goals. *Journal of Applied Psychology* 74, 1 (1989), 18–23. <https://doi.org/10.1037/0021-9010.74.1.18>
- [48] Jay G Hull. 1981. A self-awareness model of the causes and effects of alcohol consumption. *Journal of Abnormal Psychology* 90, 6 (1981), 586–600. <https://doi.org/10.1037/0021-843X.90.6.586>
- [49] Rick E Ingram. 1990. Self-focused attention in clinical disorders: Review and a conceptual model. *Psychological Bulletin* 107, 2 (1990), 156–176. <https://doi.org/10.1037/0033-2909.107.2.156>
- [50] 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. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Paris, France) (CHI 2013)*. Association for Computing Machinery, New York, NY, USA, 1071–1080. <https://doi.org/10.1145/2470654.2466137>
- [51] Daniel Kahneman and Shane Frederick. 2002. Representativeness revisited: Attribute substitution in intuitive judgment. In *T. Gilovich, D. Griffin & D. Kahneman (Eds.), Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press, 49–81. <https://doi.org/10.1017/CBO9780511808098.004>
- [52] Daniel Kahneman and Shane Frederick. 2005. A model of heuristic judgment. In *K.J. Holyoak & R.G. Morrison (Eds.), The Cambridge Handbook of Thinking and Reasoning*. Cambridge University Press, 267–293.
- [53] Pelin Karaturhan, Ecem Arıkan, Pelin Durak, Asim Evren Yantac, and Kemal Kuscı. 2022. Combining momentary and retrospective self-reflection in a mobile photo-based journaling application. In *Nordic Human-Computer Interaction Conference (NordicCHI 2022)*. Aarhus, Denmark, Article 64, 12 pages. <https://doi.org/10.1145/3546155.3546676>
- [54] Matthew Kay, Eun Kyoung Choe, Jesse Shepherd, Benjamin Greenstein, Nathaniel Watson, Sunny Consolvo, and Julie A Kientz. 2012. Lullaby: A capture & access system for understanding the sleep environment. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing (Pittsburgh, Pennsylvania) (UbiComp 2012)*. Association for Computing Machinery, New York, NY, USA, 226–234. <https://doi.org/10.1145/2370216.2370253>
- [55] Young-Ho Kim, Jae Ho Jeon, Eun Kyoung Choe, Bongshin Lee, KwonHyun Kim, and Jinwook Seo. 2016. TimeAware: Leveraging framing effects to enhance personal productivity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (San Jose, CA, USA) (CHI 2016)*. Association for Computing Machinery, New York, NY, USA, 272–283. <https://doi.org/10.1145/2858036.2858428>
- [56] Rafal Kocielnik, Lillian Xiao, Daniel Avrahami, and Gary Hsieh. 2018. Reflection companion: A conversational system for engaging users in reflection on physical activity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2, Article 70 (7 2018), 26 pages. <https://doi.org/10.1145/3214273>
- [57] Terry K Koo and Mae Y Li. 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine* 15, 2 (2016), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- [58] Julius Kuhl. 1981. Motivational and functional helplessness: The moderating effect of state versus action orientation. *Journal of Personality and Social Psychology* 40, 1 (1981), 155–170. <https://doi.org/10.1037/0022-3514.40.1.155>
- [59] Mark R Leary, Eleanor B Tate, Claire E Adams, Ashley Batts Allen, and Jessica Hancock. 2007. Self-compassion and reactions to unpleasant self-relevant events: The implications of treating oneself kindly. *Journal of Personality and Social Psychology* 92, 5 (2007), 887–904. <https://doi.org/10.1037/0022-3514.92.5.887>
- [60] Andres Ledesma, Mohammed Al-Musawi, and Hannu Nieminen. 2016. Health figures: An open source JavaScript library for health data visualization. *BMC Medical Informatics and Decision Making* 16, 1, Article 38 (2016), 19 pages. <https://doi.org/10.1186/s12911-016-0275-6>
- [61] Peter M Lewinsohn, Harry Hoberman, Linda Teri, and Martin Heutzinger. 1985. An integrative theory of depression. In *S. Reiss & R. Bootzin (Eds.), Theoretical Issues in Behavior Therapy*. Vol. 331–359. New York: Academic Press.

- [62] 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* (Atlanta, Georgia, USA) (*CHI 2010*). Association for Computing Machinery, New York, NY, USA, 557–566. <https://doi.org/10.1145/1753326.1753409>
- [63] Ian Li, Anind K Dey, and Jodi Forlizzi. 2011. Understanding My Data, Myself: Supporting Self-Reflection with Ubicomp Technologies. In *Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp 2011)*. Association for Computing Machinery, New York, NY, USA, 405–414. <https://doi.org/10.1145/2030112.2030166>
- [64] Meagan B Loerakker, Jasmin Niess, Marit Bentvelzen, and Paweł W Woźniak. 2023. Designing data visualisations for self-compassion in personal informatics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 4, Article 169 (12 2023), 22 pages. <https://doi.org/10.1145/3631448>
- [65] Sonja Lyubomirsky, Nicole D Caldwell, and Susan Nolen-Hoeksema. 1998. Effects of ruminative and distracting responses to depressed mood on retrieval of autobiographical memories. *Journal of Personality and Social Psychology* 75, 1 (1998), 166–177. <https://doi.org/10.1037/0022-3514.75.1.166>
- [66] Sonja Lyubomirsky, Fazilet Kasri, and Keri Zehm. 2003. Dysphoric rumination impairs concentration on academic tasks. *Cognitive Therapy and Research* 27, 3 (2003), 309–330. <https://doi.org/10.1023/A:1023918517378>
- [67] Sonja Lyubomirsky and Susan Nolen-Hoeksema. 1995. Effects of self-focused rumination on negative thinking and interpersonal problem solving. *Journal of Personality and Social Psychology* 69, 1 (1995), 176–190.
- [68] Sonja Lyubomirsky, Kari L Tucker, Nicole D Caldwell, and Kimberly Berg. 1999. Why ruminators are poor problem solvers: Clues from the phenomenology of dysphoric rumination. *Journal of Personality and Social Psychology* 77, 5 (1999), 1041–1060. <https://doi.org/10.1037/0022-3514.77.5.1041>
- [69] Leonard L Martin and Abraham Tesser. 1996. Some ruminative thoughts. In R.S. Wyer, Jr. (Ed.), *Ruminative Thoughts: Advances in Social Cognition, Volume IX*. Lawrence Erlbaum Associates, Inc., 1–47.
- [70] Leonard L Martin, Abraham Tesser, and William D McIntosh. 1993. Wanting but not having: The effects of unattained goals on thoughts and feelings. In D.M. Wegner & J.W. Pennebaker (Eds.), *Handbook of Mental Control*. Prentice-Hall, Inc, 552–572.
- [71] Robert G Marx, Alia Menezes, Lois Horovitz, Edward C Jones, and Russell F Warren. 2003. A comparison of two time intervals for test-retest reliability of health status instruments. *Journal of Clinical Epidemiology* 56, 8 (2003), 730–735. [https://doi.org/10.1016/S0895-4356\(03\)00084-2](https://doi.org/10.1016/S0895-4356(03)00084-2)
- [72] Kenya Mejia and Svetlana Yarosh. 2017. A nine-item questionnaire for measuring the social disfordance of mediated social touch technologies. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW, Article 77 (11 2017), 17 pages. <https://doi.org/10.1145/3134712>
- [73] Ine Mols, Elise van den Hoven, and Berry Eggen. 2016. Informing design for reflection: An overview of current everyday practices. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction* (Gothenburg, Sweden) (*NordiCHI 2016*). Association for Computing Machinery, New York, NY, USA, Article 21, 10 pages. <https://doi.org/10.1145/2971485.2971494>
- [74] Nilly Mor and Jennifer Winquist. 2002. Self-focused attention and negative affect: A meta-analysis. *Psychological Bulletin* 128, 4 (2002), 638–662. <https://doi.org/10.1037/0033-2909.128.4.638>
- [75] Alain Morin. 2002. Self-awareness review Part 1: Do you “self-reflect” or “self-ruminate”? *Science & Consciousness Review* 1 (12 2002).
- [76] Alain Morin. 2017. Toward a glossary of self-related terms. *Frontiers in Psychology* 8, Article 280 (2017), 9 pages. <https://doi.org/10.3389/fpsyg.2017.00280>
- [77] Robert F Musson and Lauren B Alloy. 1988. Depression and self-directed attention. L.B. Alloy (Ed.), *Cognitive Processes in Depression* (1988), 193–220.
- [78] 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 2020*). Association for Computing Machinery, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3379503.3405662>
- [79] Jasmin Niess and Paweł W Woźniak. 2018. Supporting meaningful personal fitness: The Tracker Goal Evolution Model. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montréal, QC, Canada) (*CHI 2018*). Association for Computing Machinery, New York, NY, USA, Article 171, 12 pages. <https://doi.org/10.1145/3173574.3173745>
- [80] Susan Nolen-Hoeksema. 2000. The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. *Journal of Abnormal Psychology* 109, 3 (2000), 504–511. <https://doi.org/10.1037/0021-843X.109.3.504>
- [81] Pablo Ortiz and D Fox Harrell. 2018. Enabling critical self-reflection through roleplay with Chimeria:Grayscale. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play* (Melbourne, VIC, Australia) (*CHI PLAY 2018*). Association for Computing Machinery, New York, NY, USA, 353–364. <https://doi.org/10.1145/3242671.3242687>
- [82] 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 (2018), 28–39. <https://doi.org/10.1016/j.ijhcs.2018.01.004>
- [83] Andrea Grimes Parker. 2014. Reflection-through-performance: Personal implications of documenting health behaviors for the collective. *Personal and Ubiquitous Computing* 18 (2014), 1737–1752. <https://doi.org/10.1007/s00779-014-0780-5>

- [84] Tom Pyszczynski and Jeff Greenberg. 1987. Self-regulatory perseveration and the depressive self-focusing style: A self-awareness theory of reactive depression. *Psychological Bulletin* 102, 1 (1987), 122–138. <https://doi.org/10.1037/0033-2909.102.1.122>
- [85] Bruce Ratner. 2009. The correlation coefficient: Its values range between +1/-1, or do they? *Journal of Targeting, Measurement and Analysis for Marketing* 17, 2 (2009), 139–142. <https://doi.org/10.1057/jt.2009.5>
- [86] Verónica Rivera-Pelayo, Angela Fessler, Lars Müller, and Viktoria Pammer. 2017. Introducing mood self-tracking at work: Empirical Insights from Call Centers. *ACM Transactions on Computer-Human Interaction* 24, 1, Article 3 (2 2017), 28 pages. <https://doi.org/10.1145/3014058>
- [87] Yvonne Rogers, William R. Hazlewood, Paul Marshall, Nick Dalton, and Susanna Hertrich. 2010. Ambient influence: Can twinkly lights lure and abstract representations trigger behavioral change?. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing (Copenhagen, Denmark) (UbiComp 2010)*. Association for Computing Machinery, New York, NY, USA, 261–270. <https://doi.org/10.1145/1864349.1864372>
- [88] Cheryl L Rusting and Susan Nolen-Hoeksema. 1998. Regulating responses to anger: Effects of rumination and distraction on angry mood. *Journal of Personality and Social Psychology* 74, 3 (1998), 790–803. <https://doi.org/10.1037/0022-3514.74.3.790>
- [89] Carol D Ryff. 1989. Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology* 57, 6 (1989), 1069–1081. <https://doi.org/10.1037/0022-3514.57.6.1069>
- [90] Gh Samaie and Hojjatollah A Farahani. 2011. Self-compassion as a moderator of the relationship between rumination, self-reflection and stress. *Procedia - Social and Behavioral Sciences* 30 (2011), 978–982. <https://doi.org/10.1016/j.sbspro.2011.10.190>
- [91] Irwin G Sarason. 1975. Anxiety and self-preoccupation. In *I.G. Sarason & C.D. Spielberger (Eds.), Stress and Anxiety*. Vol. 2. New York: Hemisphere Publishing Corporation, 129–151.
- [92] Irwin G Sarason. 1986. Test anxiety, worry, and cognitive interference. In *R. Schwarzer (Ed.), Self-related Cognitions in Anxiety and Motivation*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 19–33.
- [93] Donald A Schön. 1983. *The reflective practitioner: How professionals think in action*. Basic Books.
- [94] Orit Shaer, Oded Nov, Lauren Westendorf, and Madeleine Ball. 2017. Communicating personal genomic information to non-experts: A new frontier for human-computer interaction. *Foundations and Trends® in Human-Computer Interaction* 11, 1 (2017), 1–62. <https://doi.org/10.1561/11000000067>
- [95] Paul J Silvia, Jan Eichstaedt, and Ann G Phillips. 2005. Are rumination and reflection types of self-focused attention? *Personality and Individual Differences* 38, 4 (2005), 871–881. <https://doi.org/10.1016/j.paid.2004.06.011>
- [96] Petr Slovák, Christopher Frauenberger, and Geraldine Fitzpatrick. 2017. Reflective practicum: A framework of sensitising concepts to design for transformative reflection. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (Denver, CO, USA) (CHI 2017)*. Association for Computing Machinery, New York, NY, USA, 2696–2707. <https://doi.org/10.1145/3025453.3025516>
- [97] Jeannette M Smith and Lauren B Alloy. 2009. A roadmap to rumination: A review of the definition, assessment, and conceptualization of this multifaceted construct. *Clinical Psychology Review* 29, 2 (2009), 116–128. <https://doi.org/10.1016/j.cpr.2008.10.003>
- [98] Michael N Stagnaro, Gordon Pennycook, and David G Rand. 2018. Performance on the Cognitive Reflection Test is stable across time. *Judgment and Decision Making* 13, 3 (5 2018), 260–267. <https://doi.org/10.1017/S1930297500007695>
- [99] Stoyan R Stoyanov, Leanne Hides, David J Kavanagh, Oksana Zelenko, Dian Tjondronegoro, and Madhavan Mani. 2015. Mobile app rating scale: A new tool for assessing the quality of health mobile apps. *JMIR mHealth and uHealth* 3, 1 (2015), e27. <https://doi.org/10.2196/mhealth.3422>
- [100] Hyewon Suh, Nina Shahriaree, Eric B Hekler, and Julie A Kientz. 2016. Developing and validating the User Burden Scale: A tool for assessing user burden in computing systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (San Jose, CA, USA) (CHI 2016)*. Association for Computing Machinery, New York, NY, USA, 3988–3999. <https://doi.org/10.1145/2858036.2858448>
- [101] Keith S Taber. 2018. The use of Cronbach’s alpha when developing and reporting research instruments in science education. *Research in Science Education* 48 (2018), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- [102] Lie Ming Tang and Judy Kay. 2017. Harnessing long term physical activity data—How long-term trackers use data and how an adherence-based interface supports new insights. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2, Article 26 (6 2017), 28 pages. <https://doi.org/10.1145/3090091>
- [103] John D Teasdale and Hilary AC Green. 2004. Ruminative self-focus and autobiographical memory. *Personality and Individual Differences* 36, 8 (6 2004), 1933–1943. <https://doi.org/10.1016/j.paid.2003.08.022>
- [104] Warren S Torgerson. 1958. *Theory and methods of scaling*. John Wiley & Sons, Inc.
- [105] Paul D Trapnell and Jennifer D Campbell. 1999. Private self-consciousness and the five-factor model of personality: Distinguishing rumination from reflection. *Journal of Personality and Social Psychology* 76, 2 (1999), 284–304. <https://doi.org/10.1037/0022-3514.76.2.284>
- [106] Amos Tversky and Daniel Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* 211, 4481 (1981), 453–458. <https://doi.org/10.1126/science.7455683>
- [107] 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 (Paris, France) (CHI 2013)*. Association for Computing Machinery, New York, NY, USA, 3461–3470. <https://doi.org/10.1145/2470654.2466476>

- [108] Mona Vintilă, Otilia Ioana Tudorel, Cosmin Goian, and Carmen Bărbat. 2021. Determining the structure of smartphone addiction scale: A bifactor model analysis. *Current Psychology* 40 (2021), 1107–1114. <https://doi.org/10.1007/s12144-018-0035-0>
- [109] Andrew Ward, Sonja Lyubomirsky, Lorie Sousa, and Susan Nolen-Hoeksema. 2003. Can't quite commit: Rumination and uncertainty. *Personality and Social Psychology Bulletin* 29, 1 (2003), 96–107. <https://doi.org/10.1177/0146167202238375>
- [110] Andrew M. Webb, Rhema Linder, Andruid Kerne, Nic Lupfer, Yin Qu, Bryant Poffenberger, and Colton Revia. 2013. Promoting reflection and interpretation in education: Curating rich bookmarks as information composition. In *Proceedings of the 9th ACM Conference on Creativity & Cognition* (Sydney, Australia) (C&C 2013). Association for Computing Machinery, New York, NY, USA, 53–62. <https://doi.org/10.1145/2466627.2466636>
- [111] Keithia L Wilson, Alf Lizzio, and Paul Ramsden. 1997. The development, validation and application of the Course Experience Questionnaire. *Studies in Higher Education* 22, 1 (1997), 33–53. <https://doi.org/10.1080/03075079712331381121>
- [112] Jeri D Wine. 1971. Test anxiety and direction of attention. *Psychological Bulletin* 76, 2 (1971), 92–104. <https://doi.org/10.1037/h0031332>
- [113] Jeri D Wine. 1982. Evaluation anxiety: A cognitive-attentional construct. In *H.W. Krohne & L. Laux (Eds.), Series in Clinical & Community Psychology: Achievement, Stress, & Anxiety*. Washington, DC: Hemisphere Publishing Corporation, 207–219.
- [114] Paweł W. Woźniak, Jakob Karolus, Florian Lang, Caroline Eckerth, Johannes Schöning, Yvonne Rogers, and Jasmin Niess. 2021. Creepy technology: What is it and how do you measure it?. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI 2021). Association for Computing Machinery, New York, NY, USA, Article 719, 13 pages. <https://doi.org/10.1145/3411764.3445299>