Human–Computer Interaction

ISSN: 0737-0024 (Print) 1532-7051 (Online) Journal homepage: http://www.tandfonline.com/loi/hhci20

Personal Informatics, Self-Insight, and Behavior Change: A Critical Review of Current Literature

Elisabeth T. Kersten-van Dijk, Joyce H.D.M. Westerink, Femke Beute & Wijnand A. IJsselsteijn

To cite this article: Elisabeth T. Kersten-van Dijk, Joyce H.D.M. Westerink, Femke Beute & Wijnand A. IJsselsteijn (2017) Personal Informatics, Self-Insight, and Behavior Change: A Critical Review of Current Literature, Human–Computer Interaction, 32:5-6, 268-296, DOI: 10.1080/07370024.2016.1276456

To link to this article: https://doi.org/10.1080/07370024.2016.1276456

HUMAN–COMPUTER INTERACTION, 2017, Volume 32, pp. 268–296
Published with license by Taylor & Francis Group, LLC
ISSN: 0737-0024 print / 1532-7051 online
DOI: 10.1080/07370024.2016.1276456

Personal Informatics, Self-Insight, and Behavior Change: A Critical Review of Current Literature

Elisabeth T. Kersten-van Dijk,[1] Joyce H.D.M. Westerink,[1] Femke Beute,[1] and Wijnand A. IJsselsteijn[1]
[1] *Eindhoven University of Technology, the Netherlands*

Personal informatics (PI) systems allow users to collect and review personally relevant information. The purpose commonly envisioned for these systems is that they provide users with actionable, data-driven self-insight to help them change their behavioral patterns for the better. Here, we review relevant theory as well as empirical evidence for this self-improvement hypothesis. From a corpus of 6,568, only 24 studies met the selection criteria of being a peer-reviewed empirical study reporting on actionable, data-driven insights from PI data, using a "clean" PI system with no other intervention techniques (e.g., additional coaching) on a nonclinical population. First results are promising—many of the selected articles report users gaining actionable insights—but we do note a number of methodological issues that make these results difficult to interpret. We conclude that more work is needed to investigate the self-improvement hypothesis and provide a set of recommendations for future work.

CONTENTS

Elisabeth T. Kersten-van Dijk E.T.v.Dijk@tue.nl) is a Ph.D. candidate at Eindhoven University of Technology, the Netherlands. **Joyce H.D.M. Westerink** (J.H.D.M.Westerink@tue.nl) is a professor at Eindhoven University of Technology, the Netherlands. **Femke Beute** (F.Beute@tue.nl) is a postdoctoral researcher at Eindhoven University of Technology, the Netherlands. **Wijnand A. IJsselsteijn** (W.A.IJsselsteijn@tue.nl) is a professor at Eindhoven University of Technology, the Netherlands.

1. INTRODUCTION

The increasing pervasiveness of sensor-rich smartphones and wearable sensor technologies has facilitated a revolution of self-tracking. A dedicated community of self-trackers and life-loggers seeking "self-knowledge through numbers[1]" has been on the rise. The most visible and well-organized reflection on this community is the Quantified Self (QS) movement—an organization that has regular "meetups" around the globe to share their experiences and experiments in gathering data and collecting insight about themselves (see the QS website; Quantified Self Labs, 2015). But the trend goes beyond these "power users": A recent survey estimated that 69% of Americans keep track of a health-related parameter, either for themselves or a loved one (Fox & Duggan, 2013). And it is not just health, but also physical activity, mood, location, (social) media usage, productivity, and finances that can be and are being tracked with affordable and widely available sensor systems and (free) apps.

[1] See http://quantifiedself.com/.

In tandem with these developments, a new scientific field has emerged, focusing on technology that facilitates collection and use of personally relevant information: personal informatics (PI). In line with the QS motto of "self-knowledge through numbers," the PI community tends to focus on creating systems that not only allow users to gather data but also facilitate favorable changes in behavior. Specifically, these behavior changes are *data driven*: Users obtain (self-)insights by examining their data and subsequently change their behavior based on these insights. We refer to this idea as the *self-improvement hypothesis* of PI. Although there are exceptions (notably life-logging and other recollection- and self-documentation-focused approaches, e.g., Elsden & Kirk, 2014), the self-improvement hypothesis seems to represent the dominant way of thinking about PI and the prevailing intention in designing such systems (see, e.g., Epstein, Ping, Fogarty, & Munson, 2015 ; Li, Dey, & Forlizzi, 2010), as well as the most common reason for users to adopt such systems (Choe, Lee, Lee, Pratt, & Kientz, 2014; Epstein et al., 2015; Li et al., 2010; Whooley, Ploderer, & Gray, 2014).

In spite of the hypothesis's popularity, a systematic assessment of the scientific support for it seems to be missing from the PI landscape. Although self-monitoring of and personalized feedback on certain behaviors is known to be conductive to changes in those behaviors, it is not clear whether these changes in behavior are necessarily mediated by insights based on data (data driven). Other underlying processes may also contribute to the behavior changing power of self-tracking (e.g., the feeling of being observed, interrupting automatic decision making), and understanding how and why certain interventions work is crucial when trying to optimize these interventions (see, e.g., Campbell et al., 2000; Klasnja, Consolvo, & Pratt, 2011). The idea that self-tracking might provide self-insight and that such insights constitute a stepping stone for behavior change is intuitively appealing, but no structural review exists to reveal the current level of empirical evidence for this self-improvement hypothesis. The purposes of this article therefore are to provide such a review and assess the state of the evidence supporting the self-improvement hypothesis. In addition, this review sheds light on current best practices and possible ways forward in the PI field.

In this introduction, we first describe existing models of PI. We then discuss two types of behavior change techniques known from psychological literature (self-monitoring and feedback interventions), highlighting similarities and important differences between these techniques and PI systems. An overview of several behavior change theories and their overlap with the self-improvement hypothesis is then provided. Finally, we introduce our current approach: a structured review of the existing literature on the self-improvement hypothesis of PI, providing an overview of the current state of evidence for this hypothesis, as well as current best practices in evaluating PI systems.

1.1. Models of Personal Informatics

Two models of PI have been proposed: the stage-based model by Li et al. (2010) and the lived informatics model by Epstein et al. (2015). Li et al. (2010) were the first to propose an explicit model of PI. Their stage-based model describes the use of PI systems as a process consisting of five stages: preparation, collection, integration, reflection, and action. In the preparation stage, users decide what they want to track and how. Next comes a period of data collection, followed by an integration process, where the necessary steps are taken to transform the raw data (sometimes from multiple sources) into a coherent result (e.g., visualization) that can subsequently be reflected on. Reflection on the integrated data yields self-insight, which users can subsequently employ to change their behavior in the final stage: action.

In later work, the authors of the stage-based model further refined their model by differentiating between two subtypes of the reflection phase: discovery and maintenance (Li, Dey, & Forlizzi, 2011). Inspired by the transtheoretical model of behavior change (TTM; Prochaska & Velicer, 1997), Li et al. (2011) argued that, although some users may be using PI to identify possible courses of action (i.e., what behaviors to change to reach a certain goal: the discovery phase), others may already have implemented behavior changes and are focused not on initiating new behaviors but on maintaining adherence to an already improved behavioral regime (the maintenance phase).

Based on a survey and interviews with self-trackers, Epstein et al. (2015) recently argued that use of PI systems in practice does not adhere to the strict division of stages described in the stage-based model. They therefore propose an alternative model of PI, based on a lived informatics perspective (Rooksby, Rost, Morrison, & Chalmers, 2014). The lived informatics model highlights the messy reality of lapses and interruptions of tracking activities. In addition, Epstein et al. (2015) emphasized the integrated nature of collection of and reflection on data: Reflection often occurs in tandem with or even during collection, rather than as a strictly separate step. Despite their differences, both the stage-based model and the lived informatics model are consistent with the self-improvement hypothesis: Gathering and inspecting self-relevant data lead to self-insight, which in turn facilitates changes in behavior.

1.2. Self-Monitoring and Feedback Interventions versus PI

In this section we discuss known effects of two techniques similar to PI usage (self-monitoring and feedback interventions) that have been used successfully in the past to change people's behavior.

Self-monitoring (i.e., keeping track of a certain behavior) is a well-known intervention technique, used to change certain behaviors, namely, those behaviors that are being monitored. Effects of self-monitoring are well documented in several domains, including diet monitoring for weight reduction (for a review, see Burke, Wang, & Sevick, 2011) and eco-feedback to reduce energy usage (see, e.g., Buchanan, Russo, & Anderson, 2014; Froehlich, Findlater, & Landay, 2010). There are indications that self-monitoring of a behavior adjusts the frequency of the behavior to better suit perceived (social) norms concerning the behavior: If the behavior is seen as negative, the frequency is reduced, whereas the frequency of a more positive behavior would be increased.

Feedback interventions rely on a similar concept, providing feedback to participants about monitored behaviors with the aim of changing adverse habitual behaviors (the monitoring may or may not be done by participants themselves). For instance, feedback about students' performance is known to improve their study results (Hattie & Timperley, 2007). A recent review concluded that the effectiveness of technology-mediated feedback interventions has not yet been sufficiently examined (Hermsen, Frost, Renes, & Kerkhof, 2016), but more traditional feedback interventions have been extensively evaluated. In a review, Kluger and Denisi (1996) found that the majority of feedback interventions were effective, although it should be noted that for about one third of the interventions the effect was adverse rather than beneficial.

Several possible explanations for the effects of self-monitoring and feedback interventions exist. The effects of self-monitoring are most commonly explained in terms of Bandura's self-regulation theory (Bandura, 1991): Monitored behaviors are compared to some standard or goal, and deviations from the standard are subsequently rectified. From a more behaviorist point of view, self-monitoring may be seen as a form of operant conditioning: The observation of deviations from the norm constitutes a "punishment," discouraging the relevant behavior, whereas achievement of the norm serves as a "reward," reinforcing the relevant behavior. Alternatively, the effectiveness of self-monitoring and feedback interventions may be due (in part) to an "observer effect": The fact that one is observed (even by oneself or a piece of technology) promotes adherence to social norms (e.g., engaging in healthy behaviors). Third, from a perspective of dual-process models (e.g., the elaboration likelihood model; Petty & Cacioppo, 1986) self-monitoring and feedback interventions may serve to interrupt the flow of automatic, habitual decision making, making room for a more controlled and rational decision-making process resulting in better (healthier, more productive) decisions and behaviors. Finally, as in the self-improvement hypothesis, self-monitoring and feedback may result in specific and actionable insights leading to behavior change.

Although self-monitoring and feedback interventions may be effective at changing behavior, it may not necessarily be through data-driven insight that these changes are achieved. Understanding the underlying process, though, is crucial, because each process requires a different approach in terms of (technological) support. For instance, for self-regulation the focus should be on detection and presentation of discrepancies between the current state and goal state, whereas the dual-process perspective calls specifically for active, perhaps even effortful self-monitoring and not automated tracking. This underlines the importance of understanding not just *that* an intervention works but specifically *why* it works (or not). The importance of understanding the underlying process has been highlighted for complex health interventions (Campbell et al., 2000) and specifically for technology directed at behavior change (Klasnja et al., 2011). To understand and optimize the value of PI as a tool for self-reflection and data-driven behavior change, then, we need to examine the proposed underlying process of PI-supported behavior change: Can use of PI promote data-driven self-insight? Can such insight promote behavior change? It is specifically this proposed mediating role of insight that we investigate in this article.

2. THE SELF-IMPROVEMENT HYPOTHESIS: THEORETICAL PERSPECTIVES

The self-improvement hypothesis of PI assumes a two-step process: self-tracking leads to insight, and insight leads to behavior change. Before we move on to our review of empirical evidence for this hypothesis, in this section we discuss relevant theoretical perspectives.

2.1. Insight Through Self-Tracking

The concept of insights in PI seems to be similar to the concept of (visualization) insights in visual analytics, the scientific field concerned with visualizations of data, often with the express purpose of providing insight (e.g. Card, Mackinlay, & Shneiderman, 1999; North, 2006). Although different definitions of insight have been suggested in this context, an examination of the body of work on insight from visual analytics suggests that "insight is considered to be more or less units of knowledge" (Chang, Ziemkiewicz, Green, & Ribarsky, 2009): pieces of information extracted from data.

A specialized field of "personal visual analytics" or "personal visualizations" is now emerging, which focuses specifically on presenting personally relevant information in the most productive way (Huang et al., 2015). However, there are possible barriers to such a presentation having the intended effect of creating actionable insight and subsequent changes in behavior. In his information processing theory, McGuire (1968) described five requirements for a message to eventually achieve behavior-changing effects: exposure, attention, comprehension, yielding, and retention. First, the message must reach the perceptive channels of the audience (exposure). Second, the audience must attend to the message (attention). The message must then be understood (comprehension) and subsequently "yielded to" (i.e., believed). Finally, the change in beliefs must be stable over time (retention) in order for behavior change to occur.

One possible barrier in this process is people's natural propensity toward maintaining their existing beliefs (for a review, see Klayman, 1995). This phenomenon, generally referred to as confirmation bias, may manifest in different ways: People may selectively search for evidence supporting their existing beliefs, for instance, by believing only information that confirms their beliefs, or they may interpret information in a way that suits their beliefs (Klayman, 1995). Such processes may prevent users from "yielding" to the information provided by a PI system if the information is in conflict with their existing (self-)beliefs. So although an entire area of research is now devoted to presenting information in the most effective way, even for the clearest visualizations barriers may exist that could prevent users from obtaining insights from their data.

2.2. Behavior Change Through Insight

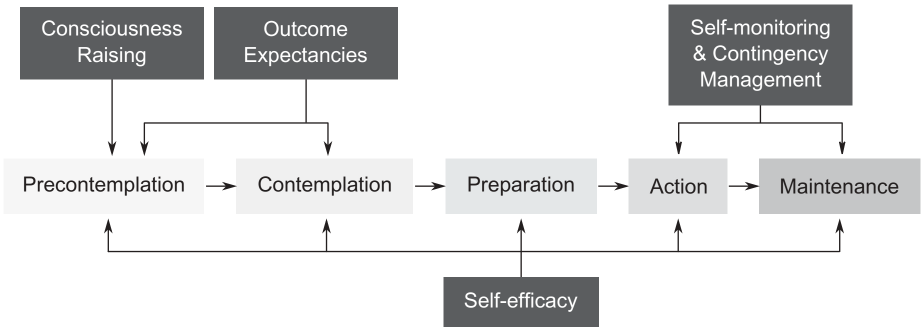
Although information processing theory (McGuire, 1968) proposes that a change in beliefs or knowledge leads to a change in behavior, this relationship may not be that straightforward. Behavior change typically requires more than simply knowing what to do; it has been shown that even if people have a specific intention to perform a certain behavior, they still only have about a 47% chance of actually performing the intended behavior (Sheeran, 2002). Add to this that behavior change is a long-term process requiring many individual behaviors and it is clear that insight is unlikely to translate directly to behavior change. In this section, we turn to models and theories about behavior and behavior change to shed light on determinants of behavior and possible routes through which insight gained through self-tracking might facilitate behavior change.

Both Li et al. (2010, 2011) and Epstein et al. (2015) noted similarities between their models of PI (the stage-based model and the lived informatics model, respectively) and the TTM (Prochaska & Velicer, 1997). The TTM is well suited to application to PI and the self-improvement hypothesis because it deals with sustained behavior change: change as a long-term process of preparation, change, and subsequent maintenance of new behavior(s). In addition, many of the concepts and ideas from the TTM are shared by a variety of other behavior (change) models, like the theory of planned behavior (Ajzen, 1985), the health belief model (Rosenstock, 1974), and various others (Bagozzi, 2000; Bandura, 1977, 1991; De Vries et al., 2003; Fishbein, 2000; Hagger & Chatzisarantis, 2014; Klein, Mogles, & Van Wissen, 2011; Kluger & DeNisi, 1998; Maes & Gebhardt, 2000; Rogers, 1975; Ryan, 2009; Schwarzer, 2008; Witte, 1992). For these reasons, we use the TTM as a starting point in our discussion of existing behavior change models and how they compare to the self-improvement hypothesis.

The TTM divides the process of behavior change into five stages: precontemplation, contemplation, preparation, action, and maintenance. According to the TTM, the process of behavior change starts well before one engages in any new behavior. At first, people have no intention of changing their behavior yet, for instance, because they are not (sufficiently) aware of the problems arising from their current behavior. This is the precontemplation stage. In the contemplation stage, people are aware of the issue, but the perceived disadvantages of changing still outweigh the perceived advantages. When people enter the preparation stage, they have made a decision to change their behavior and are making specific plans to implement their behavior change (e.g., joining a nearby gym). The fourth stage—action—demarcates the beginning of the actual change in behavior. In the action stage, the focus is on implementing changes in behavior, whereas in the final stage—maintenance—the emphasis is on preventing relapse and keeping the new behavior pattern intact.

Besides these "stages of change," the TTM proposes processes of change that can facilitate a person's progression from one stage to the next. Some of these involve the acquisition of new information and insights of the sort that might be obtained through PI (see Figure 1 for an overview).

FIGURE 1. Overview of the stages of change from the transtheoretical model (Prochaska & Velicer, 1997) and information/insight-related processes of change that can aid progress through these stages.



Consciousness Raising

First, the TTM proposes "consciousness raising," a process that increases awareness of problems associated with the existing behavior pattern, which can help a person transition from the precontemplation to the contemplation stage. Awareness can be triggered by knowledge (e.g., about problematic behavior, likely outcomes and possible alternative behaviors; De Vries et al., 2003) and cues that make the goal of behavior change more salient (e.g., an acquaintance having a heart attack; De Vries et al., 2003; Klein et al., 2011).

In some cases, PI might prompt awareness of previously unknown issues. Many PI systems rely on automatic measurements of parameters that are difficult to assess accurately by means of self-observation (e.g., physiological parameters, level of activity, sleep quality). For such parameters, the results obtained through the tracker may surprise users and give them new insights into matters they were not aware of (e.g., "My resting heart rate is higher than I imagined," "I sleep better than I thought").

Outcome Expectancies

A second phenomenon that can facilitate or hinder transition through the stages of change in the TTM is the "decisional balance": the balance of perceived pros and cons of behavior change. According to the TTM, the advantages of behavior change are unclear in the first stage of change (precontemplation), while the disadvantages (i.e., effort) are very salient. To move forward through the stages, the advantages of change need to become clear and need to eventually outweigh the disadvantages of behavior change in order to get to the action and maintenance stages.

PI systems may provide users with information about the benefits of changing a certain behavioral pattern, if the user already engages in both the detrimental behavior and the more favorable behavior. If the user's general pattern of behavior is to perform the detrimental behavior (e.g., eating right before going to bed), but they also sometimes behave in a more favorable way (e.g., eating earlier), a PI system may help them to see the benefits of the more favorable behavior (i.e., better sleep quality). This fits with what Rooksby et al. (2014) referred to as "diagnostic tracking": users keeping track of behaviors and outcomes to determine causal chains.

Self-Efficacy

The third concept that is generally considered an important requirement for behavior change is a person's belief that they are capable of adopting a new pattern of behavior: self-efficacy. Originally proposed by Bandura (1977), self-efficacy and similar concepts like "perceived behavioral control" and "perceived competence" are now incorporated in many theories of behavior and behavior change (e.g. Bagozzi, 2000; De Vries et al., 2003; Fishbein, 2000; Hagger & Chatzisarantis, 2014; Klein et al., 2011; Maes & Gebhardt, 2000; Rogers, 1975; Ryan, 2009; Schwarzer, 2008).

Self-efficacy can be influenced by different sources of information. Perhaps the most obvious is information about past performance accomplishments (Bandura, 1977): If individuals succeeded at something before, they are likely to feel more confident that they can succeed again. PI may help in this area by making (small) accomplishments more visible to users, reminding them that they are capable of "doing the right thing."

Contingency Management and Self-Monitoring

In the final two stages of behavior change—action and maintenance—the TTM proposes "contingency management" as a crucial process: a person needs to keep track of obstacles and deal with them as they arise. Monitoring adherence, lapses, and progress toward goals is needed so that adjustments to the regime can be made where necessary.

PI systems may facilitate the processes just described by providing feedback about the user's progress. Bandura (1977) noted that feedback needs to be informative and temporally near (i.e., immediate) to the relevant behavior in order to be effective. The myriad of sensors available, coupled with mobile technology, mean that PI systems have the opportunity to provide both.

2.3. Reviewing Empirical Evidence for the Self-Improvement Hypothesis

The theoretical perspectives just outlined suggest that although barriers may exist, there are reasons to believe that PI systems might facilitate insight and that several types of insight may contribute to behavior change in different ways. The question remains, however, whether empirical work on PI systems confirms these notions. In this review, we look at empirical evidence for the self-improvement hypothesis. For reasons just outlined, we are specifically interested not only in the outcome (behavior change) but also in the underlying process: Do users obtain self-insights through use of PI systems? Are these insights that (may) help them to change their behavior?

As the weight of empirical evidence very much depends on the quality of the study methodology, in the current review we also assess best practices in evaluation of PI systems. We highlight common shortcomings and potential pitfalls and, based on our findings, provide recommendations for future empirical studies of PI systems.

3. METHOD

Four scientific databases were searched for relevant work: PsycINFO, the ACM digital library, Scopus, and ScienceDirect. The same query was used in all four databases. Entries were included if either the title or abstract contained the following: [PI keywords] OR ([self-monitoring keywords] AND [technology keywords]). See Figure 2 for the individual keywords. The selection of keywords for this review was a challenge, as not all publications relevant to our search necessarily self-identify as "personal informatics" or "self-tracking" literature. The keywords were therefore selected based on the authors' knowledge of the field: We devised a set of keywords that captured relevant publications we were aware of while avoiding keywords that are so general that they produce unmanageable volumes of results. Specifically, we combined keywords focused on technology (e.g., "application") with keywords

FIGURE 2. The keywords used to query the four databases.

PI keywords	self-tracking, self-track, self tracking, self-hacking, self-hack, self hacking, quantified self, life-logging, lifelogging, life logging, personal informatics, personal analytics, e-coach, ecoach, e-coaching, ecoaching, smart coach, smart coaching, personal visualization, personal visualisation, personal visual analytics, auto-analytics, auto analytics
self-monitoring keywords	self-assessment, self assessment, self-assess, self-monitoring, self monitoring, self-monitor, self-surveillance, self surveillance, [self, own, self-management, self-regulation *combined with* feedback]
technology keywords	technology, platform, application, mobile

***Note*. Entries were included if either the title or abstract contained the following: [PI keywords] OR ([self-monitoring keywords] AND [technology keywords]). PI = personal informatics.**

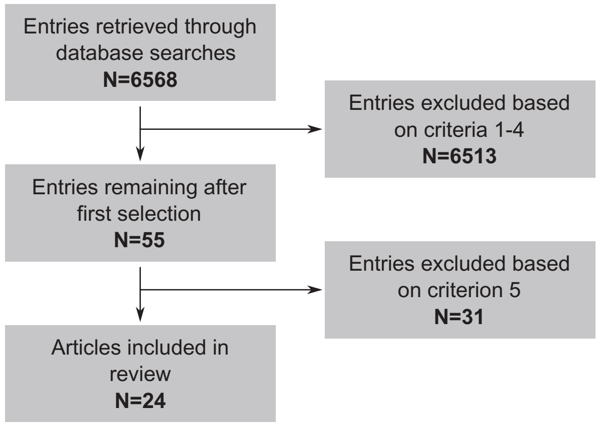
related to a focus on self-insight (e.g., "self-monitoring") to target relevant literature without automatically also including the entire body of work on computer science.

The search was conducted on January 26, 2016. In the case of Scopus, the query used produced such a large set of results (> 7,000 entries) that entries were further selected based on research area, keeping only work from relevant fields (psychology, decision sciences, social sciences, medicine, engineering, computer science, and multidisciplinary work) while excluding work from fields not relevant to our research questions (e.g., neuroscience, dentistry, physics). Across the four databases, the initial search resulted in 6,568 entries. These were narrowed further by manual inspection of titles, abstracts, and—where needed—full texts. For this selection process, five inclusion criteria were used. These criteria, as well as the motivation behind them, can be found in Figure 3.

FIGURE 3. The inclusion criteria used to select the articles for this review.

1	The entry must be peer-reviewed and report on original empirical work.
2	The entry must discuss one or more technologies that support collection of, and feedback on, personally relevant data from daily life; the goal of facilitating insight to promote data-driven behavior change must be explicitly mentioned. Note: this criterion ensures relevance to Personal Informatics and the Self-Improvement Hypothesis. Work on self-monitoring interventions that do not explicitly discuss self-insight or self-knowledge as a mediator are excluded; work that deals with feedback obtained exclusively in a lab setting is excluded.
3	The technology/technologies discussed may not contain or be accompanied by other intervention techniques besides technology-mediated self-monitoring and feedback. Note: evaluating complex interventions and applications as a whole makes it difficult to tell what the effects of certain parts of the intervention (e.g. self-monitoring) are. Work that discusses a complex intervention or application that combines self-monitoring with other techniques like coaching sessions, provision of information, tips and advice and social comparison or support is excluded.
4	Intended users and evaluation participants of the technology/technologies discussed in the entry must be part of the 'general public', not a clinical or extreme group. Note: in some groups, important factors like motivation are likely to be different compared to the general public, making results difficult to generalize. Work on, for instance, self-monitoring of blood glucose levels for diabetics is excluded, as is work on physiological monitoring for patients recovering from a heart attack or stroke.
5	The entry must contain an empirical evaluation of the technology/technologies discussed that tests whether the system(s) provide (actionable) insight to users. Note: this criterion ensures the work provides relevant input for our review. Work that describes the design for a Personal Informatics system, but provides no evaluation, or only a technical or usability evaluation, is excluded.

FIGURE 4. Overview of the selection process used in the review.



For an overview of the selection process, see Figure 4. In the first selection round (applying Criteria 1–4), 6,513 of the initial 6,568 entries were excluded because they did not concern original, peer-reviewed work relevant to PI or the self-improvement hypothesis and/or utilized a complex intervention (using additional behavior change techniques besides self-tracking) or were intended for use by clinical/extreme groups. This left 55 entries for the second selection round, where 31 studies were excluded because, although they met all other inclusion criteria, they provided no evaluation of the self-improvement hypothesis (inclusion Criterion 5). In many cases, an evaluation of the system was provided, but it focused on either technical details (e.g., accuracy of a classification algorithm) or user experience (e.g., whether users liked the interface or could find functionality easily).

Twenty-four entries met all of our inclusion criteria. The selected entries were summarized and the results gathered. Several aspects of the work were taken into account, including characteristics of the system(s) under evaluation and the evaluation methods used, and reported outcomes relevant to the Self-improvement hypothesis.

4. RESULTS

This Results section is divided into three parts: First, we describe some of the defining features of the systems evaluated in the articles under review. Second, we provide an overview of the evaluation methods used in these articles. Finally, we review the evidence the articles provide for the self-improvement hypothesis: Was insight obtained, and if so, what kind? Did participants change their behavior based on these insights? For an overview of the articles reviewed, see Figure 5.

FIGURE 5. Overview of articles reviewed.

Reference	Article type: Evaluation of new PI design	Article type: Evaluation of existing PI system	Article type: Experiences of existing PI users	Monitoring: User-controlled	Monitoring: Automatic	Feedback type: Ambient	Feedback type: Mobile app	Feedback type: Desktop app	Evaluation characteristics: N	Evaluation characteristics: Test population	Evaluation characteristics: Duration of use	Findings: Insight	Findings: Behavior change	Evaluation techniques: QS presentations	Evaluation techniques: Interview	Evaluation techniques: Custom survey	Evaluation techniques: Standard survey	Evaluation techniques: Self-tracking data	Evaluation techniques: Behavior observation
Choe, Lee and Schraefel (2015)			×	[?]	[?]	[?]	[?]	[?]	30	PI users	[?]	×		×					
Choe, Lee, et al. (2015)	×			×		×	×		22	Convenience sample	4 wks	×			×	×			
Choe et al. (2014)			×	[?]	[?]	[?]	[?]	[?]	52	PI users	[?]	×		×					
Collins, Cox, Bird and Cornish-Tresstail (2014)		×		×	×			×	23	Students	2 wks	×	×		×		×		
Collins, Cox, Bird and Harrison (2014)		×			×			×	15	Students	2 wks	×	×				×	×	
Consolvo et al. (2008)	×				×	×	×		28	Research panel	12 wks	×	×		×			×	
Cuttone and Larsen (2014)	×				×		×		45	Students	16 wks	×				×			
Cuttone et al. (2013)	×				×		×		45	Students	16 wks	×	×			×			
De Maeyer and Jacobs (2013)		×		×	×		×		10	Convenience sample	5 wks	×			×	×			
Doumit et al. (2015)	×			×				×	76	Students	3 days	×	×		×			×	
Epstein et al. (2015)			×	[?]	[?]	[?]	[?]	[?]	22	PI users	[?]	×			×	×			
Epstein et al. (2014)	×				×		×		14	PI users	4 weeks	×			×	×			
Fritz et al. (2014)			×	[?]	[?]	[?]	[?]	[?]	30	PI users	>12 wks	×	×		×				
Fujinami (2010)	×				×	×			8	Colleagues	12 wks	×			×				
Hori et al. (2013)	×				×		×		9	Business workers	1 day	×	×					×	
Kay et al. (2012)	×				×		×		4	Convenience sample	2 wks	×	×		×		×		
Khovanskaya et al. (2013)	×				×			×	19	Convenience sample	2 wks	×			×				
Li et al. (2011)			×	[?]	[?]	[?]	[?]	[?]	15	PI users	>8 wks	×	×		×				
Li et al. (2010)			×	[?]	[?]	[?]	[?]	[?]	68	PI users	[?]	×			×	×			
Park et al. (2015)	×				×			×	5	Colleagues	2 wks	×			×	×			
Rooksby et al. (2014)		×	×	[?]	[?]	[?]	[?]	[?]	22	(intended) PI users	[?]	×			×				
Snyder et al. (2015)	×				×	×			30	Students, colleagues	1 day	×			×				×
Verdezoto and Grönvall (2015)		×		×				×	10	Older adults	2 days	×			×				×
Wac (2014)		×		×	×			×	1	Students	8 wks	×			×				

Because most of the articles we reviewed employed a qualitative approach to evaluation of insight and behavior change, a formal meta-analysis of the results was not possible. The results of this review with regard to the self-improvement hypothesis are therefore presented in a thematic way, supported by quotes and supplemented with indicators of the strength of evidence (e.g., number of articles mentioning a certain finding) where possible.

4.1. Features and Use of Evaluated Systems

Of the studies we reviewed, seven (29%) investigated the experiences of existing users who adopted PI systems of their own volition. Six (25%) articles provided an evaluation of existing PI systems available on the consumer market by providing these systems to participants. One of these articles (4%) combined both approaches. In 12 (50%) articles, the focus was on evaluating a new design proposed by the authors, either of a complete PI system (eight articles, 33%) or specifically of a data capture technique (one article, 4%) or data visualization (three articles, 13%).

In four studies, an ambient display was discussed (embedded in a mirror, Fujinami, 2010; expressed in light, Snyder et al., 2015; or as a "glanceable" widget or wallpaper on a smartphone, Choe, Lee, Kay, Pratt, & Kientz, 2015; Consolvo et al., 2008). Most systems provided anywhere, anytime access to data and feedback via a mobile app, whereas others used an application intended for desktop use (e.g., Collins, Cox, Bird, & Cornish-Tresstail, 2014; Doumit et al., 2015; Khovanskaya, Baumer, Cosley, Voida, & Gay, 2013). How and when data were accessible in the systems used by existing PI users was generally not reported on.

Of the 17 studies that provided a PI system to previously non-self-tracking participants, 14 (82%) used a system that gathered data automatically, without the need for explicit user interaction (e.g., activity monitoring, tracking social network usage, sleep tracking). Six studies (35%) required participants to provide manual input about certain parameters.

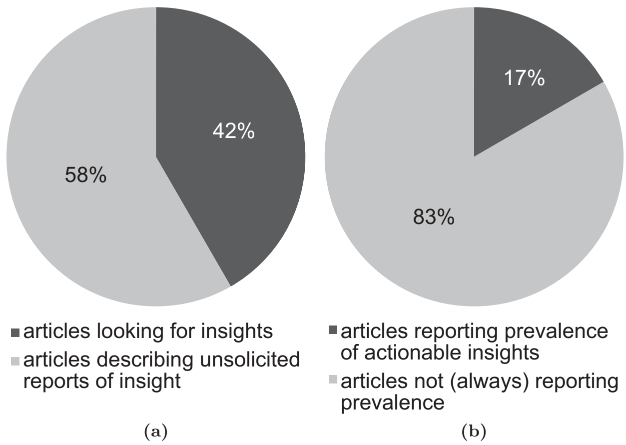
4.2. Evaluation Methods

In this section we discuss several aspects of the evaluations used in the studies we reviewed.

Measuring Insight and Prevalence

The evaluation technique most commonly employed in the studies under review was interviews (18 studies, 75%), followed by custom surveys (8, 33%). A small number of studies employed standardized surveys (3, 13%), recorded presentations about self-tracking experiences (2, 8%), behavioral observations (2, 8%) or used the tracked data to detect changes in behavior (3, 13%).

FIGURE 6. The proportions of studies in our review that (a) actively look for actionable insights in their study, and (b) report on prevalence of actionable insights.



Although all of the articles report on insight-related findings, closer inspection reveals that in only 10 studies (42%) was insight an explicit evaluation topic (also see Figure 6). For instance, Choe et al. (2015) asked participants, "What did you learn while using SleepTight?"; Park, Pedro, and Oliver (2015) included questions like "Did the tool help you discover your habits that you were not aware of" and "Did the findings motivate you to consider changing your browsing habits"; and Cuttone and Larsen (2014) asked participants "if they discovered something new about their own behavior." Collins, Cox, Bird, and Cornish-Tresstail (2014) used a more objective measure, investigating whether participants' estimations of their social network usage improved in response to feedback on the same (also see Collins, Cox, Bird, & Harrison, 2014). Three studies asked about participants' responses to the feedback given by the system but focused on utility rather than specifically on insights gained (e.g., Kay et al., 2012, asked participants, "What data [if any] they found useful"). In the remaining 11 studies (46%), participants were not explicitly probed about data-driven insights, or at least the authors do not report asking these questions. Nevertheless, participants seem to have volunteered insights gained without being prompted to in these studies.

On a related note, only a few of the articles (4, 17%) reported not only on the existence of insights or data-driven behavior change but also on the prevalence of these effects (e.g., how many participants report them). Three additional articles reported prevalence for some insights, but not all. Many of the other articles reported prevalence of other matters (e.g., barriers to usage, Li et al., 2010; motives for usage, Rooksby et al., 2014) but not on insights or data-driven behavior change. Choe, Lee, and Schraefel (2015) reported prevalence of different types of insight, but the taxonomy they used for insights is based on the type of fact participants extract from the data (e.g., noticing outliers, observing a correlation), making it impossible to extract prevalence specifically of actionable insights. We note that a number of articles reported quantitative results on behavior change (e.g., Consolvo et al., 2008; Hori, Tokuda, Miura, Hiyama, & Hirose, 2013), but for these changes it was unclear whether the effect was driven by insights from the data.

Recruitment and Sample

Articles about existing PI users got their participants from the QS forum (Choe et al., 2014, 2015), forums related to specific PI products (Li et al., 2010), or by targeting a larger audience (e.g., via flyers or Amazon's Mechanical Turk) and then selecting on PI use (Epstein et al., 2015; Fritz, Huang, Murphy, & Zimmermann, 2014; Li et al., 2011; Rooksby et al., 2014). Other articles mostly used student populations, although some used colleagues (Fujinami, 2010; Park et al., 2015; Snyder et al., 2015), reports of their own experiences (Wac, 2014), or a convenience sample obtained through a snowball technique or by recruiting from their direct social network (Choe et al., 2015; De Maeyer & Jacobs, 2013; Kay et al., 2012; Khovanskaya et al., 2013). Two articles targeted specific populations (business workers, Hori et al., 2013; active older adults, Verdezoto & Grönvall, 2015), and one used a market research recruitment agency (Consolvo et al., 2008).

An average of 25 people (range = 1–76) participated in the evaluations relevant to the self-improvement hypothesis. Three articles report recruitment in stages, beginning with a larger sample for an initial evaluation (e.g., a general survey) and a further selection for follow-up (e.g., interviews; Epstein et al., 2015; Li et al., 2010, 2011). The selection of these subsets of participants seems to rely partly on self-selection of participants (i.e., only a subset of earlier participants making themselves available for follow-up) but also in part on the experimenters' discretion ("36 [out of 168] respondents were identified as having representative tracking motivations, behaviors, and experiences and were verbose in responses, of which 6 responded"; Epstein et al., 2015, p. 734).

Study Duration, Dropout, and System Usage

For those studies that supplied systems to their participants, the duration of use of the PI system under evaluation was typically less than 2 months (5 weeks on average), with some using only one or two brief sessions (Snyder et al., 2015; Verdezoto & Grönvall, 2015), whereas others deployed the systems under evaluation for several months prior to evaluation (e.g., 4 months, Cuttone & Larsen, 2014). For evaluations based on existing PI users, the duration of previous use was varied and typically reported only on an aggregate level, if at all (e.g. more than 3 months, maximum 54 months; Fritz et al., 2014).

For nine studies, the study design excluded the possibility of dropout, either because they investigated experiences of existing PI users or because the study involved only one or two sessions. Of the remaining 15 studies, four (27%) explicitly discuss dropout, reporting dropout rates of 7% (Epstein, Cordeiro, Bales, Fogarty, & Munson, 2014), 17% (Khovanskaya et al., 2013), 19% (Doumit et al., 2015), and 44% (Park et al., 2015). In nine studies (60%), dropout rates were not explicitly reported, but results imply that all participants completed the study (i.e., 0% dropout).

Seven studies (47%) reported statistics of system usage, providing an alternative measure for (loss of) interest. Most of these studies reported issues of low engagement with the system (Collins, Cox, Bird, & Cornish-Tresstail, 2014; Collins, Cox, Bird, & Harrison, 2014; Cuttone & Larsen, 2014; Cuttone, Lehmann, & Larsen, 2013). By contrast, Kay et al. (2012) reported much higher engagement, with participants spending roughly 10 min per day interacting with the system. Three articles report on users' persistence at manually providing data. Results vary: De Maeyer and Jacobs (2013) reported that although participants were initially eager to provide input about their diet, they typically stopped tracking their food intake after 2 weeks of use. Two studies report on adherence to sleep diaries, but although Choe et al. (2015) reported relatively high adherence (73%–92%), Kay et al. (2012) reported quite low adherence rates (29%).

4.3. Support for the Self-Improvement Hypothesis

Because of the methodological features just noted, the support the articles reviewed provide for the self-improvement hypothesis consists mainly of participants' quotes. Where possible we also note the number or percentage of participants reporting certain effects. Based on the opportunities for data-driven behavior change identified in the introduction (see Figure 1), we report on support for four types of insight (consciousness raising, outcome expectancies, self-efficacy, and contingency management), as well as behavior change based on insight. The results reported in each of the studies under review (i.e., participant quotes and generalized findings such as themes identified by the authors) were scanned for statements relevant to each of these topics; the supporting evidence thus identified is reported for each topic next. Most of the articles reviewed (21, 88%) report on insight gained by users through a PI system. Nine articles (38%) report on changes in participants' behavior as a result of using the PI system. Six articles (25%) report on aspects of both.

Consciousness Raising

The most commonly reported type of insight in the reviewed articles is related to consciousness raising: Participants become more aware of certain aspects of their life (reported in 15 articles, 63%). Fujinami (2010) wrote that several participants report becoming aware of daily walking. Similarly, some of De Maeyer and Jacobs's (2013) participants became aware of how active they were, in either a positive (more than they thought) or negative (less than they thought) sense. In the study by Doumit et al. (2015), participants became aware of portion sizes of food consumed through food tracking. Hori et al. (2013) also noted that most of their participants became aware of the parameter that was being tracked—in this case, smiling. Epstein et al. (2014) provided several examples of awareness-related quotes from their participants, including, "If [I have to travel] over 3 miles, I usually drive. It's interesting to see the breaking point between where I decide it's [too far] to walk" (Epstein et al., 2014, p. 674)

However, many authors also reported that although users became aware of certain things through PI use, these insights were not always actionable. Some insights were simply trivial or "old news": "It's really not very useful and it's kind of annoying. I mean, I walk a lot. What else do I really want to know?" (Li et al., 2010, p. 562) and "It really did just confirm what I already knew" (Collins et al., 2014, p. 376). Park et al. (2015) noted that none of their participants felt that the feedback they received on mobile browsing habits provided any new insights or made them want to change things. In some studies, awareness insights were obtained, but participants expressed no desire to act on these insights, either because the insight did not highlight an actual problem (e.g., participants express surprise at how much they move during the night; Kay et al., 2012) or because the problem that was highlighted was simply not deemed to be in need of solving ("Google already knows everything about me"; Khovanskaya et al., 2013, p. 3407).

In addition, many authors made note of issues concerning consciousness raising. First, some data can be difficult to interpret; Verdezoto and Gro¨Nvall (2015) reported that participants had trouble understanding what their blood pressure measurements meant. Second, some authors express a concern about whether self-reported awareness necessarily reflects long-term and actionable awareness. Cuttone and Larsen (2014), for instance, reported that although participants became aware of their sedentary behavior, "it was not clear if this would lead to actual behavior change or the sustainability of this increased awareness longer-term" (p. 694). Third, awareness may not always be positive; Snyder et al. (2015) reported that the awareness created by giving ambient feedback on stress made some participants feel uncomfortable, as expressed in these participant quotes: "Looking at the light makes me stressed, so I am not going to look at it!" (p. 149); "I already know I'm stressed and I have these … lights that are reflecting how I feel and I already know I'm having a bad day" (p. 149).

Outcome Expectancies

Six studies (25%) reported that users gaining insight about outcome expectancies: How does behavior X affect life outcome Y? Most of the reported insights of this type are found in work about existing PI users, for example, "P35 realized that driving and drinking coffee were triggers for his panic attacks, and eliminated coffee altogether from his diet, which resulted in a decrease in frequency and severity of the attacks" (Choe et al., 2014, p. 1150), "P36 … attributed the changes in his heart rate to a vacation" (Choe et al., 2015, p. 33), "She understood her blood sugar fluctuations better and she was able to set appropriate food and physical activity goals." (Li et al., 2011, p. 409), "P3 felt he was able to use trackers to show that his stomach problem resulted from combining medication with particular foods" (Rooksby et al., 2014, p. 1168). Even for existing PI users, though, this type of insight seems to be relatively rare: Rooksby et al. (2014) noted that of the 22 participants interviewed, only two reported what the authors call "diagnostic tracking," which relates to insights about how one factor affects another.

Only one study using previous PI nonusers as participants reports on how some participants gained specific, actionable insights into how their actions during the day affect their sleep quality (Choe et al., 2015): "If I go to bed directly from doing homework, my sleep is worse" (p. 129); "I sleep better when I have less sugar and eat more earlier [*sic*] in the day" (p. 129). On the other hand, another study reports on how this kind of insight was *not* supported by the PI system used: "This would tell me maybe that my restlessness or my sleep interrupt is coming from noise, but it wouldn't tell me that that noise is happening … when I'm snoring" (Kay et al., 2012, p. 235).

Self-Efficacy

In one of the studies we reviewed a participant reports that they gained insight about previous achievements, boosting their self-efficacy: "And I could see the [feedback] and think, 'I did it last week, you can do it again this time'" (Consolvo et al., 2008, p. 59). This type of insight was not reported in any of the other articles reviewed.

Contingency Management

In six (25%) of the studies reviewed, reports are made of participants using the PI system for contingency management. Li et al. (2010) noted that "short-term reflection is valuable because it makes the user aware of their current status. For example, pedometers show a current aggregate count of steps" (p. 563). Similarly, Rooksby et al. (2014) reported that "directive tracking" (i.e., tracking progress toward goals) was quite common in their sample.

Quotes from participants in various studies support this notion, mainly in the context of activity tracking: "And if I get home at night and I've done 7000 I'll go out and do another 2000. So it's keeping me on that sort of 10,000 track" (Rooksby et al., 2014, p. 1170); "I could see my progress if I was—how much more I needed to do to get to my goal" (Consolvo et al., 2008, p. 59); "I go through between, let's say, 10,000 and 11,000 steps [daily], and I aim for 10 flights of stairs. So I know that if I'm gonna fall short of that … I do work a little harder" (Fritz et al., 2014, p. 491). This type of insight is also noted by Epstein et al. (2015) in a different domain, namely, financial tracking: "I do keep an eye on trends. If I'm trending negative, or trending positive in my accounts" (p. 673).

Behavior Change

Five articles (21%) reported on objective evaluations of behavior change. Of these, two articles reported significant positive results: Hori et al. (2013) noted increased smile rates for their participants after using their smile-feedback system

(p < .001), whereas Consolvo et al. (2008) found that participants who used their activity monitor widget maintained their activity levels over the holidays where the control group did not, $F(1, 312) = 6.51$, $p = .0112$.

Collins, Cox, Bird, and Cornish-Tresstail (2014) reported improved time management behaviors in their participants after using a PI efficiency tool ($t = -4.38$, $p < .01$), but only in one of their three experiments; their other outcome measures show no significant results. In a similar study, Collins, Cox, Bird, and Harrison (2014)found no effects at all. Fujinami (2010) found a significant increase in steps taken when using their system for ambient activity feedback ($p < .01$), but only for one of their six participants.

Besides these objective evaluations, several articles reported on participants' self-reported changes in behavior. As already noted in the previous section, several authors mentioned participants making immediate changes to their behavior in response to insights about progress toward their goals (Consolvo et al., 2008; Epstein et al., 2015; Fritz et al., 2014; Fujinami, 2010; Rooksby et al., 2014).

In two articles, examples are given of how reflection on more long-term data can lead to behavior change: "P77 decided to buy a bike instead of taking the bus because taking the bus was more costly and time-consuming than he had expected" (Choe et al., 2014, p. 1151); "Dianna's temperature data showed that there were several nights where the room was hotter than the suggested maximum. Because she did not have air conditioning, she used a fan on those nights to cool down" (Kay et al., 2012, p. 231). De Maeyer and Jacobs (2013) wrote that participants reported "trying to have more activity in terms of taking stairs more, park a bit further away, drink more water, drink less coffee" (p. 15), although it is unclear whether and how these changes relate to data-driven insights.

As was already mentioned in the section on consciousness raising, insight did not always necessarily lead to changes in behavior. Specifically, Collins, Cox, Bird, and Cornish-Tresstail (2014) mentioned the idea of "acceptance epiphanies": Participants gain awareness of an issue, but the insight leads to acceptance of the status quo rather than an increased motivation to change it. Examples are found in two studies: "It made me accept that the unproductive period just is what it is" (Collins et al., 2014, p. 377) and "Google already knows everything about me" (Khovanskaya et al., 2013, p. 3407).

5. DISCUSSION

5.1. The Self-Improvement Hypothesis

For this review, we looked at articles investigating whether PI systems can provide users with actionable, data-driven self-insight. The results shed light on the current state of evidence for this self-improvement hypothesis but also highlight remaining questions and directions for future work.

Cautious Optimism

At first glance, the results of our review are promising: Almost all of the articles reviewed reported either that participants gained new insights by means of the PI system tested or that participants were able to make changes in their behavior based on their data. Specifically, many participants seem to have reported becoming aware of whatever they are tracking. In addition, several reports are made of participants keeping track of progress toward their goals to help them adjust their behavior along the way. A few participants also reported specific insights about how their behavior influences life outcomes, suggesting specific avenues for improvement, and we found one report of a participant's self-efficacy improving as a result of feedback from a PI system.

However, many authors also reported that although users became aware of certain aspects of their life or behavior through self-tracking, this newfound insight was often not actionable. Many of these insights evoked responses like "Huh, I didn't know that" rather than "Wow, I should do something about that." In addition, some participants indicated that the self-tracking data really did not provide them with any new information. Finally, reports of specific if-this-then-that insights are rare and are balanced by reports of participants who indicated that their PI data were not sufficiently rich to provide this kind of insight.

The type of insight that seems most promising are insights about progress toward a goal. Similar to what Bandura (1991) described in his self-regulation theory, participants tested their self-observations against a norm (goal), and when a discrepancy was detected (e.g., being 2,000 steps short) adjustments to behavior were made (walking up and down the stairs while reading a book). Although this type of insight was not reported as often as those related to consciousness raising, these insights were generally found to be clearly actionable and to lead to immediate changes in behavior.

Coming back to Li et al.'s (2010) stage-based model of PI and the TTM (Prochaska & Velicer, 1997), the results from our review suggest that PI technology in its current state is mainly effective in the action stage and for maintenance of a new behavioral regime, rather than in earlier stages (contemplation/preparation in the TTM or Li et al.'s, 2010, "discovery phase"). This is in contrast to previous work on adoption patterns of the technology, which showed that users earlier in the process of behavior change are more likely to adopt and keep using a PI application (Gouveia, Karapanos, & Hassenzahl, 2015). This suggests that although users look for support from PI tools more often in early stages, the applications currently available support users best in later stages.

Future Directions

The findings just discussed suggest several avenues for future work. Current PI systems seem to be most effective at making users aware of problems (e.g., being physically inactive) and at helping users maintain new behaviors (e.g., walking more steps a day). However, the steps in between are not supported as effectively. Awareness may bring users from the precontemplation stage to the contemplation stage, but that does not help them to come up with specific strategies for tackling the problem or to build the confidence needed to translate those plans into action. To prevent users getting stranded along the way to behavior change, support is needed at all stages.

One important determinant of progress throughout all stages of change is self-efficacy: Users' belief in their own capacity to change is a powerful predictor of success. In our review, we have seen only one instance of data supporting a user's self-efficacy. We believe the potential of feedback about past successes as a way to boost users' self-efficacy is currently underdeveloped and deserves more attention.

Based on the prevailing view on self-tracking, this review has focused on data-driven self-insight as a facilitator of behavior change. But as previously noted, a number of other aspects of self-tracking may also mediate its effects on users. For instance, tracking may interrupt habitual decision making or provide avenues for social support and (positive) social pressure. These effects may compliment the effectiveness of data-driven insight but may also require a different approach to the design of PI systems. To make the most of PI, we need to understand why, how, when, and for whom they work.

Based on these insights, we propose the following directions for future work in the PI field:

1. Supporting users throughout all stages of change: Change is a journey, and although current PI systems provide users with support in some stages of this journey, support for other stages is underdeveloped. Specifically, how do we help users transition from contemplation to preparation and action?
2. A renewed focus on self-efficacy: Self-efficacy is an important predictor of successful behavior change throughout all stages of change. How can we optimize PI systems' potential to boost users' confidence in their ability to change?
3. Structured exploration of complementary approaches: Self-tracking may help users understand and change their behavior via different underlying processes (insight, conscious decision making, observer effects). Which processes work under what circumstances, and for whom? How can we optimize the impact of these potentially complementary processes?

5.2. Evaluating Evaluations of PI Systems: Lessons Learned

The second aim of our review was to examine the way in which PI systems were evaluated in these studies. The articles we reviewed clearly revealed a preference for qualitative evaluation, mostly by means of interviews. This seems an appropriate approach, as insight is a topic not easily measured otherwise, especially given the broad definition of insight in this context. Generally, such a qualitative approach is suited for more exploratory work, which again seems appropriate given that PI is a relatively young field, and many of the questions about these systems have not yet crystallized into testable hypotheses. However, in mapping out current practices in evaluating PI systems, our review also reveals a number of common methodological shortcomings and pitfalls.

Recommendations for the Short Term

Most of the work we reviewed did not report the amount of support for certain themes or findings (i.e., how many participants report a certain experience). In addition, many of the insights reported in the work we reviewed seem to have been volunteered by participants themselves rather than being elicited by targeted questioning. Although participants apparently find insights worth mentioning even without being explicitly prompted, unsolicited reports about *not* gaining any insight seem less likely. Consequently, these accidental findings are likely to paint a skewed picture because they are not balanced by reports of results (i.e., insights) being absent. Overall, this means the findings reported earlier indicate only that certain experiences *exist*, not that they are pervasive, or even common.

An additional issue is that the findings in several of the articles were based on the experiences of self-selected PI users (often members of the QS community), and self-selection plays a role in the selection of participants on a more general level as well (both in initial recruitment and later on through dropout). As previously mentioned, existing models of PI are based on self-selected PI users as well, but their experiences may not generalize to PI usage in other groups. If the goal of PI systems is to target those with a natural penchant for PI, this is not a problem. It seems likely, however, that these interventions will be aimed at a broader population, explicitly including those who are *not* naturally highly motivated to engage in behavior-change interventions. If this is the case, tests with a representative sample from that target population are needed to provide a fair evaluation of the system.

On a broader level, many articles had to be discarded in the selection process prior to the review, because the system proposed or evaluated was too complex, incorporating several behavior change techniques (e.g., information, advice, coaching, sharing) into one intervention. Such interventions may well be effective—perhaps even more so than simpler ones—but an evaluation of such a complex system can never reveal what aspects of the system lead to that success, and why. On a related note, in contrast to the 24 articles reviewed, 31 articles had to be discarded because, although they proposed a PI system with the purpose of facilitating reflection and behavior change, they did not provide any evaluation of whether insight or data-driven change were achieved. Although it would be unrealistic to expect a full efficacy and long-term usage evaluation for every proposed (sub-)system, a small-scale qualitative evaluation can be done at relatively low cost and still provide profound first insights into the way users engage with a new system and the extent to which it achieves the goals for which it was created.

Based on the issues previously noted, we propose several recommendations for evaluations of PI systems for the near future. These methodological recommendations complement the directions for future work outlined in the previous section:

1. Start with a simple system: Investigate the effects of simple interventions to understand how different techniques and components affect users.
2. Start with a simple evaluation: Use small-scale qualitative evaluations to quickly and efficiently evaluate basic ideas; only deploy more extensive (and expensive) evaluations for more mature technologies.
3. Investigate the process, not just the outcome: To better understand how and why certain techniques work and thereby find ways to optimize interventions, evaluate intermediate steps like insight, rather than just outcomes like behavior change or health outcomes.
4. Investigate and report on prevalence of effects: To get a feel for not only the existence but also the prevalence of an effect, explicitly investigate the intended effects (process and/or outcome) of a system and look out for—and report on—(percentages of) users not experiencing these effects as well.
5. Involve the right sample: Only tests with a representative sample from the target population can provide an accurate picture of future use of a system; take care to avoid a strong self-selection bias (e.g. through dropout) as it can skew a sample even if a broader population is targeted in recruitment.

Future Ambitions

The recommendations provided in the previous section are aimed at work in the near future, based on the observation that the PI field in its current (i.e., young) state necessarily focuses on exploratory work. However, as the PI field matures, so should our approach to evaluation of PI systems. Generally, the field would benefit from a more theory-driven approach. Decades of work on the origins and determinants of behavior and behavior change are available, with the potential of providing valuable insight into how behavior can be influenced—also by technology. In the introduction to this review we have given an overview of ways in which self-tracking might theoretically aid behavior change but also highlighted that the underlying mechanisms through which change is accomplished are (a) generally not studied while (b) crucial to optimization of interventions. This issue can be tackled with, on one hand, explicit investigations of underlying processes of change and, on the other hand, structured comparisons of the effects of systems with and without certain components (e.g., a system with social support and without; a system with weekly feedback reports and without).

In line with a more structured, theory-driven approach to PI, the field would also be helped by the development of standardized (outcome) measures to supplement the custom surveys and unstructured interviews that are currently most commonly used. How do we measure (data-driven) insight or awareness? What counts as behavior change? Standardized measures promote trust that reported outcomes are reliable and, of import, allow comparison of different studies.

6. CONCLUSION

The field of PI centers around technologies that allow users to collect and review personally relevant information. The purpose commonly envisioned for these systems is that they provide users with actionable, data-driven self-insight to help them change their behavioral patterns for the better. Here, we have reviewed relevant theory as well as empirical evidence for this self-improvement hypothesis. First results are promising: In the articles we reviewed, participants reported various types of data-driven insights, including those types of insight that might facilitate behavior change—in some cases, data-driven behavior change was even directly reported.

There are some methodological issues, however, that have probably skewed the results toward a more positive picture: In a number of studies, self-selection resulted in samples with a bias toward those with high motivation (i.e., by targeting existing PI users, or through dropout), and in many cases authors did not report on the prevalence of effects and did not look for negative results (i.e., participants *not* finding insights). Thus, although the promise of PI—the self-improvement hypothesis—is by no means disproved and first results are promising, more work is needed to understand how these systems can best be designed to optimally support users in changing their behavior based on self-insight.

Based on our findings, we have proposed a number of future directions for work in the PI field, as well as a set of recommendations for evaluations of PI systems aimed at avoiding common shortcomings and pitfalls observed in the current PI literature. As the PI field matures, we suggest that a more structured, theory-driven approach is needed to help us understand the underlying processes of behavior change through PI. Knowing whether it works is only the start: Only by understanding how, why, and when it works can we identify the best way to help users find what they need to find, and change what they want to change.

NOTES

HCI Editorial Record. First received March 31, 2016. Revisions received November 16, 2016. Accepted by Dr. Jodi Forlizzi. Final manuscript received December 21, 2016. — *Editor*

ORCID

Elisabeth T. Kersten-van Dijk http://orcid.org/0000-0001-8567-0764
Joyce H.D.M. Westerink http://orcid.org/0000-0002-2391-0387
Femke Beute http://orcid.org/0000-0002-4767-0648
Wijnand A. IJsselsteijn http://orcid.org/0000-0001-6856-9269

REFERENCES

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckmann (Eds.), *Action control* (pp. 11–39). Berlin, Germany: Springer.

Bagozzi, R. P. (2000). The poverty of economic explanations of consumption and an action theory alternative. *Managerial and Decision Economics*, *21*, 95–109. doi:10.1002/mde.975

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*, 191–215. doi:10.1037/0033-295X.84.2.191

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, *50*, 248–287. doi:10.1016/0749-5978(91)90022-L

Buchanan, K., Russo, R., & Anderson, B. (2014). Feeding back about eco–feedback: How do consumers use and respond to energy monitors? *Energy Policy*, *73*, 138–146. doi:10.1016/j.enpol.2014.05.008

Burke, L. E., Wang, J., & Sevick, M. A. (2011). Self-monitoring in weight loss: A systematic review of the literature. *Journal of the American Dietetic Association*, *111*, 92–102. doi:10.1016/j.jada.2010.10.008

Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, A. L., Sandercock, P., Spiegelhalter, D., & Tyrer, P. (2000). Framework for design and evaluation of complex interventions to improve health. *British Medical Journal*, *321*(7262), 694–696. doi:10.1136/bmj.321.7262.694

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (Eds.). (1999). *Readings in information visualization: Using vision to think*. San Francisco, CA: Morgan Kaufmann.

Chang, R., Ziemkiewicz, C., Green, T., & Ribarsky, W. (2009). Defining in–sight for visual analytics. *IEEE Computer Graphics and Applications*, *29*(2), 14–17. doi:10.1109/MCG.2009.22

Choe, E. K., Lee, B., Kay, M., Pratt, W., & Kientz, J. A. (2015). Sleeptight: Low-burden, self-monitoring technology for capturing and reflecting on sleep behaviors. *Proceedings of the UbiComp 2015 International Joint Conference on Pervasive and Ubiquitous Computing*. New York, NY: ACM.

Choe, E. K., Lee, B., & Schraefel, M. C. (2015, July). Characterizing visualization insights from quantified selfers' personal data presentations. *IEEE Computer Graphics and Applications*, *35*(4), 28–37. doi:10.1109/MCG.2015.51

Choe, E. K., Lee, N. B., Lee, B., Pratt, W., & Kientz, J. A. (2014). Understanding quantified-selfers' practices in collecting and exploring personal data. *Proceedings of the CHI 2014 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

Collins, E. I. M., Cox, A. L., Bird, J., & Cornish-Tresstail, C. (2014). Barriers to engagement with a personal informatics productivity tool. *Proceedings of the OzCHI 2014 Australian Computer–Human Interaction Conference on Designing Futures*. New York, NY: ACM.

Collins, E. I. M., Cox, A. L., Bird, J., & Harrison, D. (2014). Social network–ing use and rescuetime: The issue of engagement. *Proceedings of the UbiComp 2014 International Joint Conference on Pervasive and Ubiquitous Computing*. New York, NY: ACM.

Consolvo, S., Klasnja, P., McDonald, D. W., Avrahami, D., Froehlich, J., LeGrand, L., … Landay, J. A. (2008). Flowers or a robot army?: En- couraging awareness & activity with personal, mobile displays. *Proceedings of the UbiComp 2008 International Conference on Ubiquitous Computing*. New York, NY: ACM.

Cuttone, A., & Larsen, J. E. (2014). The long tail issue in large scale deployment of personal informatics. *Proceedings of the UbiComp 2014 International Joint Conference on Pervasive and Ubiquitous Computing*. New York, NY: ACM.

Cuttone, A., Lehmann, S., & Larsen, J. E. (2013). A mobile personal informatics system with interactive visualizations of mobility and social interactions. *Proceedings of the PDM 2013 International Workshop on Personal Data Meets Distributed Multimedia*. New York, NY: ACM.

De Maeyer, C., & Jacobs, A. (2013). Sleeping with technology–designing for personal health. In *Shikakeology: Designing triggers for behavior change: Papers from the 2013 AAAI Spring Symposium* (pp. 11–16). Retrieved from http://www.aaai.org/ocs/index.php/SSS/SSS13/paper/view/5726

De Vries, H., Mudde, A., Leijs, I., Charlton, A., Vartiainen, E., Buijs, G., … Kremers, S. (2003). The European Smoking Prevention Framework Approach (EFSA): An example of integral prevention. *Health Education Research*, *18*, 611–626. doi:10.1093/her/cyg031

Doumit, R., Long, J., Kazandjian, C., Gharibeh, N., Karam, L., Song, H., … Zeeni, N. (2015). Effects of recording food intake using cell phone cam- era pictures on energy intake and food choice. *Worldviews on Evidence-Based Nursing*, *13*, 3. doi:10.1111/wvn.12123

Elsden, C., & Kirk, D. S. (2014). A quantified past: Remembering with personal informatics. *Proceedings of the DIS 2014 Companion Publication on Designing Interactive Systems*. New York, NY: ACM.

Epstein, D. A., Cordeiro, F., Bales, E., Fogarty, J., & Munson, S. (2014). Taming data complexity in lifelogs: Exploring visual cuts of personal informatics data. In *Proceedings of the DIS 2014 Conference on Designing Interactive Systems*. New York, NY: ACM.

Epstein, D. A., Ping, A., Fogarty, J., & Munson, S. A. (2015). A lived informatics model of personal informatics. *Proceedings of the UbiComp 2015 International Joint Conference on Pervasive and Ubiquitous Computing*. New York, NY: ACM.

Fishbein, M. (2000). The role of theory in HIV prevention. *AIDS Care*, *12*, 273–278. doi:10.1080/09540120050042918

Fox, S., & Duggan, M. (2013, January 28). *Tracking for health*. Available from http://www.pewinternet.org/2013/01/28/tracking-for-health/

Fritz, T., Huang, E. M., Murphy, G. C., & Zimmermann, T. (2014). Persuasive technology in the real world: A study of long-term use of activity sensing devices for fitness. *Proceedings of the CHI 2014 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

Froehlich, J., Findlater, L., & Landay, J. (2010). The design of eco-feedback technology. *Proceedings of the CHI 2010 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

Fujinami, K. (2010). A case study on information presentation to increase awareness of walking exercise in everyday life. *International Journal of Smart Home*, *4*(4), 11–26. doi:10.1007/978-3-540-68504-328

Gouveia, R., Karapanos, E., & Hassenzahl, M. (2015). How do we engage with activity trackers?: A longitudinal study of Habito. *Proceedings of the UbiComp 2015 International Joint Conference on Pervasive and Ubiquitous Computing*. New York, NY: ACM.

Hagger, M. S., & Chatzisarantis, N. L. (2014). An integrated behavior change model for physical activity. *Exercise and Sport Sciences Reviews*, *42*(2), 62–69. doi:10.1249/JES.0000000000000008

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 81–112. doi:10.3102/003465430298487

Hermsen, S., Frost, J., Renes, R. J., & Kerkhof, P. (2016). Using feedback through digital technology to disrupt and change habitual behavior: A critical review of current literature. *Computers in Human Behavior*, *57*(100), 61–74. doi:10.1016/j.chb.2015.12.023

Hori, Y., Tokuda, Y., Miura, T., Hiyama, A., & Hirose, M. (2013). Communication pedometer: A discussion of gamified communication focused on frequency of smiles. *Proceedings of the 2013 Augmented Human International Conference*. New York, NY: ACM.

Huang, D., Tory, M., Aseniero, B. A., Bartram, L., Bateman, S., Carpendale, S., … Woodbury, R. (2015). Personal visualization and personal visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, *21*, 420–433. doi:10.1109/TVCG.2014.2359887

Kay, M., Choe, E. K., Shepherd, J., Greenstein, B., Watson, N., Consolvo, S., & Kientz, J. A. (2012). Lullaby: A capture & access system for understanding the sleep environment. *Proceedings of the UbiComp 2012 Conference on Ubiquitous Computing*. New York, NY: ACM.

Khovanskaya, V., Baumer, E. P., Cosley, D., Voida, S., & Gay, G. (2013). "every–body knows what you're doing": A critical design approach to personal informatics. *Proceedings of the CHI 2013 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

Klasnja, P., Consolvo, S., & Pratt, W. (2011). How to evaluate technologies for health behavior change in hci research. *Proceedings of the CHI 2011 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

Klayman, J. (1995). Varieties of confirmation bias. In J. Busemeyer, R. Hastie, & D. L. Medin (Eds.), *Decision making from a cognitive perspective: Advances in research and theory* (Vol. 32, pp. 385–418). New York, NY: Academic Press.

Klein, M., Mogles, N., & Van Wissen, A. (2011). Why won't you do what's good for you? using intelligent support for behavior change. In A. Salah & B. Lepri (Eds.), *Human behavior understanding* (Vol. 7065, pp. 104–115). Berlin, Germany: Springer.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*, 254–284. doi:10.1037/0033-2909.119.2.254

Kluger, A. N., & DeNisi, A. (1998). Feedback interventions: Toward the under–standing of a double-edged sword. *Current Directions in Psychological Science*, *7*(3), 67–72. doi:10.1111/1467-8721.ep10772989

Li, I., Dey, A., & Forlizzi, J. (2010). A stage-based model of personal informat- ics systems. *Proceedings of the CHI 2010 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

Li, I., Dey, A. K., & Forlizzi, J. (2011). Understanding my data, myself: Sup–porting self-reflection with ubicomp technologies. *Proceedings of the International Conference on Ubiquitous Computing*. New York, NY: ACM.

Maes, S., & Gebhardt, W. (2000). Self-regulation and health behavior: The health behavior goal model. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 343–368). New York, NY: Academic Press.

McGuire, W. J. (1968). Personality and attitude change: An information- processing theory. In A. G. Greenwald, T. C. Brock, & T. M. Ostrom (Eds.), *Psychological foundations of attitudes* (pp. 171–196). New York, NY: Academic Press.

North, C. (2006, May). Toward measuring visualization insight. *IEEE Computer Graphics and Applications*, *26*(3), 6–9. doi:10.1109/MCG.2006.70

Park, S., Pedro, J. S., & Oliver, N. (2015). Mobitop: Bringing topic-based re– flection to mobile browsing habits. *Proceedings of the 2015 International Joint Conference on Pervasive and*

Ubiquitous Computing and Proceedings of the 2015 International Symposium on Wearable Computers. New York, NY: ACM.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In *Communication and persuasion* (pp. 1–24). New York, NY: Springer. doi:10.1007/978-1-4612-4964-11

Prochaska, J. O., & Velicer, W. F. (1997). The transtheoretical model of health behavior change. *American Journal of Health Promotion*, *12*, 38–48. doi:10.4278/0890-1171-12.1.38

Quantified Self Labs. (2015, December). Quantified self — self knowledge through numbers. Retrieved from http://www.quantifiedself.com

Rogers, R. W. (1975). A protection motivation theory of fear appeals and attitude change. *The Journal of Psychology*, *91*, 93–114. doi:10.1080/00223980.1975.9915803

Rooksby, J., Rost, M., Morrison, A., & Chalmers, M. C. (2014). Personal tracking as lived informatics. *Proceedings of the CHI 2014 Conference on Human Factors in Computing Systems*. New York, NY: ACM.

Rosenstock, I. M. (1974). Historical origins of the health belief model. *Health Education & Behavior*, *2*, 328–335. doi:10.1177/109019817400200403

Ryan, P. (2009). Integrated theory of health behavior change: Background and intervention development. *Clinical Nurse Specialist*, *23*, 161–172. doi:10.1097/NUR.0b013e3181a42373

Schwarzer, R. (2008). Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors. *Applied Psychology: An International Review*, *57*, 1–29. doi:10.1111/j.1464-0597.2007.00325.x

Sheeran, P. (2002). Intention-behavior relations: A conceptual and empirical review. *European Review of Social Psychology*, *12*, 1–36. doi:10.1080/14792772143000003

Snyder, J., Matthews, M., Chien, J., Chang, P. F., Sun, E., Abdullah, S., & Gay, G. (2015). Moodlight: Exploring personal and social implications of ambient display of biosensor data. *Proceedings of the CHI 2015 Conference on Computer Supported Cooperative Work & Social Computing*. New York, NY: ACM.

Verdezoto, N., & Grönvall, E. (2015). On preventive blood pressure self-monitoring at home. *Cognition, Technology & Work*, *18*, 267–285. doi:10.1007/s10111-015-0358-7

Wac, K. (2014). Beat-by-beat getting fit: Leveraging pervasive self-tracking of heart rate in self-management of health. In *Big data becomes personal: Knowledge into meaning: Papers from the AAAI spring symposium* (pp. 102–103). Palo Alto, CA: The AAAI Press.

Whooley, M., Ploderer, B., & Gray, K. (2014). On the integration of self-tracking data amongst quantified self members. *Proceedings of the BCS 2014 Human Computer Interaction Conference on HCI*. Southport,UK: BCS.

Witte, K. (1992). Putting the fear back into fear appeals: The extended paral–lel process model. *Communication Monographs*, *59*, 329–349. doi:10.1080/03637759209376276