AIS Transactions on Human-Computer Interaction

Volume 12 | Issue 1 Article 1

3-31-2020

User Satisfaction with Wearables

Raquel Benbunan-Fich
City University of New York, rbfich@baruch.cuny.edu

Recommended Citation

Benbunan-Fich, R. (2020). User Satisfaction with Wearables. *AIS Transactions on Human-Computer Interaction, 12*(1), 1-27. https://doi.org/10.17705/1thci.00126
DOI: 10.17705/1thci.00126

Transactions on Human-Computer Interaction

Volume 12 Issue 1

3-2020

User Satisfaction with Wearables

Raquel Benbunan-Fich

Zicklin School of Business, Baruch College, City University of New York, rbfich@baruch.cuny.edu

Transactions on Human-Computer Interaction

Research Paper DOI: 10.17705/1thci.00126 ISSN: 1944-3900

User Satisfaction with Wearables

Raquel Benbunan-Fich

Zicklin School of Business, Baruch College, City University of New York

rbfich@baruch.cuny.edu

Abstract:

This study investigates user satisfaction with wearable technologies. It proposes that the integration of expectation confirmation theory with affordance theory sheds light on the sources of user's (dis)confirmation when evaluating technology performance experiences and explains the origins of satisfaction ratings. A qualitative and quantitative analysis of online user reviews of a popular fitness wristband supports the research model. Since the band lacks buttons and numeric displays, users need to interact with the companion software to obtain the information they need. Findings indicate that satisfaction depends on the interaction's quality, the value of digitalizing physical activity, and the extent to which the informational feedback meets users' needs. Moreover, the results suggest that digitalizing physical activity has different effects for different users. While some appreciate data availability in general regardless of their accuracy, those who look for precision do not find such quantification useful. Thus, their evaluative judgments depend on the wearable system's actual performance and the influence that the feedback has on their pursuit of their fitness goals. These results provide theoretical and practical contributions to advance our understanding of wearable technologies.

Keywords: Affordances, Digitalization, Expectation Confirmation Theory, User Satisfaction, Wearables

Fiona Nah was the accepting senior editor for this paper.

1 Introduction

Recent developments in the wearable technology area have produced a new breed of information systems (IS). According to Benbunan-Fich (2019), two characteristics distinguish wearable IS from those that the literature has traditionally examined. First, they operate at the individual level and produce informational outputs and services primarily based on individual activity (Baskerville, 2011). Second, they have fragmented components and typically feature a sensor (worn on the body) that collects input data and transmits it to the Internet for analysis and aggregation and a software application that displays the results (Nelson, Verhagen, & Noordzijc, 2016). In this study, I focus on fitness wristbands, the most popular form of wearable compared to others, such as clip-on devices or clothing made of smart fabric. Fitness wearables monitor several health-related metrics, such as steps, distance walked, calorie consumption, and sleep quality (Statista, 2019). Due to their capabilities, fitness wearables belong in the health and wellness category. Devices in this category typically focus on improving the healthcare quality by empowering patients with information relevant to their wellbeing (Wilson & Djamasbi, 2015).

Wearable technology for health and fitness crosses several demographics. Users include young and technology-savvy consumers, fitness enthusiasts, and older non-tech users interested in achieving a more active lifestyle (Japsen, 2016). According to market forecasts, 450 million wearable gadgets will ship to retailers in 2022, which includes about 50 million wristbands and 28 million sport watches (Gartner, 2018). The convergence of technology affordability with self-quantification trends (Swan, 2013; Schüll, 2016; Etkin, 2016) explains the popularity of fitness wristbands and the market value of the companies that offer them. Google recently announced plans to buy Fitbit, a pioneer company in the wearable space, and, thus, confirmed this market's positive outlook (Copeland & Thomas, 2019). Furthermore, according to a recent study, fitness apps' availability represents a key factor in why users adopt smart watches (Adapa, Nah, Hall, Siau, & Smith, 2018).

Empirical research on fitness wearables has thus far investigated several areas, such as what drives their adoption (Fritz, Huang, Murphy, & Zimmermann, 2014; Nelson et al., 2016; Attig & Franke, 2019), the quality of the user experience (Benbunan-Fich, 2019), and the accuracy of the metrics they produce (Shah, Dunn, Huebner, & Landry, 2017). Other studies have examined their effects on wellbeing (James, Wallance, & Deane, 2019a; James, Deane, & Wallance, 2019b; Sitglbauer, Weber, & Batinic, 2019), why users abandon them (Attig & Franke, 2020), and user satisfaction's determinants (Shin & Biocca, 2017).

Satisfaction constitutes a central construct in behavioral research (Vaezi, Mills, Chin, & Zafar, 2016) and a key predictor of continuance intentions (Bhattarcherjee, 2001) and system success (DeLone & McLean, 2003). While researchers have used many different models and approaches to study user satisfaction, many have used expectation confirmation theory (ECT) in particular (Battacherjee, 2001; Battacherjee & Premkumar, 2004; Brown, Venkatesh, & Goyal, 2014; Lankton & McKnight, 2012). According to this theory, users' satisfaction with a system (or technology) arises as they compare its performance against their initial expectations of it. Thus, satisfaction represents a user's affective or emotional state about the system based on cognitively evaluating the discrepancy between their prior expectations of it and its performance post usage (Bhattacherjee, 2001). Prior studies have referred to performance as a perceptual belief (e.g., Premkumar & Battacharjee, 2008) or an actual experience (Brown, Venkatesh, Kuruzovich, & Massey, 2008; Brown, Venkatesh, & Goyal, 2012; Brown et al., 2014). The empirical literature in IS has devoted comparatively more attention to perceived performance than to actual or experienced performance with novel systems. To address this gap, I use affordance theory to investigate performance experiences in actual contexts of use, which I call performance use experiences henceforth.

Affordance theory provides a conceptual perspective to examine how human actors relate to objects given the possibilities for action that such objects afford (Gibson, 1977). Human-computer interaction (HCI) research has a rich tradition of applying this theory to technology artifacts (Kaptelinin & Nardi, 2012) albeit with a tendency to focus on interface features (Norman, 1999). Affordance is a relational concept that links IT artifacts with their usage and that usage's potential consequences (Huchtby, 2001; Savoli & Barki, 2019). Contemporary research on affordances has shown the value of incorporating not only artifacts' surface-level characteristics but also the deeper purposes they fulfill (Savoli & Barki, 2019; Mettler & Wulf, 2019). With wearable technology, where fragmented components (sensor, software, and data) need to work together to help the user/wearer meet a need, interface-focused studies would be incomplete. Since wearable information systems present a dual interface (one in the device itself and another in the companion software application), their evaluation requires one to assess how these components work together to meet users' needs (Benbunan-Fich, 2019).

However, HCI research on wearables has yet to investigate the link between performance use experiences and satisfaction in realistic contexts (i.e., “in the wild”) in depth. To this end, I use a hierarchical affordance framework to analyze actual performance and characterize the relation between user and the IT artifact as enabling (successful) or constraining (when a failure in the user-object relation occurs). Specifically, in this study, I address the following research question (RQ):

RQ: How do affordances in performance use experiences influence users’ satisfaction with wearables?

I address this research question theoretically by integrating Affordance Theory with ECT. Affordance Theory offers a lens to analyze performance experiences, while ECT provides the evaluative mechanisms to explain how satisfaction is formed as a result of these experiences. Empirically, I tested the hypotheses with a sample of online reviews related to a minimalist wrist-worn fitness tracker that lacks buttons and numeric displays. I employed qualitative and quantitative techniques to analyze the content of the reviews. The results show that satisfaction is contingent upon establishing a successful user-wearable relation with the device and with the software, as well as the user’s perception that the feedback is valuable, given the user’s motivations for using fitness wearables.

This paper proceeds as follows. In Section 2, I provide the theoretical background on affordance theory and ECT. In Section 3, I present several hypotheses based on integrating these two theories. In Section 4, I describe the dataset I obtained (a large sample of user reviews about one of the most popular minimalist fitness wristbands) and the analysis techniques (manual coding and algorithmic text analysis) I used to analyze it. In Section 5, I present the quantitative and qualitative findings and the results from testing the hypotheses. In Section 6, I discuss the findings, the paper’s limitations, and insights for advancing our current understanding of user satisfaction with wearable systems. Finally, in Section 7, I conclude the paper.

2 Theoretical Background

2.1 Affordance Theory

The notion of affordances represents a unifying lens to study the relationship between users and technology. Gibson (1977) originally developed the concept to refer to what the environment offers to an organism, to what the environment provides or furnishes, for good or ill. In Gibson’s conceptualization, the term affordance refers to the environmental properties or features that make interaction between an agent and its surroundings possible. Greeno (1994) proposes the term ability to refer to agent’s corresponding aptitude that contributes to the kind of relation that occurs. Therefore, affordances confer abilities to actors to enable or constrain activity (Gibson, 1977). In the HCI field, Norman (1988) used the term to indicate that affordances provide clues about how things operate. In this interpretation, affordances refer to an artifact’s properties, whose presence suggests functionality and use via (real or perceived) action possibilities (Norman, 1999). A real affordance concerns a device’s or interface’s physical characteristics that allow its operation. This interpretation concurs with Gibson’s original definition, which placed affordances in the environment. In contrast, a perceived affordance concerns the features in an artifact’s appearance that suggest its proper operation.

Alternative interpretations have reconciled the dichotomy between real and perceived affordances by focusing on the notion of affordances in interaction (Vyas, Chisalita, & Dix, 2006). This relational perspective recognizes that affordances emerge in the relation between agents and objects in their environment (Huchtby, 2001). When the analysis focuses on technology, affordance theory explains how technical artifacts afford possibilities for action to goal-oriented actors (Markus & Silver, 2008; Majchrzak & Markus, 2012; Leonardi, 2011). Affordances are functional and relational aspects of an artifact that frame action possibilities (Huchtby, 2001). When agents relate to technology, the artifact’s design restricts the range of possible practices. In the range of possibilities, constraints emerge from the artifact's affordances in situations with a mismatch between the affordances and the task or a misconfiguration in the relation between the agent and the artifact (Huchtby, 2001). In this study, I conceptualize affordance constraints as mismatches in the agent-affordance relationship.

The relational view considers affordance as a socially and culturally constructed relationship between users and artifacts in real-life contexts (Kaptelinin & Nardi, 2012; Baerentsen & Trettvik, 2002; Vyas et al., 2017). It focuses on the interaction between individuals and tools in their environment as actors seek to have an effect on objects (Leonardi, 2011). Researchers that have adopted the relational view have recognized that affordances do not operate independently, and they have proposed different types of arrangements, such as networks or hierarchies (Baerentsen & Trettvik, 2002; Savoli & Barki, 2019). For example, Savoli and Barki (2019) interconnected affordances through a means-ends chain (with a how-what-why triad) to extract affordance hierarchies. Using a similar logic, Baerentsen and Trettvik (2002) applied the relational view of affordances to HCI and artifact design and proposed a hierarchical classification with three levels: how one uses the IT artifact, what purposes the IT artifact fulfills, and why one needs the IT artifact. I adopt this hierarchical framework and use the terms operational, instrumental, and influential to capture the essence of affordances at each level. An *operational* affordance refers to handling an artifact or tool or how the user relates to the artifact at a physical level (e.g., using the wristband). An *instrumental* affordance refers to the artifact's purpose or the goal that the user accomplishes with it (e.g., quantifying steps). An *influential* or need-related affordance explains the motivation that prompted one to use the artifact (e.g., improve health). Table 1 summarizes this mapping.

Table 1. Hierarchical View of Affordances

Level	Question	Affordances
Need related	*Why* does one need the IT artifact?	Influential
Goal oriented	*What* purpose does the IT artifact fulfill?	Instrumental
Interaction	*How* does one use the IT artifact?	Operational

This hierarchical framework highlights the importance of using an affordance lens to investigate not only the interface characteristics and practical use of systems but also the extent to which the systems meet users' motivational needs (Baerentsen & Trettvik, 2002). Past IS research introduced the concept of need-related affordances to understand how one can design systems to meet users' needs (Zhang, 2008; Karahanna, Xu, Xu, & Zhang, 2018). Zhang (2008) defines needs as conditions in individuals that they need for personal growth and their wellbeing and argues that needs act as motivation sources for using an IT artifact.

The hierarchical view of affordances can help one to conceptualize the interplay between affordances from a holistic perspective using means-end reasoning and to analyze technology performance as each user experiences it. Consistent with the relational perspective, these affordances do not refer to the system's features but rather to possible relations between users and systems at these three levels (Baerentsen & Trettvik, 2002). As such, affordances may be associated with multiple possible outcomes and researchers can use them as lenses to study idiosyncratic user experiences (Hutchby, 2001; Evans, Pearce, Vitak, & Treem, 2017; Mettler & Wulf, 2019). However, despite these benefits, this theoretical lens lacks the evaluative mechanisms to explain how users evaluate their experiences and form satisfaction judgments.

2.2 Expectation Confirmation Theory

Expectation confirmation theory (ECT), also called expectation disconfirmation theory, refers to a theory that researchers have broadly adopted to study satisfaction's determinants. At its core, ECT comprises four constructs: expectations, performance, confirmation, and satisfaction (Anderson, 1973; Oliver, 1977). Although researchers have tested alternative models with different combinations of these constructs, Lankton and McKnight (2012) demonstrated the incremental value that one gains from using the complete model that includes performance (along with the other three constructs) vis-à-vis a simple model without performance. Similarly, Brown et al. (2014) examined an experiences-only model and an expectations-only model, which served as baselines. Their conclusions suggest that complete models provide better explanatory power and that performance is an important construct. Some ECT studies have operationalized performance as a perception or belief (e.g., Premkumar & Battacharjee, 2008), while others (Brown et al., 2008, 2012, 2014) have defined it as an actual experience after usage.

The dependent variable in the ECT model—satisfaction—refers to the state that results from comparatively evaluating a priori performance expectations with post-usage performance (Bhattacherjee, 2001). The performance-expectation comparison produces differential effects on satisfaction via dis/confirmation. Positive disconfirmation occurs when product performance exceeds expectations (i.e., over-performance). In this case, satisfaction levels reflect this positive outcome, also called a surprise effect. Negative disconfirmation occurs when a discrepancy arises and the product's performance fails to meet expectations (i.e., under-performance). This discrepancy produces a disappointment effect and reduces satisfaction. Confirmation constitutes the middle point and occurs when performance exactly matches expectations (Oliver, 1977). When users hold realistic expectations, actual product performance determines the magnitude and size of the discrepancy through the value of disconfirmation.

ECT researchers have found that large negative deviations between performance and expectations in any direction affect satisfaction more prominently than small deviations (Anderson, 1973; Brown et al., 2014). In addition, due to asymmetric effects, a large negative value in a variable has a stronger effect than a positive value of the same absolute magnitude (Lankton & McKnight, 2012). Prospect theory (Kahneman & Tversky, 1979) provides the theoretical basis for the non-equivalent effect of negative deviations. This theory proposes that, when making judgments, people hold different perceptions with respect to gains and losses; they perceive the disutility that losses cause as worse than the utility that equivalent gains produce. Asymmetric effects explain why surprise effects due to superior performance might not produce a substantial increase in satisfaction, while disappointment due to subpar performance might substantially decrease satisfaction (Cheung & Lee, 2005; Lankton & McKnight, 2012).

Figure 1 shows the basic ECT model, which I adapted from Venkatesh and Goyal (2010), as integrated with affordance theory through the performance construct. Expectations exist prior to actual use, while the model views performance use experiences with the hierarchical framework of affordances. I present prior expectations with dotted lines because I do not measure them in this study. I conceptualize disconfirmation as a continuum that ranges from negative to positive values and that represents users' assessments when they evaluate a technology artifact in relation to their pre-usage expectations. Taken together, performance use experiences and disconfirmation effects influence satisfaction levels. In Section 3, I develop three hypotheses about the relations among these constructs.

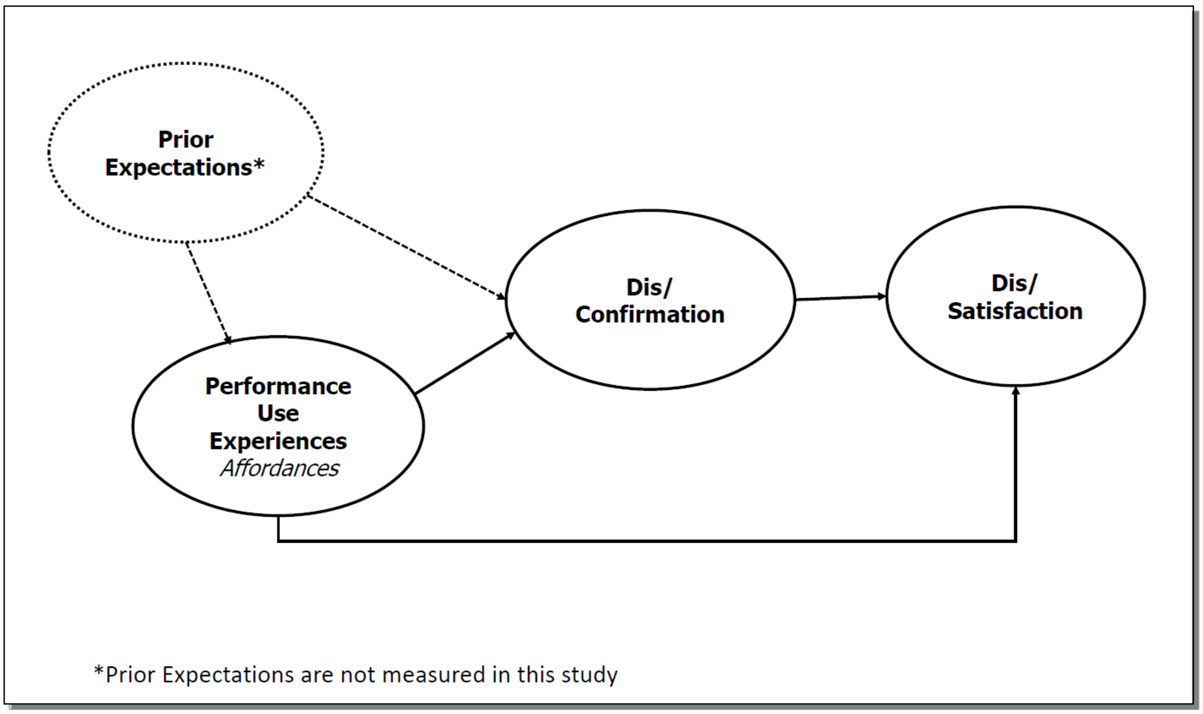

Figure 1. Expectation Confirmation Model (Adapted from Venkatesh & Goyal, 2010)

3 Hypothesis Development

My theoretical integration proposes that the analysis of performance use experiences with an affordance lens elucidates how these experiences produce dis/confirmation and leads to outcomes (dis/satisfaction). Prior research has found support for the notion that assessment and outcomes are anchored to actual experiences and that expectations constitute relatively less important drivers of satisfaction. Specifically, Brown et al. (2008) argue that, due to a recency effect (Tversky & Kahneman, 1974), experiential use is more salient than previous expectations and, therefore, more influential in the evaluation process. Based on this logic, I expect that performance use experiences will directly and positively relate to confirmation effects and that experienced performance use constraints will reduce such confirmation effects to produce (negative) disconfirmation. Thus, I hypothesize:

H1: Performance use experience constraints negatively influence disconfirmation effects.

According to ECT, large disconfirmation effects have a stronger impact on satisfaction than small deviations between performance and expectations (Anderson, 1973; Brown et al., 2014). Furthermore, due to prospect theory (Kahneman & Tversky, 1979) and the principle of asymmetric effects, large negative disconfirmation exerts a stronger influence on outcomes than positive disconfirmation of the same absolute magnitude (Lankton & McKnight, 2012). Researchers have empirically confirmed the principle that negative evaluations have more pronounced effects than positive ones due to loss aversion in other settings (e.g., Cheung & Lee, 2005). Accordingly, I expect that negative disconfirmation (due to under-performance assessments) will have a more pronounced adverse effect on satisfaction than positive disconfirmation (due to over-performance assessments). Thus, I hypothesize:

> **H2:** Negative disconfirmation has a greater impact on satisfaction than positive disconfirmation.

The expectation-disconfirmation paradigm posits that user satisfaction results from their comparing their a priori expectations (i.e., beliefs about desired product attributes) with the actual consumption experience (Oliver, 1977, 1980). From the perspective of product quality and performance, existing research also suggests that, in addition to disconfirmation, satisfaction stems from the judgment that a product or service provides a pleasurable level of consumption-related fulfillment (Oliver, 1989). Hence, satisfaction evaluations result from a process that involves multiple comparisons against standards of performance (Oliver, 1989). Users may derive their satisfaction from performance regardless of their expectations or disconfirmation (Tse & Wilton, 1988). Due to the presence of multiple comparison standards and the likelihood that some may directly affect individuals' satisfaction levels when they experience a product's performance, I expect that disconfirmation will play a mediator role in the relationship between performance and satisfaction albeit not entirely. Thus, I hypothesize:

> **H3:** Disconfirmation partially mediates the negative relationship between performance use experience constraints and satisfaction.

I tested these hypotheses in the fitness wearables context. In terms of the affordance hierarchy that I present in Table 1, fitness wearables help people achieve a healthier and more active routine (need/motivation) by extending their data-collection and information-processing capabilities (self-quantification goal) when they use them (operation). Given the array of potential needs and varied experiences in realistic use contexts, my theoretical integration of affordance theory and ECT posits that enabling and constraining affordances in actual performance leads to disconfirmation levels, which, in turn, drives satisfaction or lack thereof. In Section 3.1, I briefly introduce the research context and summarize relevant recent empirical research.

3.1 Context: Fitness Wearables

Wearables typically comprise different interacting components: a body-worn sensor-equipped device, a Web-based or mobile app gateway software, and data stored online. A wearable device collects data unobtrusively from the wearer and stores it locally until the device synchronizes with the companion software to store it online (Benbunan-Fich, 2019). Although not all wearables follow this approach, this device-to-gateway model represents one of the most popular due to its ability to support small, inconspicuous wearables (Rose, Eldridge, & Chapin, 2015). Consequently, individuals who use fitness wearables interact with three types of IT artifacts: 1) the physical wrist-worn device, 2) the digital components (algorithms and data accessible through a Web-based dashboard or mobile app) associated with the device, and 3) the feedback mechanisms in the form of badges or rewards at the individual level and the online communities created for individuals who use these devices.

Discreet wearables feature a minimalist design and offer a simplified physical interface that typically feature inconspicuous displays and few or no buttons. Therefore, users can access the data that the sensor collects only through software, and a clear separation between physical and digital interfaces exists. The availability of aggregated individual information, the rewards for achieving personal goals, and the optional participation in online communities close the "feedback loop" by reinforcing motivation (Benbunan-Fich, 2019). Users' motivations for adopting fitness technology relate to the need to empower their health (Nelson et al., 2016), improve their wellbeing (James et al., 2019a, 2019b), or achieve a more active and healthier lifestyle (Benbunan-Fich, 2019).

Contemporary research in fitness wearables has addressed the link between motivation to exercise and technology use and between use and outcomes (James et al., 2019a, 2019b; Shin & Biocca, 2017). Other studies have focused more on technology use by examining voluntary adoption (e.g., Fritz et al., 2019), the quality of the user experience (Benbunan-Fich, 2019), user mental models (Mettler & Wulf, 2019), and perceptions and use in everyday life (Rapp & Cena, 2016). Another research stream has examined the positive effects of technology use through empowerment (Nelson et al. 2016) and improvements in perceived physical health (Stiglbauer et al., 2019) and the negative impacts of tracker dependency (Attig & Franke, 2019) and eventual tracker abandonment decisions (Attig & Franke, 2020). These studies have employed various methods ranging from interviews (Fritz et al. 2014) to experiments (Stiglbauer et al., 2019) and surveys (Attig & Franke, 2019, 2020; James et al., 2019a, 2019b). Other studies have employed qualitative techniques such as content analysis (Benbunan-Fich, 2019), diary studies (Rapp & Cena, 2016), or Q-sorting (Mettler & Wulf, 2019). I summarize this literature in Appendix A.

4 Research Methods

In this study, I qualitatively and quantitatively analyzed online user reviews from a popular minimalist fitness wristband. I used online user reviews as a data source in order to combine the advantages of a large sample with the availability of qualitative data. A growing body of research has relied on online reviews as data sources to investigate their quality (Jensen, Averbeck, Zhang, & Wright, 2013; Mudambi & Schuff, 2010), economic impact (Ghose & Ipeirotis, 2011), and influence on purchase behavior (Zhou & Duan, 2016; Zhao, Stylianou, & Zheng, 2018).

4.1 Product Selection

To create a uniform referent object for the reviews, I selected a single minimalist wrist worn wearable (Fitbit Flex). The device comprises a small sensor encased in a wristband that tracks movement by turning acceleration (steps) into digital measurements. The tracker estimates steps, sleep quality, and calories burned with data collected from physical movement patterns. The rubber wristband closes via prongs in their corresponding openings. The sensor has wireless connectivity and sends the data it collects to a compatible smartphone (via Bluetooth 4.0) or to a computer (via a dongle receiver) when the device is nearby. The internal battery in the sensor requires periodic charges with a specially designed charging cable that plugs into a USB port. Since the wristband lacks a battery light indicator, one can see the battery level only through the app. The software sends email notifications when the battery runs low. To charge the tracker, the user must remove the sensor from the rubber wristband and insert it into an opening in the charging cable (see Figure 2).

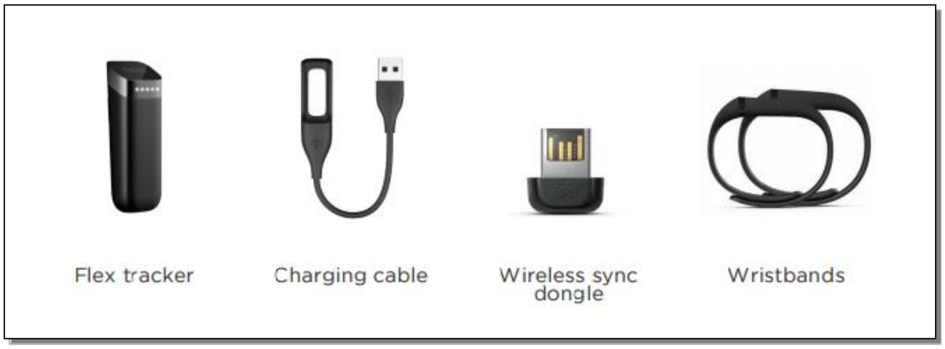

Figure 2. Fitbit Flex Device and Parts (Fitbit, n.d.)

The software enables users to set goals, customize parameters (gender, height, weight, stride), check the battery's status, log additional data (such as activities that the sensor does not track or consumed food), and visualize results. Statistics in charts and tables show the data that the sensor has collected. Users can access the software via a website or via a smartphone app available for certain Android or iOS devices (see Figure 3). Due to technical restrictions in Bluetooth connectivity, the tracker only works with some smartphone models.

Figure 2. Fitbit Flex Software (PixelPedia, 2019)

4.2 Sample and Analysis

I collected all online user reviews of the selected fitness tracker posted on Amazon from April, 2013 (the release date for the product's first version), until August,2016 (the second version's release date). The average length of all the reviews in the entire dataset was 65.77 words (SD ± 100.21), and the average rating of the product was 3.7 stars. Since the dataset comprised thousands of short reviews, I selected a manageable sample suitable for manual coding. To ensure detailed information about performance use experiences, I filtered the reviews and retained only those with at least 350 words or more (i.e., a half page of text).

To analyze the sample, I combined manual content analysis with automatic text-analysis techniques. Content analysis involves making inferences from written text either inductively using an open coding approach or deductively with a set of pre-defined content categories (Krippendorff, 2004). I employed a deductive or directed coding technique using the hierarchical conceptualization of affordances and used a sentiment analysis program to estimate disconfirmation. For the qualitative manual coding, I stripped the records from the numerical satisfaction ratings and blocked portions of the text that explicitly referenced stars or satisfaction levels. Two coders who worked independently examined the set of randomly sorted satisfaction-blind reviews following a drill-down approach. In the first pass, they identified overall enabling affordances in each level, and, in the second pass, coded the constraining affordances in more detail. After each phase, a third coder resolved the coding disagreements and examined the excerpts to identify themes. I also analyzed the reviews' content with a natural language processing (NLP) algorithm to compute the sentiment of each review with a polarity score to estimate the disconfirmation value.

5 Results

In total, 448 reviews that individuals posted between April, 2013, and August, 2016, met the selection criteria based on length (350 words or more). Table 2 shows the descriptive statistics. The average rating in the sample was 3.32. Due to the length filter, the average length was 549.65 words (SD ± 240.74).

Table 2. Descriptive Statistics

Star rating	Num. of reviews	Length mean (± SD)	Max length	Earliest-latest date
1-star (1s)	67	489.51 (± 174.67)	1396	5/13/2013-8/17/2016
2-stars (2s)	62	524.53 (± 184.85)	1160	5/04/2013-8/23/2016
3-stars (3s)	82	594.11 (± 316.30)	2326	4/15/2013-7/28/2016
4-stars (4s)	135	571.84 (± 235.75)	1947	5/02/2013-7/29/2016
5-stars (5s)	102	539.30 (± 238.32)	1986	5/03/2013-7/13/2016
Sample	448	549.65 (± 240.74)	2326	4/15/2013-8/23/2016

The reviews amounted to 296 pages of text. I used QSR NVivo Pro v.11 for coding. I present the results of the content analysis according to the iterations I describe in Section 4.2. In Section 5.1, I identify the main themes by affordance level. In Section 5.2, I more deeply analyze constraining forces at each level. In Sections 5.3 and 5.4, I describe the quantitative measures and the process I used to test the hypotheses in the research model.

Each review is a case that represents an individual user. Although the data did not contain reviewers' identities (since the database had anonymized records), some users voluntarily disclosed their demographic information in their reviews (gender, age, occupation). These voluntary self-disclosures helped frame the wearable experience from the user perspective.

5.1 Enabling Affordances by Level

In the first iteration of the qualitative analysis, the coders considered all the reviews together (regardless of their star ratings) and identified different patterns in how users used the wearable. While the majority of users mainly focused on monitoring steps, some focused on other things. For instance, some users mostly sought to monitor sleep and, hence, found wearing the device comfortably at night important. However, others wanted an activity-tracking tool to compute the number of calories they burned. These users had to manually log their food consumption for the software to compute their caloric intake and add any exercises that the sensor did not automatically detect (such as biking, swimming, etc.) to adjust the number of calories they expended.

In the first iteration, the coders focused on categorizing affordances. To this end, they coded statements in the reviews according to how I defined operational, instrumental, or influential affordance. For example, they coded relations between the user and the wrist-worn device (successful or unsuccessful) as operational. They coded statements regarding the (accurate or inaccurate) measurement capabilities as instrumental. Finally, they coded excerpts related to positive or negative influences of the wearable on the user's state of mind as influential. After this high-level coding, the coders and I examined the excerpts in each category to identify fine-grained topics or themes.

5.1.1 Operational Affordances

The statements at this level referred to how users found wearing the device on the wrist every day. We (the coders and I) categorized the statements into three themes: wearability, maintainability, and understandability.

Statements concerning wearability referred to how comfortable users found the device. The sensor had a comfortable rubber band encasing and a discreet five-light display. However, users sometimes found the closure mechanism challenging the first time, but they tended to get better with it over time.

Statements concerning maintainability referred to the upkeep of the device in terms of charging and cleaning. To charge the sensor, the user needed to take it out of the rubber band and connect it to the special charging cable. Thus, the user could not wear the device while the sensor charged. However, charging offered the opportunity for users to clean the wearable's different parts.

Finally, statements concerning understandability referred to the communication between the user and the wristband. This user-device dialogue relied entirely on sensory affordances (visual and tactile). The device offered two main functions (step tracking and sleep monitoring), and, given the lack of buttons, the user changed the mode by tapping. After a series of rapid taps to change from awake to sleep mode (or vice versa), the device vibrated to indicate that the mode switched. Tapping also activated the light indicators to display progress towards the daily step goal, but, in this case, the user needed to tap only twice. I provide supporting quotes for these themes in Table B1 in Appendix B.

5.1.2 Instrumental Affordances

The affordances at this level concerned the extent to which the system tracked movement (or sleep patterns) and the extent to which the software provided the visibility that the band lacked. We categorized the statements into three themes: data visibility, tracking reliability, and connectivity.

Statements concerning data visibility referred to the step count (or sleep patterns display) and to the battery charge status. Users accessed the software via a Web-based dashboard or a mobile app. While slight variations between the two platforms and between the iOS and Android mobile app versions existed, the interface in them all presented a set of tiles to display metrics and allows users to check progress at a glance.

Statements concerning tracking reliability referred to how the sensor tracked activity and which kind of activities it detected. On the one hand, since the device was a wrist-worn pedometer, it could only approximate steps based on arm movements. It sometimes did not perfectly estimate steps as natural hand motions influence the counts. In active time, arm movements counted as additional steps, whereas, in sleep tracking, the device interpreted arm movement as awake time. The algorithms in the sensor could only track typical steps (or absence of arm movement for sleep monitoring), and it did not correctly track other types of exercises. Users needed to manually log these additional physical activities in the wearable system software to fix the sensor's limitations.

Finally, statements concerning connectivity referred to the communication between the device and the software. For automatically collected data, the software could only show updated metrics if the user paired (or synchronized) the device either with the computer (via a dongle) or with a compatible smartphone. Users needed to connect the device in this way for the wearable to transmit new tracking data and so they could adjust the metrics stored online. Due to the type of connectivity used, the device only worked with specific types of phones. I provide supporting quotes for these themes in Table B2 in Appendix B.

5.1.3 Influential Affordances

The highest level of affordances encompassed statements that speak to the power of the wearable device to help users meet their needs. We categorized the statements into three themes: accountability, quantifiability, and sociability. Statements concerning accountability referred to the device reminding users to do something (e.g., move) given their general need to achieve a healthier or more active lifestyle. Statements concerning quantifiability referred to users' ability to access data about physical activity, which informed their choices and progress toward their fitness goals. Statements about sociability referred to connecting with other users to compete or share progress in the context of an online community. I provide supporting quotes for these themes in Table B3 in Appendix B.

5.2 Constraints by Affordance Level

The affordances identified in the previous phase enable the user-wearable relationship and provide a foundation to identify constraints (i.e., failures in the user-wearable relationship). At the operational level, the three enabling affordances had three corresponding constraints: material tear, battery deterioration, and interface miscommunication. As for material tear and battery deterioration, statements mentioned that the silicone band material had a tendency to tear and deteriorate, which imposed limits on its wearability. Others explained the breakage as resulting from taking the sensor out of the band to charge it. Hygiene issues also impeded proper charging when the sensor did not make proper contact with the special charging cable. As for miscommunication, the user-device communication presented difficulties for users due to the tap conventions (double vs. multiple) and the responses with flashing lights and vibrations. When the device interprets vigorous hand movements as taps, it erroneously changed the tracking mode and created further user-device "misunderstandings".

At the instrumental level, the affordances at this level had several constraints that arose from the tracker's technical characteristics: data inaccuracy, tracking limitations, manual logging, software glitches, and disconnectivity. As for data inaccuracy, given the tracker's nature as a wrist-worn pedometer, it sometimes inaccurately recorded data, which meant that, for some users, step count or sleep patterns lacked credibility. Users also mentioned its tracking limitations because the sensor tracked only walking and lacked an altimeter to track when users walked up stairs. Further, the sensor did not track activities such as running or riding a bike. To completely record their activity, users could manually log their additional exercises in the dashboard and enter their food intake for calorie estimation. While these features compensated for the tracker's limitations, they require extra effort from the users. The device also had shortcomings related to software glitches (in both the Web dashboard and the app) that prevented users from monitoring their progress, consulting charts, or manually logging information. Finally, the disconnectivity constraint arose either due to the device's incompatibility with specific phone models or from a malfunction in the pairing procedures.

At the influential level, the constraints include returns or switch to other models, total malfunction or loss, ineffective or a hassle to use, and quantification obsession. These constraints arose when the device became inconsequential and no longer motivated users due to technical or behavioral reasons. As for returns or switch to other models, users sometimes returned the device or switched it for more advanced trackers. As for malfunction or loss and ineffective or hassle to use constraints, these concern quantifiable failure, such as a dead battery (i.e., it no longer held charge) or the user lost the device (e.g., by accidentally opening its strap). Constraints at this level also include cases where, rather than building accountability, users decided to abandon the device because it did not fit with their lifestyle or they found it a hassle to use. At the other extreme, accountability and tracking become addictive and lead to users developing a quantification obsession wherein the means (tracking) replaced the goal (achieving a more active lifestyle). Table 3 shows the constraints by affordance level and indicates the percentage of reviews with each theme. Appendix C shows an expanded version of the table with selected illustrative quotes.

Table 3. Themes of Constraints by Level

Label	Enablers	Constraints	Incidence of Constraints
Operational	Wearability	Material tear	33%
	Maintainability	Battery deterioration	24%
	Understandability	Interface miscommunication	31%
Instrumental	Tracking reliability	Data inaccuracy	29%
		Tracking limitations	19%
	Data visibility	Manual logging	11%
		Software glitches	16%
	Connectivity	Disconnectivity	18%
Influential	Accountability Quantifiability Sociability	Returns or switch to other models Total malfunction or loss Ineffective or a hassle to use Quantification obsession	13% 9% 8% 2%

Incidence percentages indicate that tear and miscommunication occurred most often followed by data inaccuracy and battery deterioration. Of these four themes, three belong to the operational level, while only one (data inaccuracy) belongs to the instrumental level.

To gain insights into how users assessed technology performance overall, coders tagged each review with a 0-1 variable indicating overall disappointment effect (i.e., 0 for no disappointment, and 1 for overall disappointment). Recall that I stripped numerical satisfaction ratings and blocked text that explicitly referenced stars or satisfaction levels for the qualitative coding. After solving coding disagreements, about a third of the reviews (32%) received a 1 in the disappointment variable. In comparing this code with the rest of the variables, I found that overall disappointment mostly resulted from product malfunction and lack of fit with a user's lifestyle or goals.

5.3 Measures

I operationalized the constructs in the model as follows: I calculated performance use experience constraints (PUEC) by adding the number of constraints that the coders identified in the qualitative coding for each review in the sample. I measured disconfirmation effects via a polarity score calculated with a sentiment analysis algorithm for each review. These sentiment scores offer more granular and refined evaluations than star ratings (Lak & Turetken, 2017). Since polarity scores evaluate performance with a continuous variable that can take positive or negative values in a more nuanced way, I used these scores to operationalize dis/confirmation effects in the reviews. Consistent with other studies on online reviews (Engler, Winter, & Schulz, 2015), I used each review's star rating to operationalize user satisfaction. These ratings offer an overall measure of satisfaction and summatively evaluate the user experience given the issues reported qualitatively in the text (Mudambi & Schuff, 2010).

5.4 Hypothesis Testing

Before reporting the results I obtained from testing the hypotheses, I provide summary statistics of the key measures and their correlations. For the empirical analysis, I considered the entire sample but only quantified performance use experience constraints that the coders identified. Polarity measures, which the sentiment analysis program calculated, ranged from -1.08 to 4.03 with an average of 0.34 (SD ± 0.5). The correlation between PUEC and polarity was negative (-.29) and significant at $p < .0001$. The average satisfaction for the reviews in the sample was 3.32 (SD ± 1.36) stars. The correlation between polarity and star ratings was positive (.43) and significant at $p < .0001$, while the correlation between PUEC and satisfaction was negative (-0.70) and significant at $p < .0001$.

H1 posits that PUEC negatively influences disconfirmation effects. To test H1, I ran a regression model with disconfirmation as the dependent variable and PUEC as the independent variable and found that the beta-coefficient was -0.15 (SE = 0.02, t = -6.35, $p < .0001$). Thus, I found support for H1.

H2 posits that negative disconfirmation has a greater impact on satisfaction than positive disconfirmation. To test H2, I separated the sample into two subsamples according to the sign of the polarity score. I examined these two subsamples via an analysis of variance. Results shows that the satisfaction mean in the negative disconfirmation subsample was significantly lower than in the positive polarity (i.e., positive disconfirmation) group (mean 2.27 (± 1.34) vs. 3.63 (± 1.21); F = 98.68; $p < .0001$). In addition, I ran two regression models—one for each subsample with satisfaction as the dependent variable (DV) and disconfirmation as the independent variable (IV)—and a joint regression for the entire sample. Table 4 shows these results.

Table 4. Regression Analysis of Polarity Subsamples

Dep. var Satisfaction	Polarity samples	N	Intercept	SE	T-value	Beta-coefficient	SE	t-value	R^2	Model F
Model 1	Negative	99	2.67	0.20	13.01***	1.53	0.59	2.59*	6%	6.69*
Model 2	Positive	349	3.26	0.09	32.77***	0.69	0.15	4.55***	5.6%	20.68***
Model 3	All	448	2.99	0.07	41.67***	1.16	0.12	9.93***	18%	98.68***

Chow test results: F-value = 10.08 ***
Significance levels *p < .01; **p < .001; *** p < .0001

The beta-coefficient in the negative polarity regression was 1.53 (SE = 0.59; intercept = 2.67; R^2 = 6%), and the coefficient in the positive polarity regression was 0.69 (SE = 0.15; intercept = 3.26; R^2 = 5%). A Chow test of the difference in the beta-coefficients between the first and second models was significant (F = 10.08; $p < .0001$), which indicates that a unit change in polarity in the negative subsample had a stronger effect on satisfaction that a unit change in polarity in the positive subsample. Thus, I found support for H2.

To provide further insights into the source of these polarity results, I analyzed the constraint profile of each rating category. To this end, I designed a radar chart with five overlapping shapes, one for each rating level with the top four performance use experience constraints (tear, miscommunication, data inaccuracy, and battery deterioration) (see Figure 4). For one-star reviews, the most prominent constraint was lack of band durability (i.e., tear). For two- and three-star reviews, the most prominent constraint was battery deterioration. However, these two categories differ in the inaccuracy vertex, which was most prominent in two-star than in three-star reviews. For above-average reviews (four and five stars), the shapes were similar with a slight difference in miscommunication, which was more prominent for four-star reviews.

The star ratings concur with how the coders qualitatively coded the disappointment effect in the reviews' text. This code appeared in 84 percent of one-star, 76 percent of two-star, 38 percent of three-star, four percent in four-star, and four percent in five-star reviews. Note that the percentage of two-star reviews expressing disappointment was twice as more than the percentage of three-star reviews that did (76% vs. 38%). While a comparative analysis of these two categories shows that they had a similar average number of constraints (3.10 in two-star reviews vs. 2.89 in three-star reviews), they had a noticeable difference regarding data accuracy. In other words, these two rating categories had a similar constraint profile in all aspects except that the two-star reviews more prominently assessed the data that the wearable collected compared to the three-star reviews. In comparing the disappointment code, I found that star ratings contained both cognitive and affective components.

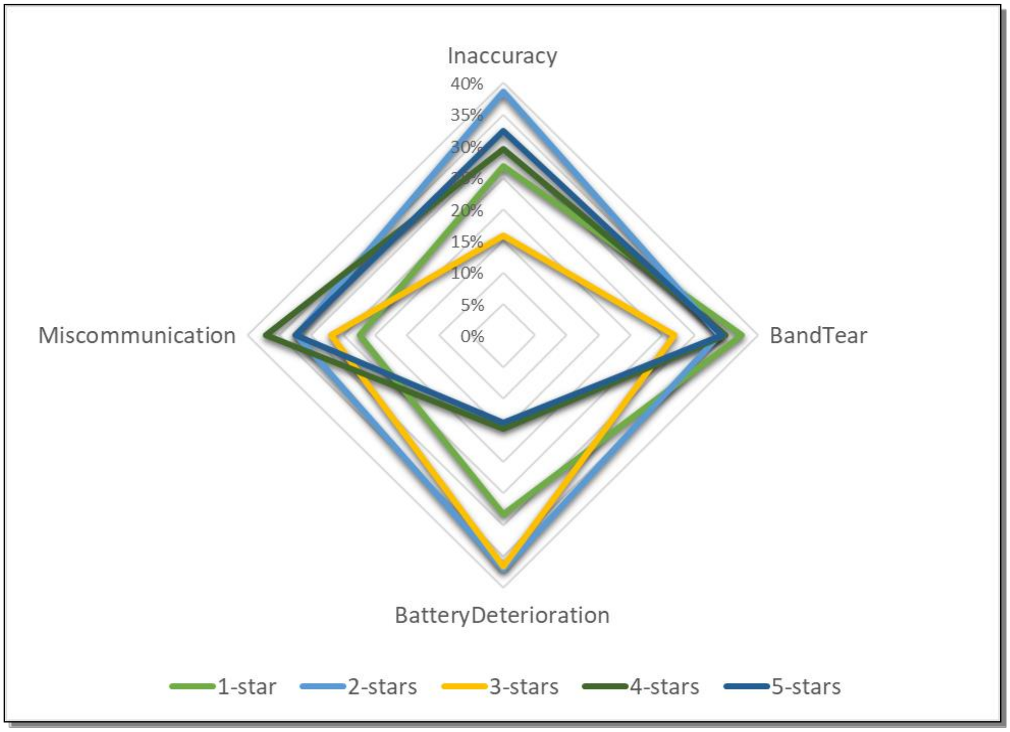

Figure 3. Profile Analysis of Constraints by Satisfaction Level

H3 posits that disconfirmation partially mediates the negative relationship between performance use experience constraints and satisfaction. To test for mediation, I followed the approach that Sobel (1982) suggest and the four steps that Baron and Kenny (1986) outline and ran three regression models. In the first model (first step) with the IV as PUEC and the DV as satisfaction, the beta-coefficient of performance was -1.01 (SE = 0.05, t = -21.02, p < .0001). In the second model (second step) with the IV as PUEC and the DV as disconfirmation, the beta-coefficient of disconfirmation was -0.15 (SE = 0.02, t = -6.35, p < .0001). In the third model (third and fourth steps) with PUEC and disconfirmation as the explanatory variables and the DV as satisfaction, the beta-coefficient of disconfirmation was 0.66 (SE = 0.09, t = 7.32, p < .0001) and the beta-coefficient of performance was -0.91 (SE = 0.05, t = -19.17, p < .0001). I note that the beta-coefficient of disconfirmation in this model differed from the one I report in the third model in Table 4 because this regression has two explanatory variables (PUEC and disconfirmation). I found that the effect that PUEC had on satisfaction controlling for disconfirmation was not zero, which suggests partial mediation (see Table 5).

Table 5. Regression Models for Mediation Tests

	Step	Criterion DV	Predictor(s) IV	Beta coefficient	SE	t-value	R^2	Model F
Model 1	Step 1	Satisfaction	Intercept PUEC	4.89 -1.01	0.087 0.05	55.79*** -21.02***	49.7%	441.72***
Model 2	Step 2	Disconfirmation	Intercept PUEC	0.57 -0.15	0.04 0.02	13.17*** -6.35***	8%	40.35***
Model 3	Step 3 Step 4	Satisfaction	Intercept Disconfirmation PUEC	4.52 0.66 -0.91	0.097 0.09 0.05	4.68*** 7.32*** -19.17***	55%	273.36***

Sobel test statistic -4.79**
Significance levels: **p < .001; ***p < .0001

The mediation test was significant (Sobel test statistic = -4.79; p < .001), and the critical ratio was 0.099 (total effect = -0.10). These results indicate that the indirect effects that PUEC had on satisfaction via the mediator (i.e., dis/confirmation measured with polarity scores) significantly differed from zero, which supports H3. Figure 5 summarizes the results of the mediation tests.

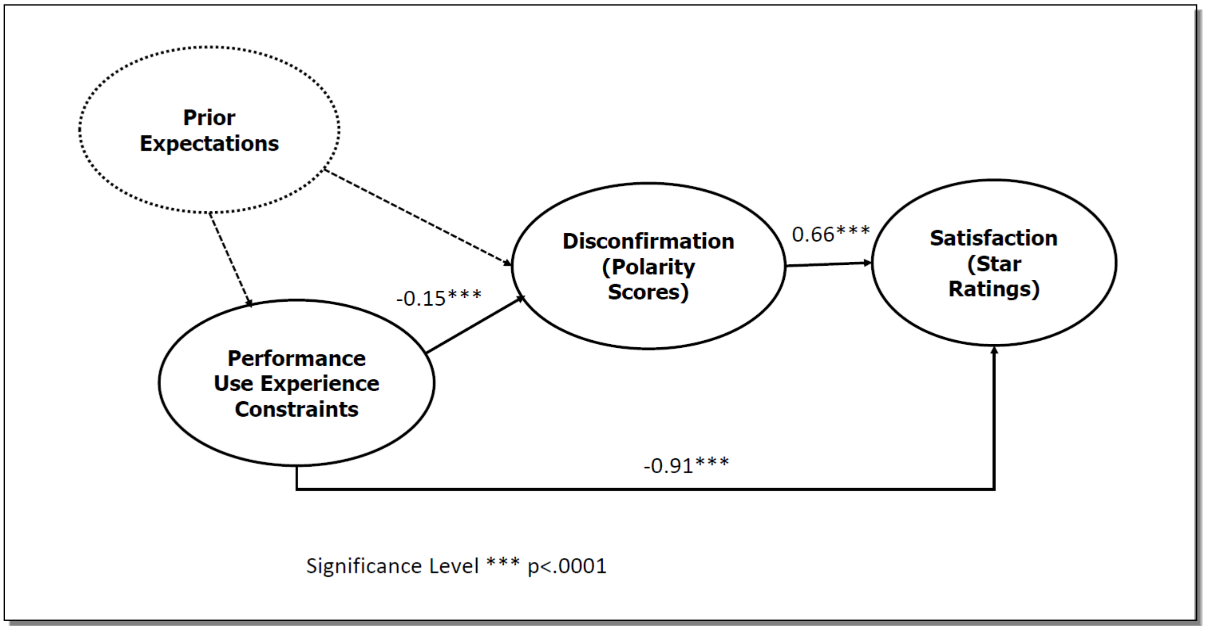

Figure 4. Results of the Mediation Tests

6 Discussion

In this study, I integrate affordance theory with ECT to investigate users' satisfaction with wearables. My theoretical integration proposes a hierarchical view of affordances to analyze performance use experiences in actual usage contexts. The wearable technology I chose for this study (a minimalist wristband fitness tracker) has different potential uses (i.e., pedometer, sleep monitor, and/or calorie counter). The variety of potential uses coupled with specific user needs yields idiosyncratic performance use experiences, where similar satisfaction ratings could represent different situations. While practices might differ by user, identifying enabling and constraining affordances offers a lens to investigate the relation between actual performance use experiences and user satisfaction with real users (e.g., "in the wild"). Furthermore, incorporating (dis)confirmation as a mediator in the relation between negative performance experiences (or constraints) and satisfaction provides additional insights on the determinants of user satisfaction.

The quantitative analyses support the hypotheses I present and validate the approach of quantifying performance with an affordance lens. I operationalized performance constraints based on qualitatively analyzing the reviews' text, whereas I operationalized the disconfirmation continuum with a sentiment analysis program that calculated each review' polarity score. Consistent with prior studies of online reviews, I measured overall satisfaction with star ratings. Using these measures, I showed that performance use experience constraints related to disconfirmation (H1), disconfirmation had an asymmetric effect on satisfaction (H2), and disconfirmation partially mediated the negative relation between performance constraint and satisfaction (H3). I enhanced these results via qualitatively analyzing the reviews.

Findings from the content analysis indicate that the most prevalent constraints (tear and miscommunication) directly relate to the wearable product's physical design (specifically material durability and interface conventions, which belong at the operational level). Lack of data accuracy (due to the sensor's placement on the wrist to quantify feet movement) also appeared among the top constraints. However, this issue belongs in the instrumental level given the product's the minimalist design in that users could not immediately see data it collected and could access it only through the companion software (a Web-based or mobile app). Despite potential data inaccuracy, users evaluate device performance differently. Users who relied on precision in step or sleep quantification (i.e., accurate digitalization) did not find this wearable tracker effective. In contrast, users who accepted a relative count of their movements did appreciate digitalization's benefits.

The hierarchical affordance conceptualization also underscores the importance of analyzing cross-level effects with a means-end chain. The wearable I selected separates the primary interaction (with the device) from access to the digitalization results, which requires a separate interaction with the companion software. When users experienced an interaction breakdown either because the device battery failed or because they could not interact with the software, they had a subpar performance use experience, which resulted in high negative disconfirmation and lower satisfaction levels. In contrast, users who overcame interaction constraints and tolerated data imprecision experienced higher (i.e., more positive) disconfirmation levels and, thus, higher satisfaction (via higher ratings). These results concur with ECT's tenets.

In minimalist devices, the fragmentation between tracking and data visibility separates operational from instrumental affordances. Unclear interaction negatively affects digitalization. More importantly, digitalization's value differs for each user. Some users appreciate data's availability in general regardless of its accuracy. However, those who look for precision do not find such quantification effective. Thus, their evaluative judgments depend on the system's actual performance and the influence that the feedback has on their pursuit of their fitness goals. The hierarchical affordance framework offers a layered lens to evaluate actual performance from a research perspective and from a practical standpoint as well. Accordingly, constraints at the lower levels of the framework prevent users from experiencing affordances at a higher level and, thus, break the means-end chain. Even when the chain remains intact, my findings indicate that users experience different performance situations and contexts that produce different disconfirmation effects (positive or negative. These effects lead to the satisfaction ratings that users ultimately assigned.

6.1 Limitations

As with any empirical study, this work has several limitations that result from the theoretical basis and methodological choices. I used affordance theory and ECT as the main lenses to understand technology performance experiences based on qualitatively and quantitatively analyzing reviews. Work that used an alternative theoretical framework or operationalized the constructs differently could obtain different results. With respect to methodology, my selecting a specific wristband with a minimalist interface as the focal IT artifact restricts the extent to which one can generalize the results to other wearables with different form (worn elsewhere) or function (used for other tracking purposes).

Additionally, because I used secondary data from online reviews, I could not seek supplementary information through questioning or probing users' expectations. Further, by filtering the review dataset to retain longer reviews in the timeframe, I may have affected the incidence level for each issue, and self-reporting experiences through online reviews may produce incomplete reports. However, since users typically document the most salient negative issues they encounter in their online reviews, the focus on performance constraints partially mitigates concerns that the sample underestimated incidence percentages. Moreover, potential inaccuracies in reported percentages do not threaten my cross-level affordance results regarding interaction, digitalization, and feedback.

While these limitations suggest caution when generalizing, they do not invalidate the results. Rather, my findings offer fruitful avenues to extend this research by incorporating other types of wearables and building comparisons between minimalist versus non-minimalist devices or between different types of digitalization goals (e.g., self-quantification vs. monitoring of others, such as in police body-worn cameras).

6.2 Theoretical and Practical Contributions

With this study, I contribute to the literature in several ways. First, I integrate affordance theory with ECT to analyze satisfaction's determinants with wearables. I use a hierarchical classification of affordances at the operational, instrumental, and influential levels based on a means-end chain as a lens to investigate actual performance. Since this conceptual lens does not include the assessment mechanisms to differentiate satisfaction levels, I related performance experiences to satisfaction according to ECT's tenets. The integrated model proposes that user experiences in terms of enabling vis-à-vis constraining forces influence disconfirmation levels and shed light on satisfaction's drivers. Second, I elaborate on the concept of complex IT artifacts/tools and to show how applying AT and ECT helps one theorize about the dynamics of interaction, digitalization, and feedback. Unlike traditional information systems that integrate these three levels, wearable systems that comprise minimalist devices and feedback via supplementary software do not integrate them. As a result, they feature a fragmented and breakdown-prone interaction-digitalization-feedback loop. These failures explain low satisfaction and eventual technology abandonment and contribute to the recent literature on this topic (Attig & Franke, 2020).

From a practical standpoint, individuals who design and research wearables and other self-quantification technologies will find my results valuable. As contemporary research has shown (e.g., James et al. 2019a, 2019b), users have varied needs and goals, and, as a result, they value digitalization differently. For example, digitalization cannot influence users who take data accuracy literally when the tracking system lacks precision. Conversely, when users see digitalization as a means to obtain a relative quantification of their physical activity, the wearable meets their needs. Adapa et al. (2018) has explored the connection between attributes-consequences-value in wearables using a laddering technique. Given my results suggesting that users value the benefits of wearable systems differently, the laddering technique provides a fruitful avenue to extend this research. This study also has another practical implication that mostly applies to minimalist devices, which, by design, impose two constraints on the users: the need to master non-obvious interaction mechanisms and the additional burden to access digitalization. Any design modifications that address these constraints via better explaining interaction conventions or providing more reliable software for users to access data will improve not only the integration of otherwise fragmented components but also overall user satisfaction.

7 Conclusion

The new era of wearable technology calls for novel research approaches. In this study, I integrate affordance theory with expectation confirmation theory to propose a new way to understand satisfaction with contemporary IT artifacts developed for individual use. I revealed new insights from applying these conceptual lenses to a fitness wearable through quantitatively and qualitatively analyzing a sample of online reviews. Findings show how the dynamic interplay of interaction, digitalization of physical movement, and feedback determines whether fitness wearables succeed or fail. In minimalist devices, the fragmentation between tracking and data visibility separates interaction from digitalization and implies that unclear interaction compromises digitalization. However, users have different expectations regarding digitalization. While some appreciate data availability in general regardless of their accuracy, those who look for precision do not find such quantification beneficial. For researchers interested in wearable devices, this study lays the theoretical foundation for future work to examine satisfaction with wearables by integrating affordance theory with ECT. When viewed through these conceptual lenses, user interaction with wearable technology enters the foreground and connects to the underlying needs that motivate users to use technology.

Acknowledgments

This work was partially supported by a PSC-CUNY Grant (#61665-0049). I thank the research assistants who spent many hours coding the reviews and to Melissa Fich for her competent proofreading and editing help. Sincere thanks to the senior editor, Fiona Nah, and the review team for their constructive and helpful feedback to improve this paper.

References

Adapa, A., Nah, F., Hall, R., Siau, K., & Smith, S. (2018). Factors influencing the adoption of smart wearable devices. *International Journal of Human-Computer Interaction*, *34*(5), 399-409.

Anderson, R. E. (1973). Consumer dissatisfaction: The effect of disconfirmed expectancy on perceived product performance. *Journal of Marketing Research*, *10*(1), 38-44.

Attig, C., & Franke, T. (2020). Abandonment of personal quantification: A review and empirical study investigating reasons for wearable activity tracking attrition. *Computers in Human Behavior*, *102*, 223-237.

Attig, C., & Franke, T. (2019). I track, therefore I walk—exploring the motivational costs of wearing activity trackers in actual users. *International Journal of Human-Computer Studies*, *127*, 211-224.

Baerentsen, K. B., & Trettvik, J. (2002). An activity theory approach to affordance. In *Proceedings of NordiCHI*.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182.

Baskerville, R. (2011). Individual information systems as a research arena. *European Journal of Information Systems*, *20*(3), 251-254.

Benbunan-Fich, R. (2019). An affordance lens for wearable information systems. *European Journal of Information Systems*, *28*(3), 256-271.

Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation confirmation model. *MIS Quarterly*, *25*(3), 351-370.

Bhattacherjee, A., & Premkumar, G. (2004). Understanding changes in belief and attitude toward Information Technology usage: A theoretical model and longitudinal test. *MIS Quarterly, 28*(2), 229-254.

Brown, S. A., Venkatesh, V., Kuruzovich, J. N., & Massey, A. P. (2008). Expectation confirmation: An examination of three competing models. *Organizational Behavior and Human Decision Processes 105*(1), 52-66.

Brown, S. A., Venkatesh, V., & Goyal, S. (2014). Expectation confirmation in information systems research: A test of six competing models. *MIS Quarterly*, *38*(3), 729-756.

Brown, S. A., Venkatesh, V., & Goyal. S. (2012). Expectation confirmation in technology use. *Information Systems Research*, *23*(2), 474-487

Cheung, C. M. K., & Lee, M. K. O. (2005). The asymmetric effect of web site attribute performance on web satisfaction: An empirical study. *E-Service Journal*, *3*(3), 65-86.

Copeland, R., & Thomas, P. (2019). Google to buy Fitbit, amping up wearables race. *Wall Street Journal*. Retrieved from https://www.wsj.com/articles/fitbit-to-be-acquired-by-google-llc-11572613473

DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: A ten year update. *Journal of Management Information Systems*, *19*(4), 9-30.

Engler, T. H., Winter, P., & Schulz, M. (2015). Understanding online product ratings: A customer satisfaction model. *Journal of Retailing and Consumer Services*, *27*, 113-120.

Etkin, J. (2016). The hidden cost of personal quantification. *Journal of Consumer Research*, *42*(6), 967-984.

Evans, S. K., Pearce, K. E., Vitak, J., Treem, J. W. (2017). Explicating affordances: A conceptual framework for understanding affordances in communication research. *Journal of Computer-Mediated Communication*, *22*, 35-52.

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. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 487-496).

Fitbit. (n.d.). *Fitbit Flex user manual* (version 1.1). Retrieved from https://staticcs.fitbit.com/content/assets/help/manuals/manual_flex_en_US.pdf

Gartner. (2018). *Gartner says worldwide wearable device sales to grow 26 percent in 2019.* Retrieved from https://www.gartner.com/en/newsroom/press-releases/2018-11-29-gartner-says-worldwide-wearable-device-sales-to-grow-

Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, *23*(10), 1498-1512.

Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 67-82). Hillsdale, NJ: Erlbaum

Greeno, J. G. (1994). Gibson's affordances. *Psychological Review*, *101*(2), 336-342.

Hutchby, I. (2001). Technologies, texts and affordances. *Sociology*, *35*(2), 441-456.

James, T. L., Wallace, L., & Deane, J. K. (2019a). Using organismic integration theory to explore the associations between users' exercise motivations and fitness technology feature set use. *MIS Quarterly*, *43*(1), 287-312.

James T. L, Deane J. K, & Wallace L. (2019b). An application of goal content theory to examine how desired exercise outcomes impact fitness technology feature set selection. *Information Systems Journal*, *29*, 1010-1039.

Japsen, B. (2016). Wearable fitness devices attract more than the young and healthy. *Forbes*. Retrieved from https://www.forbes.com/sites/brucejapsen/2016/07/11/wearable-fitness-devices-attract-more-than-young-healthy/#28bbd45657df

Jensen, M. L., Averbeck, J. M., Zhang, Z., & Wright, K. B. (2013). Credibility of anonymous online product reviews: A language expectancy perspective. *Journal of Management Information Systems*, *30*(1), 293-324.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263-292.

Kaptelinin, V., & Nardi, B. (2012). Affordances in HCI: Toward a mediated action perspective. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.*

Karahanna, E., Xu, S. X., Xu, Y., & Zhang, N. (2018). The needs-affordances-features perspective for the use of social media. *MIS Quarterly*, *42*(3), 737-756.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage.

Lak, P., & Turetken, O. (2017). The impact of sentiment analysis output on decision outcomes: An empirical evaluation. *AIS Transactions on Human-Computer Interaction*, *9*(1), 1-22.

Lankton, N. K., & McKnight, H. D. (2012). Examining two expectation disconfirmation theory models: Assimilation and asymmetry effects. *Journal of the Association for Information Systems, 13*(2), 88-115.

Leonardi, P. M. (2011). When flexible routines meet flexible technologies: Affordance, constraint, and the imbrication of human and material agencies. *MIS Quarterly*, *35*(1), 147-167.

Majchrzak, A., & Markus, M. L. (2012). Technology affordances and constraints in management information systems (MIS). In E. Kessler (Ed.), *Encyclopedia of management theory*. Thousand Oaks, CA: Sage.

Markus, M. L., & Silver, M. S. (2008). A foundation for the study of IT effects. *Journal of the Association for Information Systems*, *9*(10/11), 609-632.

Mettler, T., & Wulf, J. (2019). Physiolytics at the workplace: Affordances and constraints of wearables use from an employee's perspective. *Information Systems Journal*, 29(1), 245-273.

Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quarterly, 34*(1), 185-200.

Nelson, E. C., Verhagen, N. T., & Noordzijc, M. L. (2016). Health empowerment through activity trackers: An empirical smart wristband study. *Computers in Human Behavior, 62*, 364-374.

Norman, D. A. (1988). *The psychology of everyday things*. New York, NY: Basic Books.

Norman, D. A. (1999). Affordances, conventions, and design. *Interactions*, *6*(3), 38-41.

Premkumar, G., & Bhattacherjee, A. (2008). Explaining information technology usage: A test of competing models. *Omega*, *36*(1), 64-75.

Oliver, R. L. (1977). Effect of expectation and disconfirmation on postexposure product evaluations: An alternative interpretation. *Journal of Applied Psychology*, *62*(4), 480-486.

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, *17*(4), 460-469.

Oliver, R. L. (1989). Processing of the satisfaction response in consumption: A suggested framework and research propositions. *Journal of Consumer Satisfaction, Dissatisfaction, and Complaining Behavior, 2*, 1-16.

PixelPedia. (2019). Fitbit—the best fitness app you can trust. Retrieved from https://thepixelpedia.com/fitbit-the-best-fitness-app-you-can-trust/

Rapp, A., & Cena, F. (2016). Personal informatics for everyday life: How users without prior self-tracking experience engage with personal data. *International Journal of Human-Computer Studies*, *94*, 1-17.

Rose, K., Eldridge, S., & Chapin, L. (2015). Internet of things: An overview. *The Internet Society.* Retrieved from https://www.internetsociety.org/wp-content/uploads/2017/08/ISOC-IoT-Overview-20151221-en.pdf

Savoli, A., & Barki, H. (2019). Affordance networks: An approach for linking IT features-in use to their effects. In *Proceedings of the International Conference on Information Systems.*

Schüll, N. D. (2016). Data for life: Wearable technologies and the design of self-care. *Biosocieties, 11*, 317-333.

Shah, Y., Dunn, J., Huebner, E., & Landry, S. (2017). Wearables data integration: Data-driven modeling to adjust for differences in Jawbone and Fitbit estimations of steps calories and resting heart-rate. *Computers in Industry*, *86*, 72-81.

Shin, D.-H., & Biocca, F. (2017). Health experience model of personal informatics: The case of a quantified self. *Computers in Human Behavior*, *69*, 62-74.

Statista. (2019). *Fitness and activity trackers—statistics & facts.* Retrieved from https://www.statista.com/topics/4393/fitness-and-activity-tracker/

Stiglbauer, B., Weber, S., & Batinic, B. (2019). Does your health really benefit from using a self-tracking device? Evidence from a longitudinal randomized control trial. *Computers in Human Behavior, 94*, 131-139.

Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equations models. In S. Leinhart (Ed.), *Sociological methodology* (pp. 290-312). San Francisco, CA: Jossey-Bass.

Swan, K. (2013). The quantified self: Fundamental disruption in big data science and biological discovery. *Big Data*, *1*(2), 85-99.

Tse, D. K., & Wilton, P. C. (1988). Models of consumer satisfaction formation: An extension. *Journal of Marketing Research*, *25*, 204-212.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124-1131.

Vaezi, R., Mills, A., Chin, W., & Zafar, H. (2016). User satisfaction research in information systems: Historical roots and approaches. *Communications of the Association for Information Systems*, *38*, 501-532.

Venkatesh, V., & Goyal, S. (2010). Expectation disconfirmation and technology adoption: Polynomial modeling and response surface analysis. *MIS Quarterly*, *34*(2), 281-303.

Vyas, D., Chisalita, C., & Dix, A. (2017). Organizational affordances: A structuration theory approach to affordances. *Interacting with Computers*, *29*(2), 117-131.

Wilson, E. V., & Djamasbi, S. (2015). Human-computer interaction in health and wellness: Research and publication opportunities. *AIS Transactions on Human-Computer Interaction*, *7*(3), 97-108.

Zhang, P. (2008). Motivational affordances: Reasons for ICT design and use. *Communications of the ACM*, *51*(11), 145-147.

Zhao, K., Stylianou, A. C., & Zheng, Y. (2018). Sources and impacts of social influence from online anonymous user reviews. *Information & Management*, *55*(1) 16-30.

Zhou, W., & Duan, W. (2016). Do professional reviews affect online user choices through user reviews? An empirical study. *Journal of Management Information Systems*, *33*(1), 202-228.

Appendix A: Summary of Selected Empirical Literature on Wearables

Table A1. Literature Review

Study	Objective	Theory	Method	IT artifact
Attig & Franke (2020)	Identify usage barriers and psychological mechanisms resulting in tracker abandonment decisions	Personal quantification and habit formation	Online survey of former users recruited from social media platforms (Facebook, Linkedin, Instagram)	Wrist-worn trackers, worn on clothing, or non-wearable via mobile app)
Attig & Franke (2019)	Investigate if quantified feedback of gamified systems create a dependency that harms motivation when tracker is not available	Self-determination theory and motivation	Scenario and questionnaire-based survey recruited from interest groups in social media platforms	Various wearable trackers
Benbunan-Fich (2019)	Quality of user experience in Wearable Information Systems	Activity theory	Content analysis of online reviews	Wristband wearable
Fritz et al. (2014)	Investigate voluntary adoption and long-term use of wearables in the wild	Personal informatics and persuasive technologies	In depth semi-structured interviews	Fitbit, Fuelband, Jawbone, and Striiv
James et al. (2019a)	How individuals' motivations influence their fitness technology use and the impact of features on wellness outcomes	Organismic interaction theory and self-determination	Survey administered to an Amazon Mechanical Turk sample	Various fitness technologies
James et al. (2019b)	Examine how exercise goals are related to fitness technology use and how use is related to well-being	Goal content theory	Survey administered to an Amazon Mechanical Turk sample	Various fitness technologies
Mettler & Wulf (2019)	Explore the mental models of employees who experience the introduction of physiological analytics in corporate wellness programs	Affordance theory	Mixed methods (Q-methodology)	Various wearables
Nelson et al. (2016)	Whether individuals using activity trackers are more empowered to set and stick to personal health goals	Self-regulation theory	Survey posted on online communities of wristband users	Wristbands
Rapp & Cena (2016)	Investigate how users perceive and use tracking tools in everyday life	Self-monitoring and Personal Informatics	Diary study	Wearable (Jawbone up) vs. Mobile app
Shin & Biocca (2017)	Examine quantified self health experience (motivation, quality, feedback, change in behavior)	Trans-theoretical model of behavior change and expectation confirmation theory	Multiple methods (Survey and Experiment)	Smartwatch (Sony) in experiment
Stiglbauer et al. (2019)	Study of effects on individuals who use quantified self tracking tools	Self-regulation and Self-evaluation	Longitudinal randomized control trial	Various (activity tracker, mobile apps)

Appendix B: Illustrative Quotes of Enabling Affordances by Level

Table B1. Operational Affordances and Illustrative Quotes

Affordance	Quote
Wearability	*Flex is, well, flexible, everywhere except the top for about an inch across. Very comfortable, and I usually forget that I'm wearing it.* *I had to fasten it several times, but once I broke it in a teeny bit, it's been very easy to clasp one-handed and still stays securely fastened around my wrist.*
Maintainability	*This tracker is removable which is a real perk for cleaning and charging.* *The unit is fairly easy to get in and out of the band for charging, and the charger is easy to use.* *From a hygiene aspect it doesn't get any easier. Pop out the Flex dongle and wash the rubber band.*
Understandability	*Yes, the tapping to activate and deactivate sleep takes getting used to but once you get the hang of it, its all cake.* *The sensor has five LEDs that light up as you get 20% closer to reaching your step goal. These lights are activated by simply tapping the device twice making it convenient to get a real time indication of your progress throughout the day.*

Table B2. Instrumental Affordances and Illustrative Quotes

Affordance	Quote
Data visibility/ availability	*The accompanying dashboard through the website is a pretty awesome. All sorts of tiles to click on and go more in detail. More deeper settings you can mess with, like turning off estimated calories which was the first thing I did.* *You set your goal either through the iPhone or the Web interface along with setting details, recording body changes, personal physical details for calorie estimation, and a whole slew of sleep, activity, etc., charts and graphs to appeal to those that such things appeal to. It also plots your sleep cycle and presents it graphically.*
Tracking reliability	*It's a wrist-based pedometer, and that means that if you're doing any activity that over or under involves your hands/arms, it's going to throw off the accuracy of the step counting, regardless of how good fitbit's accelerometer algorithms are.* *They said that the "Dominant" setting makes the device less sensitive and the "Non- dominant" setting is more sensitive. If you are experiencing problems with it recording steps with arm movements I would highly recommend making sure the setting is set to "Dominant". This made a big difference for me :)* *It cannot record other physical actives very well such as biking, cycling, cardio fitness, canoeing/kayaking, stairs, weight training or yoga. It is mainly for walking and hiking. It can detect movement and it can tell when you are being active and for how long but if you are not walking it may or may not record any data.*
Connectivity	*Syncing is very easy on the Flex. You can either open up the app on your phone and it will auto sync, or the device comes with a USB device that will detect the Flex if you're within 20ft of your PC for any length of time. You can also manually sync with the PC device.* *This device works flawlessly out of the box and will work with a number of smart phone models, however check the compatible device list for Android compatible models: …. The device will only work with iPhone 4S and 5 devices. This is not due to a limitation of the Flex device but rather the case that it uses newer Bluetooth technology which is supported by newer smart phones.*

Table B3. Influential Affordances and Illustrative Quotes

Affordance	Quote
Accountability	*My fitbit, a black one to be exact, became my best friend, a reminder to be healthy and aware. It's my workout buddy and accountability partner.* *It is not a training device, it is a lifestyle device, in my opinion, but that's ok. It is a tangible reminder to engage in healthy habits.*
Quantifiability	*What my Fitbit has done for me is provide motivation. I've developed an awareness of how sedentary my job/life has become and am now taking steps (pun intended) to move more. When I look at my graph for the day, I'm now aware that I have large blocks of time where I do nothing!*
Sociability	*For a reasonable price, I am able to track my daily health, which is very important to me in order to improve my overall wellness. My favorite part of this product is that I am able to connect with my friends on Fitbit so that I can see how many steps they took that day. This allows for extra motivation and friendly competition. Overall, by purchasing this product I have been able to set high goals and actually see progress on my mobile device.*

Appendix C: Themes and Quotes of Affordance Constraints by Level

Table C1. Themes and Quotes of Affordance Constraints by Level

Constraint	Quote
Operational level	
Material tear	*The plastic/rubber band holding the device is very weak. Mine developed a major tear in a few weeks, and it's not like you can just pick up any old band.* *You really need to rethink the charging mechanism, as well as the design of the bands- all of mine are tearing now where the rubber part meets the window on the inside part of the band- I guess when you pop the tracker out to charge it or to change the band color, this puts wear and tear on the band.*
Battery deterioration	*When I gently push the FitBit into the cradle and hold it, the lights flash as they are supposed to showing charging. As soon as I let it go, the device goes dead. I have cleaned the contacts but the problem seems to be with the fit of the device into the cradle. I have tried two different cradles with the same result.*
Interface miscommunication	*Double tapping to get the current progress usually works ok, but sometimes I must not be tapping it fast enough or hard enough to register and have to do it again. I have this problem with the 5 quick tap to enter and exit sleep mode. Any activity with minor vibrations can shift the device in and out of sleep mode. I have to remove the device while attending any events that may lead to clapping, and I cannot wear the device while pushing a cart at the grocery store, as the device keeps shifting in and out of sleep mode while adding too many extra steps. But if I don't wear the device, I lose a large number of steps.*
Instrumental level	
Data inaccuracy	*This is in NO WAY accurate. I did find out how to setup your strides for better accuracy, but it still counts more steps than I would take. Again, THIS IS NOT AN ACCURATE READING DEVICE. It can't be if it's on my wrist. reads my driving as steps, make a turn and you get 5 steps!! wave at a neighbor, counts as steps. (I waved my hand while on the couch and got 5 steps...but I was on my couch.)You get the idea. Just know that this device is meant to give you a general idea of how many steps you are generally taking.*
Tracking limitations	*[It] does not know I have elevated the treadmill to 15 degrees. It does not know I am on a bike. It only knows my earth movement and assumes I am walking on flat ground.* *It does not track biking, climbing stairs (which we all know is much more intense than just walking quickly), not swimming, (even though the Flex is submersible to 10m) and not weight training.*
Manual logging	*The activities that the Flex doesn't measure via movement can be manually entered, and if you are honest with yourself as to the degree of exertion, provides further useful info.*
Software glitches	*The app is fine but with glitches. You can't clean up your food selections and it leaves loads of blank entries for some odd reason that you have to continue to scroll through as you try to select foods you have entered in the past. That is a pretty big glitch and I'm an iOS developer so they should have caught that in testing.* *Unfortunately the software has so many glitches it winds up being more frustrating than anything else.*

About the Author

Raquel Benbunan-Fich is an Associate Professor in the Paul H. Chook Department of Information Systems and Statistics at the Zicklin School of Business, Baruch College, City University of New York. Her research interests are in the area of User Behavior and Human-Computer Interaction. Her prior work has been published in many highly recognized peer-reviewed journals such as *ACM Transactions on Human-Computer Interaction, Communications of the ACM, Decision Support Systems, European Journal of Information Systems, Information & Management, International Journal of Electronic Commerce,* and *Journal of Strategic Information Systems*. She received her PhD in Management Information Systems from Rutgers University, Graduate School of Management.