Full citation: Patel, P., Licorish, S. A., Savarimuthu, B.T.R., and MacDonell, S. G. 2016. Studying expectation violations in socio-technical systems: A case study of the mobile app community, in proceedings of the 24th European Conference on Information Systems (ECIS 2016) (Istanbul, Turkey, June 12-15, 2016).

STUDYING EXPECTATION VIOLATIONS IN SOCIO-TECHNICAL SYSTEMS: A CASE STUDY OF THE MOBILE APP COMMUNITY

Patel, Priyanka, University of Otago, Dunedin, New Zealand, priyanka.patel@otago.ac.nz

Licorish, Sherlock A., University of Otago, Dunedin, New Zealand, sherlock.licorish@otago.ac.nz

Savarimuthu, Bastin Tony Roy, University of Otago, Dunedin, New Zealand, tony.savarimuthu@otago.ac.nz

MacDonell, Stephen G., University of Otago, Dunedin, New Zealand, stephen.macdonell@otago.ac.nz

Abstract

With information technology mediating most aspects of contemporary societies, it is important to explore how human-oriented concepts may be leveraged to explore human actions in this new dispensation. One such concept is expectation violations. Expectations govern nearly all aspect of human interactions. However, while this phenomenon has been studied in human-human contexts where violations are expressed through verbal or non-verbal forms, little effort has been dedicated to the study of expectations in human-software contexts in socio-technical systems. We have thus studied expectation violations in one such instance, the mobile app community. Using Expectation Violation and Expectation Confirmation theories, we studied users' reviews of four apps in the health and fitness domain to understand how this app community responds to expectation violations, and if users in a similar domain will express dissatisfaction about similar expectation violations. Our outcomes confirm that the mobile app community responded to expectation violations, just as individuals do in human-human settings. In addition, we observed that users of different health and fitness apps reported similar expectation violations, as is the case for individuals and groups sharing culture-specific expectations. Beyond being of practical relevance for the app community, our outcomes also highlight opportunities for extending the abovementioned theories.

Keywords: text mining, analytics, socio-technical systems, expectancy violation theory (EVT), expectation confirmation theory (ECT), norms

1. Introduction

Interactions between and among individuals in societies, both ancient and modern, have been governed by expectations (Gamble 1982; Schrire 2009). Such expectations arise from societal norms and other rules or conventions of interaction that affect many facets of life. For instance, long established gender-based role allocations have created anticipations (as well as tensions) about the nature of tasks those of a specific gender should undertake (Wood and Eagly 2002). In the same vein, the norm of gift-giving at Christmas and birthday celebrations has been quite commonplace over a period of centuries. A more modern norm (or rule in some contexts) is seen in the *non-smoking practice in public places* of certain states. Furthermore, some norms are enforced, becoming rules, as in the *Equal Employment Opportunities (EEO) for everyone irrespective of one's gender or race* rule. Numerous behavioural expectations of this kind exist in ancient and modern life (Fehr and Fischbacher 2004; Gamble 1982; Schrire 2009; Sugden 1998).

Such expectations are now encountered in the socio-technical domain, such as virtual communities and online platforms. On the basis that individuals are the main drivers of interactions in such systems, *expectations* regarding interaction behaviours of other entities, even in these more recently emergent contexts, naturally arise (Cranefield 2014; Petter 2008). This is fitting given the pervasive adoption and use of information technology (IT) in modern societies, and the associated globalization and decentralization of organizations. This rapid adoption of IT means that interactions between globally distributed individuals and groups (both formal and informal) are increasingly mediated through information and communication technology (ICT) infrastructure.

Socio-technical systems are at the heart of such interactions (Trist 1981). Applications such as wikis, blogs, review portals and discussion boards are used by individuals and groups to start new friendships, explore new ideas, and provide feedback for products and services, instantiating socio-technical systems. In such systems it is common for communication traces from interactions to be archived, providing opportunities for researchers to gain insight into a range of emergent social, cultural and behavioural issues. Of such issues, and given the way societies are governed by expectations, one might seek to learn about commonly expected norms and expectations, penalties for violating these, and more importantly, how to reduce the consequences of violations. Explorations aimed at addressing such issues could provide insights that would lead to functional socio-technical systems (i.e., with satisfied members). Such outcomes could also extend current theoretical constructs and underpinnings, and move the literature on expectation violation forward in these new contexts.

To contribute to such theoretical advancement we study expectation violations in mobile apps, considered among the most engaging of socio-technical systems, where the construction of a technical system (an app) is shaped extensively by the feedback provided by the users of the system. Mobile apps have revolutionized the personal and social computing space wherein multitudes of apps, of which a significant proportion are free, are available to mobile device users. It seems no exaggeration to say that growth in the mobile app market has been phenomenal. The Google Play store, the official and largest Android application repository, offers over 1 million apps to potential users, and the total number of app downloads has exceeded 50 billion (Welch 2013). This community thus provides unique opportunities for research enquiries into users' expectations in socio-technical systems. We have exploited this opportunity in this work. Using Expectation Violation and Expectation Confirmation theories, we studied users' reviews of four apps in the health and fitness domain to understand how the app community responds to expectation violations and if users in a given app domain have similar expectations. We contribute to the advancement of Expectation Violation and Expectation Confirmation theories, and we provide practical insights for the mobile app community.

In the next section we present the study's background and motivation, and also outline our specific research hypotheses (Section 2). We then describe our methodology, introducing our measures in this section (Section 3). Thereafter, we present our results (in Section 4), before discussing our findings and outlining their implications (in Section 5). We then consider the threats to our study in Section 6, before finally providing concluding remarks in Section 7.

2. The Study of Expectations

Expectations are beliefs about future events and/or outcomes (i.e., a belief that something will happen or is likely to happen[1]). Human-human expectations are those expectations held by individuals regarding the behaviour of others (Gamble 1982; Schrire 2009). From a human-human viewpoint, expectation as a concept has been studied from various perspectives, including norms, conventions, policies and pledges (Cranefield 2014). Although possessing subtle differences depending on the field of study, generally, when an expectation is not fulfilled (or, is violated) those affected are disappointed, and their reactions may be expressed – perhaps publicly – in some form of resentment. This could result in an external behavioural change (e.g., a verbal or material sanction) or an internal record of violation (e.g., loss of reputation) (Burgoon 1993; Burgoon and Hale 1988). For example, an advisor expects her students to be on time for their research meetings, and a chairwoman expects her board members to be prepared for their monthly meetings. Violation of either expectation/norm would result in some form of reaction or sanction (e.g., a written warning).

There has been substantial research effort aimed at understanding various forms of human-human expectations in the social sciences (Fehr and Fischbacher 2004; Gamble 1982; Schrire 2009; Sugden 1998). For instance, in the examination of the way social norms are formed, of the forces determining their content, and of the cognitive and emotional requirements that enable individuals to establish and enforce social norms, it was observed that sanctions are decisive for norm enforcement, and that they are largely driven by non-selfish motives (Fehr and Fischbacher 2004). In addition, Fehr and Fischbacher (2004) noted that studying social norms and expectations provides insights into the proximate and ultimate forces behind human cooperation. Oliver and Burke (1999) studied the role and persistence of expectation and expectation-related effects within the expectancy disconfirmation and performance model and found that expectation-initiated performance comparisons (disconfirmation) and performance judgments were important satisfaction influences. In particular, these authors showed that expectation manipulation had an immediate but declining effect on consumption. Others have examined this issue from other perspectives, including for example, from a normative viewpoint (Sugden, 1998) and that of culture adherence (Gamble, 1982).

While some attention has been directed to the study of the fulfilments and violations of expectations that are expressed in socio-technical systems and ICT-mediated interactions (Bailenson et al. 2001; Bonito et al. 1999; Obaid et al. 2012; Poggi et al. 2005), attention has not been given to expectation fulfilment in the mobile app space. Prior work on computer-mediated expectations exists in the area of human-computer interaction (HCI), where the focus has been on designing user interfaces that meet the expectations of intended users of conventional software systems (e.g., windows-based applications) (Bonito et al. 1999; Gürkök et al. 2011). However, a growing range of data sources are now available which contain evidence of expectations and their fulfilment or violation in a range of other socio-technical domains, including the mobile app community (Dam et al. 2015). The hidden expectations (and their violations) in these sources can be extracted and analysed to obtain valuable insights. In fact, recent works have indeed noted that large numbers of requests for software improvements are logged by online communities (Iacob and Harrison 2013; Licorish et al. 2015a; Licorish et al. 2015b; Licorish et al. 2015c; Pagano and Maalej 2013). These are likely to be fuelled in part by stakeholders' expectations. Violations of expectations, in particular, are important to study as they may point towards potential reparations or other consequences that may result if stakeholders' expectations are continually violated. Beyond practical implications, such insights would also advance expectation theories. We consider this issue next. Thereafter, we review the way expectation violations have been studied in socio-technical systems, before developing our hypotheses.

2.1. Expectation Violation

As noted in the previous section, the issue of expectation violation has generated considerable research interest (Fehr and Fischbacher 2004; Gamble 1982; Schrire 2009; Sugden 1998). In the specific context of human communication, the Expectancy Violation Theory (EVT) has been produced (Barry and Crant 2000; Burgoon 1993; Burgoon and Hale 1988). EVT notes that individuals

[1] http://www.merriam-webster.com/dictionary/expectation

form expectations based on societal rules which drives their beliefs about future events, and should the expectation of an individual be violated, they react in compensation (Milewski 2004; Petter 2008). For example, in the field of proxemics, if someone stands too close while having a conversation with another person, the second individual may feel uncomfortable, changing their gesture and posture (Bailenson et al. 2001; Bailenson et al. 2003; Yee et al. 2007). This is because their expectation of the inter-personal distance, arising from a culture-specific norm, has been violated during the conversation. Moving beyond non-verbal cues, EVT has been studied in the context of verbal communication, e.g., teamwork in organisations (Barry and Crant 2000) and higher education (Houser 2006). In these and other disciplines there has been support for EVT in explaining individuals' behaviours and responses to expectations and norms.

Similar to EVT, conventions, norms, contracts, policies, trust, pledges and commitments also drive vendor-purchaser behaviours, and these concepts are thus embedded in consumer satisfaction theories such as the Expectation Disconfirmation Theory (EDT). EDT, also known as Expectation Confirmation Theory (ECT), was initially proposed in the marketing domain (Oliver 1977). ECT notes that post-purchase fulfilment is influenced by previous expectations, perceived performance, and disconfirmation of beliefs. This theory has been successfully employed to study the role of customer expectations on the adoption, usage and continuity of Information Systems (Bhattacherjee 2001; Lankton and McKnight 2006). Such actions are quite related to those that may be prevalent among mobile app users.

In general then, ECT and EVT could provide useful bases for exploring expectations fulfilment or violation in socio-technical domains. In modern societies and globally connected organisations where ICT enables the instant sharing of information, violations of expectations may be reported in a range of social media outlets such as social networks, discussion and review boards and emails (McLaughlin and Vitak 2012; Ramirez and Wang 2008). Thus, extending the predominantly human-human notion of expectations to the domain of software (and its use) would provide insights into how individuals and groups have extended this concept to the requirement of software predictability during use (Lin et al. 2012; MacMillan and Koenig 2004). We examine this issue further in the following section.

2.2. Expectation Violation in Socio-technical Systems

The socio-technical domain comprises humans using applications and/or software entities to achieve specific goals (whether organisation or personal). With the rapid adoption and use of ICT in modern society, ICT-mediated interactions embody expectations (Cranefield 2014; Petter 2008): humans have naturally extended the notion of expectation to the functioning of software (Lin et al. 2012; MacMillan and Koenig 2004). They expect software to behave in a certain way, particularly if they have had previous experience with similar software (e.g., they would likely expect similar menu conventions from software products belonging to the same vendor). Individuals' expectations may also arise due to recommendations from others, or from the reading of product reviews (Bonito et al. 1999; Gürkök et al. 2011). These expectations give rise to beliefs, and anticipated behaviours of 'software entities'. Violations of those beliefs have implications for customer satisfaction.

Expectations may also arise through the ubiquitous usage of mobile apps across domains, including life logging, social networking, banking and health (among others). As a result, there is growing interest in providing insights into the nature of expectation violations faced by individuals increasingly using apps to address personal and work obligations (Iacob and Harrison 2013; Licorish et al. 2015a; Pagano and Maalej 2013). It would thus be insightful to understand the scale of this phenomenon. A prominent source from which such knowledge may be drawn is the feedback portal that is commonly provided to app users. Information conveyed on such portals is publicly available, providing a potentially rich and valuable source of details regarding stakeholders' expectations, and how these are (or are not) accommodated. We framed two hypotheses to address such an enquiry, as outlined next.

2.3. Hypotheses Development

While the consideration of EVT (Barry and Crant 2000) and ECT (Oliver 1977) has to date been limited to human-human expectation violations, we contend here that such theories may be applied to the study of expectation violations in socio-technical systems. Users must *increasingly* interact with software in contemporary societies, giving rise to expectations about what software 'should do'. As with the way humans develop expectations about specific societal conventions and norms, which are driven by an innate system of reasoning, so too are individuals likely to develop expectations about how specific software should behave (Milewski 2004; Petter 2008). In the previous sections it is noted that when an individual's expectation is not fulfilled (or, is violated) they become disappointed, and their reactions may be publicly expressed in some form of resentment (Burgoon 1993; Burgoon and Hale 1988). Given the many channels now available (e.g., social networks, wikis, blogs, review portals and discussion boards) to individuals to express such resentment, responses (i.e., reactions as a form of compensation (Milewski 2004; Petter 2008)) to expectation violations could be numerous. Users may also have expectations that are specific to a given domain, as with the expectation of maintaining inter-personal distance for specific cultures (Bailenson et al. 2001; Bailenson et al. 2003; Yee et al. 2007). Insights into these reactions could be useful for the mobile app community, but may also have implications for theory. In providing such insights and validating the relevance of EVT and ECT to a specific socio-technical instance, that of mobile apps, we thus outline the following hypotheses:

H1. *The mobile app community responds to expectation violations (with resentment) in their reviews, and these expectation violations may be about specific features and issues.*

H2. *Users of a specific app domain will express similar expectation violations given their expectations of how software features should work in that domain.*

3. Methodology

To facilitate our inquiry we extracted reviews from Google Play, using the android-market-api[2]. This API allows the extraction of reviews (4,500 most recent) that are logged by users for apps on Google Play. The reviews comprising author ID, creation time, rating and comment text, were extracted using a script written by the first author of the paper. Thereafter, the reviews were pre-processed by removing punctuation, tags and non-English characters so as to avoid confounding our analysis (Licorish and MacDonell 2013). The cleaned reviews were then stored in a database, where exploratory data analysis involving querying the structure of the reviews and assigning the most suitable datatypes to specific fields was conducted. Through these exercises we noted that reviews sometimes discussed multiple issues in different sentences. We thus partitioned each review into individual sentences, with the sentence forming our unit of analysis. We obtained a total of 16,896 reviews containing 37,633 sentences (refer to Table 1). This data was suitable for conducting our planned analyses. The steps taken in this regard are discussed in the following two subsections.

3.1. Measuring Expectation Violations in Reviews (H1)

In measuring expectation violations (and so testing H1) we employed a multi-phase approach, depicted diagrammatically in Figure 1. We first filtered the sentences to include all those associated with reviews that rated apps poorly (between 1 and 3 out of 5). Given our goal to examine reactions (with resentment) as a form of compensation for expectation violations, we anticipated that users would not rate apps and their features highly if they were dissatisfied with their offerings. In fact, Fu et al. (2013) observed that reviews that were rated =< 3 indeed largely contained expressions of users' dissatisfaction, though, the nature of violations was not considered. In addition, Vasa et al. (2012) observed that reviews rated 1, 2 and 3 tended to have longer comments when compared to those that were rated higher, as they contained descriptive complaints about apps features.

Of that subset of sentences we next extracted those that contained negative words (i.e., either negative emotion words or generic negative words). Negative words, such as *dislike*, *sad*, and *hate*, have been observed to signal dissatisfaction of individuals (Pennebaker et al. 2001). Thus, in order to support our analysis such words were obtained from the Linguistic Inquiry and Word Count (LIWC) library (Pennebaker et al. 2001). The LIWC tool library has been developed by researchers over the past 30 years and has been widely used in similar work; e.g., (Tausczik and Pennebaker 2010). A total of 431 negative words were extracted to support our analyses. For a full list of such words, we refer the

[2] https://code.google.com/p/android-market-api/

reader to LIWC online[3]. Having selected our bag of negative words (431 in total), we tokenized the low-ranked review sentences (i.e., rated =< 3) into words, and then employed a pattern matching approach to extract those that contained expressions of dissatisfaction. If the review sentence contained at least one negative word the sentence was classified as a sentence expressing dissatisfaction.

We anticipated that, beyond the LIWC tool's negative words (e.g., *sad* and *dislike*), other words and phrases may also combine to indicate dissatisfaction (e.g., *does not*, *cannot* and *must not*). We indeed observed this during our preliminary exploratory analysis. Thus, another bag of words containing 57 such phrases, including negative modal verbs (e.g., *must not*, *should not*, *may not*) and common slangs such as "*ain't"*, and "*shan't"* was constructed. This way, even if a review sentence did not contain negative words as captured by the LIWC dictionary, it would still be labelled as a violation sentence if it contained one or more of the 57 additional phrases in the second bag of words. In perusing the literature to further understand word use and its relationship to the study of norms and expectations, we noted that modal verbs (e.g., *must* and *should*) have been used previously to identify obligations in emails and business contracts (Gao and Singh 2014; Kalia et al. 2013). Negations also tend to correspond to expectation violations arising from obligations and prohibitions (Meyer and Wieringa 1994). Obligations describe what someone is expected to do, while prohibitions describe what individuals are not expected to do (Savarimuthu and Dam 2014). We anticipated that both forms of violation may exist in app reviews, and thus, our inclusion of the negations of such verbs (e.g., *must not*) would indeed capture additional violations beyond those extracted through the use of the LIWC negative words. Of the 37,633 total pre-processed sentences, 4,523 were classified as containing expectation violations (i.e., 12% violations).

To initially validate that these sentences indeed expressed users' dissatisfaction as a result of expectation violations, two of the authors coded 400 randomly selected sentences from the 4,523. The 400 sentences were labelled with Y or N depending on if they were assessed as representing dissatisfaction as a result of a violation or not. We found that only 17 sentences were wrongly selected, indicating a 96% agreement. We are thus confident that the approach used above indeed reliably selected review sentences that contained expressions of dissatisfaction as a result of expectation violations.

Prior research has established that noun terms in unstructured text reflect the main concepts in the subject of a clause (Dengel 2014; Yu et al. 2011). From parts-of-speech (POS) and linguistic perspectives, nouns are indeed reflective of specific objects or things, and often form the subjects and objects of clauses or verb phrases, respectively. These and other understandings have been embedded as rules in natural language processing (NLP) tools, including the Stanford parser which performs POS tagging (Toutanova et al. 2003). We next created a software program that incorporated the Stanford API[4] to enable us to extract noun phrases from the 4,523 review sentences to identify app features that resulted in expectation violations. We then counted the frequency of each noun unigram (e.g., if "notification" appeared at least once in 20 review sentences our program would output notification = 20). This allowed us to rank the app features that resulted in most dissatisfaction (i.e., those that most often violated users' expectations).

Finally, we conducted analyses of association between the most problematic features and other unigrams in the review sentences in exploring the scale of expectation violations for specific features and issues. We first examined features that were mentioned regularly with others (e.g., GPS and Battery) (Licorish et al. 2015a). We then examined the specific dissatisfaction or issue that was complained about (e.g., GPS and lost). These analyses utilized two metrics for measuring association (or co-occurrence): (1) frequency of features and/or issues in sentences identified as expectation violation sentences, and (2) conditional probabilities of the occurrences for a pair of entities (feature-feature and feature-issue). Finally, the results obtained from these analyses were carefully examined to draw appropriate inferences. We illustrate the above processes, with examples, in Figure 1.

---

[3] http://www.liwc.net/

[4] http://nlp.stanford.edu/nlp/javadoc/javanlp/

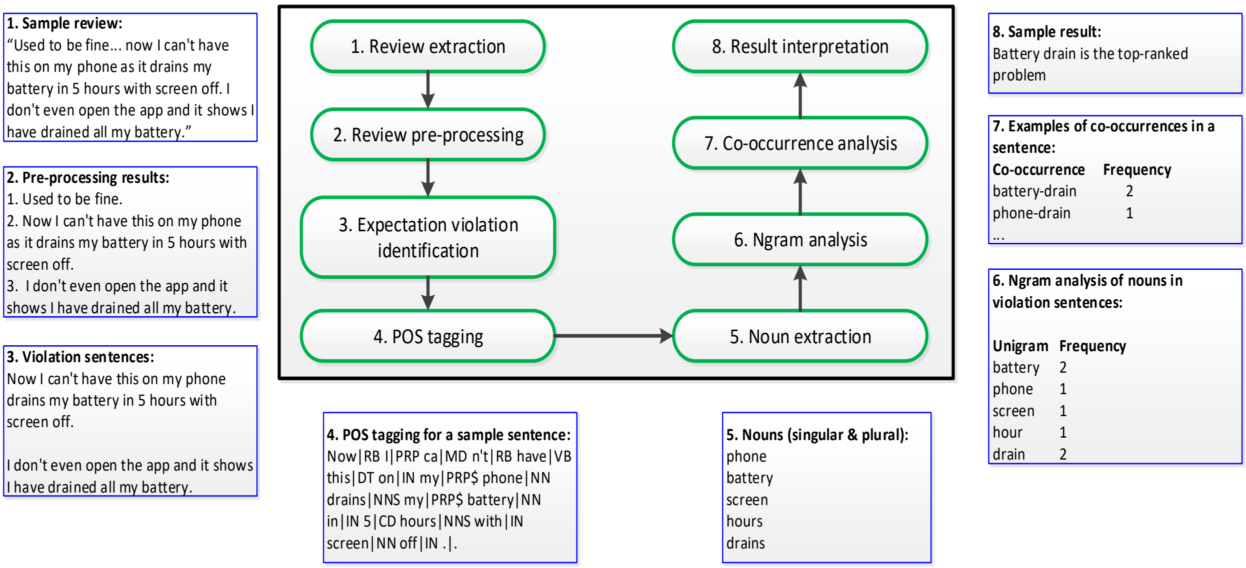

Figure 1. Process for extracting and analysing expectation violations

App	Reviews	Sentences	Sentences/Review	Expectation Violations (EV)	EV/Sentence
My Tracks	4,471	8,491	1.9	1,051	0.12
Endomondo	4,442	8,623	1.9	1,271	0.15
Nooms Cardio Trainer	4,477	9,301	2.1	1,538	0.17
Zombies Run!	3,506	5,609	1.6	663	0.12
∑	**16,896**	**37,633**	**2.2**	**4,523**	**0.12**

Table 1. Summary statistics for health and fitness apps considered

3.2. Expectation Violations in a Specific App Domain (H2)

In order to test H2 we extracted features that signaled expectation violations for multiple apps in the same domain. The above process in Section 3.1 was executed on the review sentences of four apps (My Tracks, Endomondo, Nooms Cardio Trainer and Zombies Run!) belonging to the health and fitness domain. The four apps were chosen based on their popularity in the domain over the past year. The popularity score was based primarily on usage statistics and rankings, which were both available through the use of the Google API (described in Section 3.1). Table 1 provides summary statistics for the apps considered, which shows that similar numbers of reviews were logged for three of the apps (My Tracks, Endomondo, Nooms Cardio Trainer), with the last app (Zombies Run!) attracting fewer reviews. The average sentence lengths of the reviews were also very similar for the first three apps in the table, with users logging slightly shorter reviews for 'Zombies Run!'. In addition, Table 1 shows that between 12% and 17% of the review sentences for these apps contained expressions of dissatisfaction. Given the size of the samples in Table 1, our dataset was more than adequate to support our enquiry. We present the outcomes of our analyses in order to test H1 and H2 in the following section.

4. Results

We present our results to test the two hypotheses in this section. We first present three sets of outcomes towards testing H1. Thereafter, we report results of how expectation violations were similar or differed across the four apps in testing H2.

4.1. Expectation Violations in Reviews (H1)

Top Features with Expectation Violations: Having ranked the app features that resulted in the most expectation violations, we noted that these were too numerous to present in a readable way given space restrictions. We thus present the top-10 features that resulted in expectation violations for the four apps selected from Google Play (My Tracks, Endomondo, Nooms Cardio Trainer and Zombies

Run!). Our analysis shows that the top-10 features that resulted in expectation violations were (in descending order): *GPS*, *Time*, *Map*, *Signal*, *Data*, *Distance*, *Battery*, *Support*, *Location* and *Version* (refer to Figure 2).

Taking a closer look at these results it is noted that *GPS* accounts for 34% of the expectation violations in the top-10 list. Given that the app domain selected concerns the tracking of exercise (e.g., walks and runs) it is understandable that *GPS* functionality would be of interest to users, as this functionality is generally used to identify an individual's location over time. The scale of the problems reported, however, leaves much to be desired. Tracking of exercise also involves recording how long individuals trained (i.e., *Time*). Our results in Figure 2 show that the working of this feature also violated users' expectations, accounting for the second highest number of complaints (14% of issues in the top-10 list). *Map* and *Signal* problems rank third and fourth respectively in Figure 2, and the other top-10 features that violated users' expectations ranked similarly (with counts close to 100). A two-way ANOVA test was conducted to formally examine the distribution of problems reported in Figure 2, which confirmed that there were indeed significantly more violations related to *GPS*, *Time* and *Map* features ($P<0.01$), with Tukey post-hoc test results confirming that *GPS* was particularly troublesome ($P<0.05$ for all comparisons).

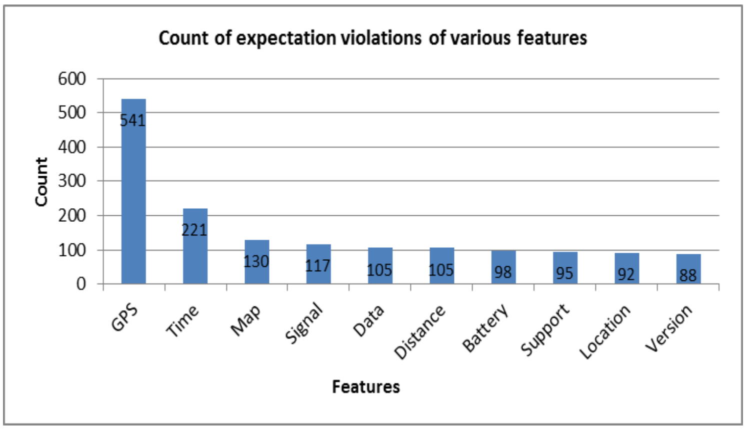

Figure 2. Counts of top 10 features violating user expectations

Associated Features with Expectations Violations: It was previously established that there are often interdependencies between problematic features (Licorish et al. 2015a). Thus, in order to investigate if there were interdependencies between features that resulted in expectation violations, we compute the conditional probabilities of co-occurring features using the formula: $P(A|B) = P(A \cap B)/P(B)$. To explain the formula, let us consider an example. A total of 4,523 review sentences were logged because of expectation violations. Of these sentences, 117 contained the feature *Signal*. Out of these 117 *Signal* entries, *GPS* and *Signal* co-occurred 103 times. In other words, the probability of *GPS* co-occurring with *Signal* is 0.88[5]. This probability could be classified as high given that when users complain about *Signal*, they also complained about *GPS* (i.e., 88% of the times). Beyond testing our hypotheses in view of extending theory, this insight could be potentially useful for the health and fitness app community in terms of identifying these interconnections. We computed pairwise conditional probabilities for all combinations of violation features (a total of 10*9 = 90 combinations). Figure 3 shows the feature-feature network where paired features had a conditional probability value that was higher than the mean probability value for a particular feature co-occurring with all the other features. For example, for *GPS* and the other nine features (i.e., for the nine pairs), the conditional probabilities were computed and only those values that were higher than the mean of the nine pairs (0.27) were included to construct the network. The width of a line between a pair of features in Figure 3 reflects the strength of association between the feature pair as computed on a log scale. Note that the mean values differed for other features (e.g., the mean for pairs involving *Time* is 0.08). Apart from the high correlation between the *GPS-Signal* pair, our results show that, in considering our mean threshold value of 0.27, *GPS* also co-occurred with *Location* frequently (p=0.44). Other feature-feature relationships whose probability values were relatively high (compared to their respective

[5] $P(GPS|Signal) = P(GPS \cap Signal) / P(Signal) = (103/4523) /(117/4523) = 103/117 = 0.88$

means), indicating comparatively stronger relationships, include *Time-Distance* (p=0.17), *Map-Location* (p=0.12), *Time-Signal* (p=0.12) and *Time-Map* (p=0.12).

Beyond our conditional probability measures (refer to Figure 3), Pearson product-moment correlation tests confirm that users expressed dissatisfaction with *GPS* and three other features most frequently – *Distance*, *Signal* and *Map* (Pearson correlation coefficients, r=0.84, 0.76 and 0.75, respectively). *Map* was found to be correlated with *Battery*, *Signal* and *Data* the most (r=0.82, 0.78 and 0.75, respectively), and *Time* correlated most closely with *Data* (r=0.75). *Signal* violations were reported most when *Version*, *Distance* and *Battery* dissatisfactions were expressed (r=0.77, 0.76 and 0.73 respectively). Finally, *Data* correlated most with *Battery* (r=0.80) and *Distance* with *Version* (r=0.72). Of note is that these results are all strong, and statistically significant (P<0.05).

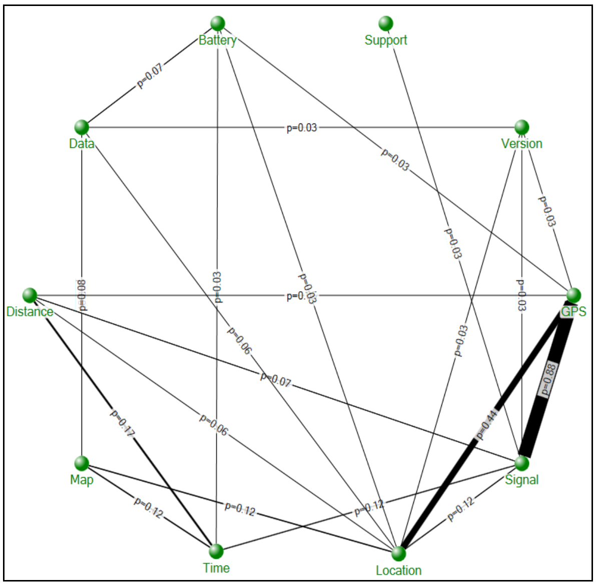

Figure 3. Network of feature-feature co-occurrences

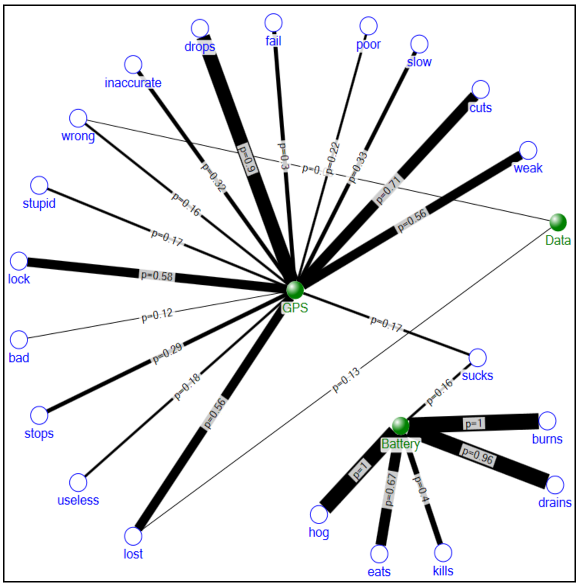

Figure 4. Network of feature-issue co-occurrences

Top Features with Expectation Violations: While we have identified the features that resulted in expectation violations, it is also important to reveal the expectations that were violated. For example, if *GPS* is problematic, the question of *what aspects of GPS* are problematic arises. To reveal the actual expectation violations, we computed the conditional probabilities between features and all other words expressed in the sentences, and then manually selected all the pairs that contained meaningful unigrams that represented issues (e.g., words such as *lost* and *drains*) whose frequency count was greater than or equal to 5. Given space limitations a sample of our results is presented in Figure 4 for three features: *GPS*, *Battery* and *Data*. We present our outcomes as a network diagram to visually depict the findings. Figure 4 shows all the issues (violations) that were associated with *Battery*, which relate heavily to *draining*. Here it is shown that verbs such as *eats*, *kills*, *drains* and *burns* frequently co-occurred with *Battery*. In fact, we believe that even the word *hog* in Figure 4 describes the battery draining issue. However, for *GPS*, a broader range of violations were recorded, as shown in Figure 4. Given this large spread of violations for *GPS*, and its high ranking in Figure 2, we plot the counts and conditional probability scores separately, in Figure 5. Here it can be seen that *lost* is the most frequent violation that was reported for *GPS*, followed by *useless*, a more generic description of a problem. The conditional probability values for the feature-issue pairs (e.g., *GPS-lost and GPS-wrong*), are shown as discrete points in Figure 5 (see the y-axis on the right for the corresponding values). The top word (violation) co-occurring with *GPS* is *drops* (80% of the time). In addition, *cuts* was complained about frequently with *GPS* (71% of the time). Furthermore, the violations *weak*, *lost* and *lock* co-occurred with *GPS* about 50% of the time.

The violations presented in Figure 5 may be further categorized into several manually constructed groups, as shown in Figure 6. From our results we may group violations as follows: (a) connectivity (*drops*, *cuts*, *stops*, *lost*, *lock*), (b) accuracy (*inaccurate*, *wrong*), (c) failure (*fail*), (d) strength (*weak*), (e) speed (*slow*), and (f) generic issues (*poor*, *stupid*, *bad*, *useless* and *sucks*). Presenting our outcomes using this classification, it can be observed in Figure 6 that *connectivity* was the top violation, accounting for 49% (109 out of 221) of the problems reported. It is also revealing to observe that *generic issues* accounted for 32% of the violations, and *failure* of GPS accounted for 19% of users' dissatisfactions.

As above, in triangulating our feature-issue conditional probability results for *GPS*-related issues (refer to Figures 4 and 5) correlation results show that *GPS-lost* was reported frequently with *GPS-fail* (r=0.79), *GPS-stop* was reported frequently with *GPS-inaccurate* (r=0.77), *GPS-lock* was reported with *GPS-inaccurate* (r=0.78) and *GPS-suck* (r=0.73), *GPS-suck* was reported with *GPS-inaccurate* (r=0.79), *GPS-wrong* was reported with *GPS-drop* (r=0.76), *GPS-drop* was reported with *GPS-slow* (r=0.78), and *GPS-poor* was reported more when *GPS-weak* was also reported the highest (r=0.75). Again, these results were all strong, and statistically significant (P<0.05).

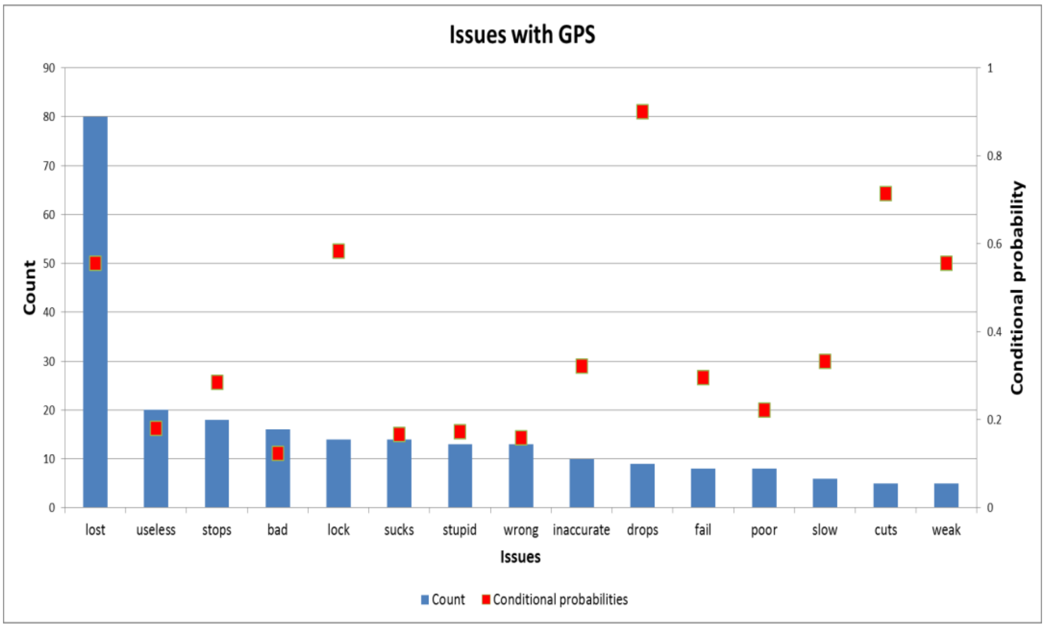

Figure 5. Issues with GPS, their counts and conditional probabilities

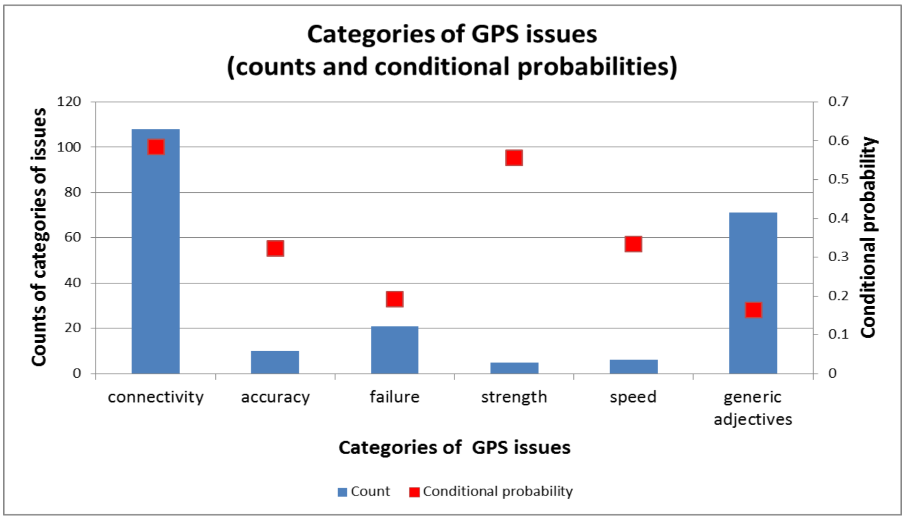

Figure 6. Categories of GPS issues (counts and conditional probabilities)

4.2. Expectation Violations in a Specific App Domain (H2)

In testing H2 we compare our results across the four apps (My Tracks, Endomondo, Nooms Cardio Trainer, Zombies Run!) that were selected from the health and fitness domain on Google Play. These results are provided in Figure 7, highlighting the distribution of the top-10 features that had expectation violations. Figure 7 reveals certain commonalities among the apps, suggesting that users of apps in the health and fitness domain had specific expectations (given the expectation violations recorded) regarding how such apps should work. For example, in Figure 7 *GPS* is the top problematic issue in all four apps, followed by *Time*. In addition, overall, there were few differences between the apps' violations. That said, analysis of the Zombies Run! app revealed a slightly different proportion of violations reported for the top-10 features when compared to the other three apps. Figure 7 shows that there were relatively more violations with the *Map* and *Location* features of this app than the other three apps (My Tracks, Endomondo and Nooms Cardio Trainer).

In fact, the other apps had similar feature violation profiles. Only subtle variations are seen for these three apps; for instance, comparatively more *Map*-related violations were reported for My Tracks than the other apps, more *GPS* violations were submitted for Nooms Cardio Trainer than for the other apps, and more *Version*-related violations were recorded for Endomondo, compared to the others (My Tracks, and Nooms Cardio Trainer). To triangulate our results in Figure 7, and to validate that these violations were indeed consistent, we performed a two-way ANOVA test. Results from this test reveal that the pattern of feature violations were not significantly different (P=0.60), tending to be relatively homogenous. This outcome confirms that community users reported similar expectation violations for the four apps, which was likely driven by their expectations of how software features should work in that domain. We discuss these findings and those above in the following section.

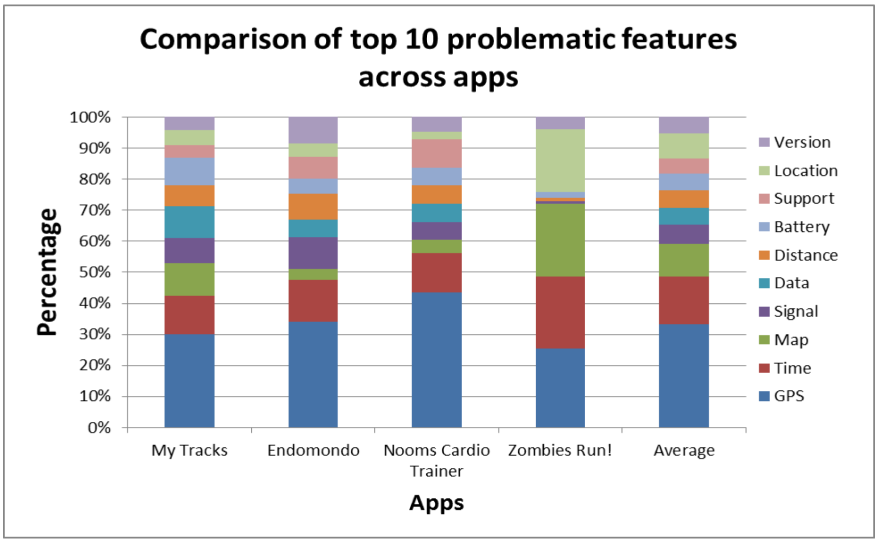

Figure 7. Comparison of top 10 problematic features across apps

5. Discussion and Implications

Expectations govern how individuals operate in societies, and how societies respond to norms and beliefs (Gamble 1982; Schrire 2009). It has been shown that expectations and norms drive cooperation and human behaviours (Fehr and Fischbacher 2004). With socio-technical systems and computer mediated interactions becoming central to the way societies operate, it is pertinent to understand how this issue is managed (Cranefield 2014). The mobile app community in particular has grown exponentially over recent years, resulting in large volumes of feedback reflecting individuals' expectations for various forms of apps. Previous theories (EVT and ECT) explaining the influence of expectations on individuals' beliefs and fulfilment suggest that stakeholders of the mobile app community should indeed be affected by their expectations of the artefacts they use. We have used these theories as a basis for testing two hypotheses H1 and H2. We revisit these in this section, in considering our results in the previous section (Section 4).

H1*. The mobile app community responds to expectation violations (with resentment) in their reviews, and these expectation violations may be about specific features and issues*. Our outcomes reported in the previous section indeed confirm that the mobile app community responded to expectation violations about specific features and issues through the use of reviews, we must thus accept H1. We noticed that, for the apps studied, some specific features did not operate as users expected they would, which gave rise to numerous complaints. **This evidence confirms the relevance of EVT and ECT**

for studying expectation violations in socio-technical systems, beyond human-human applications (Fehr and Fischbacher 2004; Gamble 1982; Schrire 2009; Sugden 1998). EVT notes that individuals form expectations based on societal rules which drives their beliefs about future events, and when the expectation of an individual is violated they react in compensation (Milewski 2004; Petter 2008). In our context app users logged reviews complaining about app features' inadequacies. For instance, concerns over *GPS*, *Time* and *Map* were raised in frequent numbers by those tracking their exercise (e.g., walks and runs) using the apps we investigated. Beyond making such complaints, users would likely discontinue using apps that do not address their expectations – such a move has previously been shown to result from users' dissatisfaction (Bhattacherjee 2001; Lankton and McKnight 2006). ECT notes that customer fulfilment is influenced by previous expectations, perceived performance, and disconfirmation of beliefs. While we did not examine the changes in adoption and continued usage of the apps over time in our inquiry, there is little doubt that *users would discontinue usage of apps on the mobile app stores if the features presented by these apps result in continuous expectation violations*. We thus believe that such an investigation is warranted given our evidence for the many expressions of expectation violation.

We observed that some specific app features were reported together, supporting the outcomes of previous work that looked into how complaints in issue logs may mirror actual software modules' interconnections (Licorish et al. 2015a). While this outcome is useful for the app development community, in terms of providing insights into how complex defects (or inadequacies) may be detected from users' complaints, it also shows that *users' expressions of dissatisfaction for mobile apps are not restricted*. In fact, we noted a range of violations that were expressed, and in some contexts users compensated for violations with added frustration and resentment (e.g., *GPS* was *useless*). Of course in the mobile app context – that is, a socio-technical system or virtual community – such an external behavioural change (i.e., a verbal sanction) is fitting, which could no doubt result in loss of reputation (Burgoon 1993; Burgoon and Hale 1988). Since there are limited if any opportunities for users to meet with app providers, users may freely express their feelings online when frustrated, which provokes the thought - *perhaps virtual communities promote greater willingness to express verbal sanctions when expectations are violated?* Confirming such a finding with follow up work could have implications for theory, leading for instance to contextualization of EVT and ECT. In the same vein, understanding how expectations in a particular mobile app domain are reported, and what issues are of most concern, could provide additional context about the suitability of applying EVT and ECT to the studying of users' expectation violations in socio-technical systems, and how such theories may be extended. We further consider this issue next.

***H2**. Users of a specific app domain will express similar expectation violations given their expectations of how software features should work in that domain*. Our results in Figure 7 show mild differences in the features that violated expectations across the four apps that were taken from the health and fitness domain. However, taken together, these differences were not significant. In fact, features that resulted in the high number of violations (i.e., *GPS* and *Time*) were consistent for all apps; we must thus accept H2. That said, only between 12% and 17% of review sentences were expectation violations. Our findings suggest that, as with how individuals may develop specific expectations of others given certain norms (e.g., culture-specific norms and conventions (Bailenson et al. 2001; Bailenson et al. 2003; Yee et al. 2007)), **app users grow to possess specific expectations and beliefs about the features that are delivered in apps given their purpose**. Our outcomes also confirm that violations of expectations may be reported in a range of media outlets, including app review portals (McLaughlin and Vitak 2012; Ramirez and Wang 2008). Since the app domain at hand concerns the tracking of exercise (e.g., walks and runs) it is understandable that users of all four apps used *GPS* to identify their location over time. Users employing such apps also track the *Time* taken to perform specific exercises. In our context, when these two features did not work as expected, users complained and expressed resentment. A requirement (i.e., an expectation) for such apps seems to be *the efficient functioning of GPS and Time features*. This finding has implications for theory as our findings here show that human-software expectations may vary across app domains, and again, also points to the applicability of EVT and ECT to the study of expectation violations in socio-technical systems. Beyond theoretical relevance, insights from this analysis could be beneficial to developers in helping them to prioritize features that need to be delivered and/or improved in future versions of apps

(Dam et al. 2015). In addition, our evidence suggests that developers of health and fitness apps should focus on delivering functional *GPS* feature and accurately recorded users' *time*. Users could also utilise our findings to identify problematic features in a given app domain, or in certain apps, thus, informing their decision to use or not use an app. Our analyses could scale to other app domains to identify the key expectations of users, and how these are addressed by the community of app developers. In fact, beyond the mobile community, such insights may be revealed for other online communities and socio-technical systems (e.g., restaurants, movies, tourist attractions, transport services) where users provide feedback.

6. Threats to Validity

Our work only considered review sentences that were rated =< 3 to be potential responses to expectation violations. While previous work has indicated that such reviews largely contain expressions of dissatisfaction (Fu et al. 2013), we concede that reviews rated 4 and 5 may also contain expectation violations which are not considered in this work. We acknowledge that we may have missed some negative words, and particularly those that may be specific to software users, but not captured as a part of the LIWC library (e.g., uninstall, freeze, timeout). That said, our feature-issue (expectation violation) analysis did not unearth a high incidence of such words, suggesting that their occurrence was limited or unlikely. In addition, our reliability assessment (our formal reliability test returned 96% agreement – refer to Section 3.1) also offers some form of triangulation for our findings. We also concede that some nouns (features) may have been missed by the POS tagger, although, this tool has been validated through repeated use across numerous domains (Guzman and Maalej 2014; Licorish et al. 2015a). Finally, we are aware that the review sentences that were summarised in this work do not necessarily represent those of all users of the apps considered, and furthermore, some introverted users may have sentiments but suppress them altogether. Thus, our findings do not necessarily provide a complete picture of app users' views. That said, we believe that our approach and the insights that are provided could be valuable to stakeholders of the app community in general.

7. Conclusion

As modern societies continue to shift towards decentralised structures, interactions between globally distributed individuals and teams are increasingly mediated through ICT. With this shift there are also growing expectations about how entities should behave, giving rise to the need to understand expectations is socio-technical systems. We have thus taken a first step towards decomposing this issue, and have studied a specific socio-technical system, the mobile app community, using EVT and ECT as our theoretical bases. We tested two hypotheses, H1 and H2, in addressing this objective. Our outcomes confirm that the mobile app community responded to expectation violations through the use of reviews, and that users of health and fitness apps have similar expectations. While our outcomes confirm the relevance of EVT and ECT for studying expectation violations in socio-technical systems, we believe that such theories may also be extended with confirmation of our findings from follow up work. For instance, since there are limited opportunities for users to meet with app providers, users in virtual communities may freely express verbal sanction when frustrated with app features. However, such willingness may diminish in face-to-face settings. The scale of (and responses to) expectation violations may thus be mediated depending on the context. Furthermore, similar to the way culture-specific norms drive individuals' expectations, specific app domains may drive users' expectations, an outcome we confirmed in this work. Beyond theoretical implications, our findings in this work may provide useful insights for app developers and the app user community. Our next step is to consider a larger sample of apps in a domain, and then multiple apps across several domains. This exercise may corroborate our outcomes, both in terms of validating the suitability of the EVT and ECT theories for studying expectations in socio-technical systems, and the practical relevance of our outcomes for the mobile app community. We also encourage future use of our approach to study expectations in other relevant domains.

References

"Iso/Iec 9126-3: Software Engineering - Product Quality - Part 3: Internal Metrics," ISO/IEC, Japan, p. 81.

Bailenson, J. N., Beall, A. C., Blascovich, J., Raimundo, M., and Weisbuch, M. 2001. "Intelligent Agents Who Wear Your Face: Users' Reactions to the Virtual Self," *Intelligent Virtual Agents*: Springer, pp. 86-99.

Bailenson, J. N., Blascovich, J., Beall, A. C., and Loomis, J. M. 2003. "Interpersonal Distance in Immersive Virtual Environments," *Personality and Social Psychology Bulletin* (29:7), pp. 819-833.

Barry, B., and Crant, J. M. 2000. "Dyadic Communication Relationships in Organizations: An Attribution/Expectancy Approach," *Organization Science* (11:6), pp. 648-664.

Bhattacherjee, A. 2001. "Understanding Information Systems Continuance: An Expectation-Confirmation Model," *MIS quarterly*), pp. 351-370.

Bonito, J. A., Burgoon, J. K., and Bengtsson, B. 1999. "The Role of Expectations in Human-Computer Interaction," *Proceedings of the international ACM SIGGROUP conference on Supporting group work*: ACM, pp. 229-238.

Burgoon, J. K. 1993. "Interpersonal Expectations, Expectancy Violations, and Emotional Communication," *Journal of Language and Social Psychology* (12:1-2), pp. 30-48.

Burgoon, J. K., and Hale, J. L. 1988. "Nonverbal Expectancy Violations: Model Elaboration and Application to Immediacy Behaviors," *Communications Monographs* (55:1), pp. 58-79.

Cranefield, S. 2014. "Agents and Expectations," in *Coordination, Organizations, Institutions, and Norms in Agent Systems Ix*. Springer, pp. 234-255.

Dam, H. K., Savarimuthu, B. T. R., Avery, D., and ghose, A. 2015. "Mining Norms in Open Source Software Repositories," *In proceedings of the 37rd International Conference on Software Engineering (To appear in NIER track)*: ACM.

Dengel, A. 2014. "Aspect-Based Sentiment Analysis of Amazon Reviews for Fitness Tracking Devices," *Trends and Applications in Knowledge Discovery and Data Mining: PAKDD 2014 International Workshops: DANTH, BDM, MobiSocial, BigEC, CloudSD, MSMV-MBI, SDA, DMDA-Health, ALSIP, SocNet, DMBIH, BigPMA, Tainan, Taiwan, May 13-16, 2014. Revised Selected Papers*: Springer, p. 50.

Fehr, E., and Fischbacher, U. 2004. "Social Norms and Human Cooperation," *Trends in cognitive sciences* (8:4), pp. 185-190.

Fu, B., Lin, J., Li, L., Faloutsos, C., Hong, J., and Sadeh, N. 2013. "Why People Hate Your App: Making Sense of User Feedback in a Mobile App Store," *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*: ACM, pp. 1276-1284.

Gamble, C. 1982. "Interaction and Alliance in Palaeolithic Society," *Man*), pp. 92-107.

Gao, X., and Singh, M. P. 2014. "Extracting Normative Relationships from Business Contracts," *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*: International Foundation for Autonomous Agents and Multiagent Systems, pp. 101-108.

Gürkök, H., Hakvoort, G., and Poel, M. 2011. "Evaluating User Experience with Respect to User Expectations in Brain-Computer Interface Games,").

Guzman, E., and Maalej, W. 2014. "How Do Users Like This Feature? A Fine Grained Sentiment Analysis of App Reviews," *Requirements Engineering Conference (RE), 2014 IEEE 22nd International*, pp. 153-162.

Houser, M. L. 2006. "Expectancy Violations of Instructor Communication as Predictors of Motivation and Learning: A Comparison of Traditional and Nontraditional Students," *Communication Quarterly* (54:3), pp. 331-349.

Iacob, C., and Harrison, R. 2013. "Retrieving and Analyzing Mobile Apps Feature Requests from Online Reviews," in: *10th IEEE Working Conference on Mining Software Repositories (MSR)*. San Francisco, CA: IEEE, pp. 41-44.

Kalia, A., Motahari Nezhad, H. R., Bartolini, C., and Singh, M. 2013. "Identifying Business Tasks and Commitments from Email and Chat Conversations," tech. report, HP Labs.

Lankton, N., and McKnight, D. H. 2006. "Using Expectation Disconfirmation Theory to Predict Technology Trust and Usage Continuance Intentions," *Eli Broad College of Business, Minnesota[Links]*).

Licorish, S. A., Lee, C. W., Savarimuthu, B. T. R., Patel, P., and MacDonell, S. G. 2015a. "They'll Know It When They See It: Analyzing Post-Release Feedback from the Android Community," in: *21st Americas Conference on Information Systems (AMCIS 2015)*. Puerto Rico: AISeL, pp. 1-11.

Licorish, S. A., and MacDonell, S. G. 2013. "The True Role of Active Communicators: An Empirical Study of Jazz Core Developers," in: *17th International Conference on Evaluation and Assessment in Software Engineering (EASE 2013)*. Porto de Galinhas, Brazil: ACM, pp. 228-239.

Licorish, S. A., MacDonell, S. G., and Clear, T. 2015b. "Analyzing Confidentiality and Privacy Concerns: Insights from Android Issue Logs," in: *19th International Conference on Evaluation and Assessment in Software Engineering*. Nanjing, China: ACM.

Licorish, S. A., Tahir, A., Bosu, M. F., and MacDonell, S. G. 2015c. "On Satisfying the Android Os Community: Users' Feedback Still Central to Developers' Portfolio," in: *24th Australasian Software Engineering Conference (ASWEC 2015)*. Adelaide, Australia: IEEE Computer Society, p. TBC.

Lin, J., Amini, S., Hong, J. I., Sadeh, N., Lindqvist, J., and Zhang, J. 2012. "Expectation and Purpose: Understanding Users' Mental Models of Mobile App Privacy through Crowdsourcing," *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*: ACM, pp. 501-510.

MacMillan, K., and Koenig, T. 2004. "The Wow Factor Preconceptions and Expectations for Data Analysis Software in Qualitative Research," *Social Science Computer Review* (22:2), pp. 179-186.

McLaughlin, C., and Vitak, J. 2012. "Norm Evolution and Violation on Facebook," *New Media & Society* (14:2), pp. 299-315.

Meyer, J.-J. C., and Wieringa, R. J. 1994. *Deontic Logic in Computer Science: Normative System Specification*. John Wiley and Sons Ltd.

Milewski, A. E. 2004. "Software Engineers and Hci Practitioners Learning to Work Together: A Preliminary Look at Expectations," *Software Engineering Education and Training, 2004. Proceedings. 17th Conference on*: IEEE, pp. 45-49.

Obaid, M., Damian, I., Kistler, F., Endrass, B., Wagner, J., and André, E. 2012. "Cultural Behaviors of Virtual Agents in an Augmented Reality Environment," *Intelligent Virtual Agents*: Springer, pp. 412-418.

Oliver, R. L. 1977. "Effect of Expectation and Disconfirmation on Postexposure Product Evaluations: An Alternative Interpretation," *Journal of applied psychology* (62:4), p. 480.

Oliver, R. L., and Burke, R. R. 1999. "Expectation Processes in Satisfaction Formation: A Field Study," *Journal of Service Research* (1:3), pp. 196-214.

Pagano, D., and Maalej, W. 2013. "User Feedback in the Appstore: An Empirical Study," in: *21st IEEE International Requirements Engineering Conference*. Rio de Janeiro: IEEE, pp. 125-134.

Pennebaker, J. W., Francis, M. E., and Booth, R. J. 2001. "Linguistic Inquiry and Word Count (Liwc): A Computerized Text Analysis Program," *Mahwah (NJ)* (7).

Petter, S. 2008. "Managing User Expectations on Software Projects: Lessons from the Trenches," *International Journal of Project Management* (26:7), pp. 700-712.

Poggi, I., Pelachaud, C., de Rosis, F., Carofiglio, V., and De Carolis, B. 2005. "Greta. A Believable Embodied Conversational Agent," in *Multimodal Intelligent Information Presentation*. Springer, pp. 3-25.

Ramirez, A., and Wang, Z. 2008. "When Online Meets Offline: An Expectancy Violations Theory Perspective on Modality Switching," *Journal of Communication* (58:1), pp. 20-39.

Savarimuthu, B. T. R., and Dam, H. K. 2014. "Towards Mining Norms in Open Source Software Repositories," in *Agents and Data Mining Interaction*. Springer, pp. 26-39.

Schrire, C. 2009. *Past and Present in Hunter Gatherer Studies*. Left Coast Press.

Sugden, R. 1998. "Normative Expectations: The Simultaneous Evolution of Institutions and Norms," *Economics, values, and organization* (73).

Tausczik, Y. R., and Pennebaker, J. W. 2010. "The Psychological Meaning of Words: Liwc and Computerized Text Analysis Methods," *Journal of language and social psychology* (29:1), pp. 24-54.

Toutanova, K., Klein, D., Manning, C. D., and Singer, Y. 2003. "Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network," in: *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*. Edmonton, Canada: Association for Computational Linguistics, pp. 173-180.

Trist, E. 1981. "The Evolution of Socio-Technical Systems," *Occasional paper* (2), p. 1981.

Vasa, R., Hoon, L., Mouzakis, K., and Noguchi, A. 2012. "A Preliminary Analysis of Mobile App User Reviews," *Proceedings of the 24th Australian Computer-Human Interaction Conference*: ACM, pp. 241-244.

Welch, C. 2013. "Google: Android App Downloads Have Crossed 50 Billion, over 1m Apps in Play. The Verge [July 24, 2013]."

Wood, W., and Eagly, A. H. 2002. "A Cross-Cultural Analysis of the Behavior of Women and Men: Implications for the Origins of Sex Differences," *Psychological bulletin* (128:5), p. 699.

Yee, N., Bailenson, J. N., Urbanek, M., Chang, F., and Merget, D. 2007. "The Unbearable Likeness of Being Digital: The Persistence of Nonverbal Social Norms in Online Virtual Environments," *CyberPsychology & Behavior* (10:1), pp. 115-121.

Yu, J., Zha, Z.-J., Wang, M., and Chua, T.-S. 2011. "Aspect Ranking: Identifying Important Product Aspects from Online Consumer Reviews," *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*: Association for Computational Linguistics, pp. 1496-1505.