Analyzing Users' Narratives to Understand Experience with Interactive Products

Alexandre N. Tuch
a.tuch@unibas.ch

Rune N. Trusell
runetrusell@gmail.com

Kasper Hornbæk
kash@diku.dk

Department of Computer Science, University of Copenhagen
Njalsgade 128, 2300 Copenhagen, Denmark

ABSTRACT

Recent research in user experience (UX) has studied narratives, users' account of their interaction with technology. It has emphasized specific constructs (e.g., affect, needs, hedonics) and their interrelation, but rarely analyzed the content of the narratives. We analyze the content and structure of 691 user-generated narratives on positive and negative experiences with technology. We use a multi-method approach consisting of manual (structural analysis of narratives) as well as of automated content analysis methods (psycholinguistic analysis and machine learning). These analyses show converging evidence that positive narratives predominantly concern social aspects such as family and friends. In addition, technology is positively experienced when it enables users to do things more efficiently or in a new way. In contrast, negative narratives often express anger and frustration due to technological failures. Our multi-method approach illustrates the potential of automated (as opposed to manual) content analysis methods for studying text-based experience reports.

Author Keywords

User experience.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Human Factors; Design; Measurement.

INTRODUCTION

User experience (UX) is an emerging field of research that studies both users' experience with interactive products and how we may design such products to set about certain experiences [3, 11]. UX has taken a holistic view [19], emphasizing all aspects of product use, including expectations and experiences following the use situation. One implication of this view is that users' experience is indivisible and needs to be studied as a whole.

CHI 2013, April 27–May 2, 2013, Paris, France.

The holistic approach to UX can be seen in recent research on user-generated descriptions of experiences with interactive products [e.g., 9, 15, 21, 23]. We name such descriptions of personal experiences *narratives,* and assume that they either directly represent experience or that they are a key component of remembering and making sense of experience. Analysis of users' narratives has made several contributions to our understanding of UX. For instance, Hassenzahl et al. [9] had users report stories on positive experiences with interactive products and rate need fulfillment, hedonic/pragmatic quality, and emotion. Thereby, the study was able to link need fulfillment and emotion. Partala and Kallinen [23] extended the study of Hassenzahl by also assessing negative experiences and providing some qualitative data from the narratives. Korhonen et al. [15] took another approach. They used open-ended stories of experiences with personal devices such as smart phones to identify key aspects of UX and the role of context in shaping those experiences. Karapanos et al. [12] used narratives in conjunction with quantitative measurements to study UX over time.

Despite these contributions, the analysis of narratives is often limited to the listing of simple descriptive information (e.g., what kind of device was used). As a consequence, we have little systematic knowledge of the actual content of the narratives. Hassenzahl et al. [9], for instance, noted that "we were not successful in further classifying the content of experiences. The descriptions were just too different in length, style and depth." (p. 357). Partala and Kallinen [23] did some analyses of the content of narratives, but noted that "current qualitative results highlight the need for further developing systematic methods for qualitative reflections on personal experiences" (p. 32). Thus, it appears that we can learn more from the content of narratives.

The present study analyses the content and structure of 691 narratives obtained from an online questionnaire about experience with interactive products. We combine manual and automated processing to contrast positive and negative experiences in regard to content, narrative structure, need fulfillment, affect, and product qualities. Two contributions are made: (1) we provide a systematic overview of the content of a large sample of authentic user experiences and (2) we offer a methodological approach that enables the analysis of large-scale qualitative data.

RELATED WORK

UX research strives to describe and understand people's experience with interactive products [3, 11]. It takes a holistic view [11, 19], by studying anticipation, use, storytelling around use, and recommendations of products. Methodologically, UX ranges from qualitative research of experiences with in-depth interviews to quantitative data from questionnaires on the correlation among ratings of affect and technology perceptions. It has been observed, though, that different methodologies are rarely mixed [3]. One aim of the present paper is hence to combine different methods for the analysis of narratives on user experience.

Storytelling is natural and people do it all the time when sharing personal experiences of their everyday lives. Especially emotion-arousing events or episodes are frequently narrated to third parties [27]. In the context of UX research, information about events or episodes that elicit emotional responses are of great interest. Such information allows us to study the circumstances under which users experience significant events with technology. Consequently, asking users to speak or write about such an event might serve as a rich and meaningful data source for studying UX.

Several studies have investigated UX by means of user narratives. For instance, Hassenzahl and colleagues [9] conducted an online study where 500 participants reported a recent positive experience with an interactive product. Participants also had to evaluate the reported experience in terms of need fulfillment and affect, and rate the product on pragmatic and hedonic quality. The study could thereby investigate the interplay between product qualities and psychological aspects of user experience employing a large sample of diverse and authentic experiences with technology. Although the study gave important insights into the role of need fulfillment in UX, it (1) studied only positive experiences and (2) analyzed the content of the narratives in little detail only. It is interesting to understand positive user experiences (as designers aim at creating positive experiences with their products), but focusing only on positive experiences may bias our understanding of UX and may confound the content of narratives in general with the characteristics of positive experiences.

In contrast to the study of Hassenzahl and colleagues, Partala and Kallinen [23] also assessed negative experiences. They had 45 participants report a recent positive and negative experience where technology formed a substantial part. Participants evaluated each experience on the same questionnaires as in [9]. Partala and Kallinen were therefore able to contrast positive and negative experiences to specific affective responses and need fulfillment. Their findings offer insights on which needs and emotions are most important within positive and negative experiences. In addition, the content of the narratives were analyzed in a more detailed way, providing interesting insight into the difference between positive and negative experiences: (1) positive experiences were more often related to fist-time usage, stressing the importance of novelty and surprise within positive user experiences, (2) negative experience typically occur in a more hurried context, and (3) technical and usability problems are related to negative experiences. However, as mentioned in the introduction, Partala and Kallinen [23] recognized the need for richer analyses of narratives.

Karapanos et al. [12] also used narratives to describe how the experience of six iPhone users developed over a five-week period. Based on an analysis of narratives they proposed a model of how user experience develops over time. In particular, they showed how identification with the technology became increasingly prominent as users became more experienced. Also, learnability issues decreased over time. This study is tied closely to the experience of these particular individuals and to the iPhone as a product.

Other research has also collected narratives [2, 15, 21]. But similarly to the above-mentioned studies, it does not link ratings and narratives and rarely analyze the content of narratives in detail. Thus, we pursue the hypotheses that analysis of narratives may give us additional and potentially important information about user experience.

METHOD

The data on experiences were collected with a web-based questionnaire comprising 73 questions (see Table 1). The goal of the questionnaire was to collect qualitative data on both positive and negative experiences. In addition, we wanted to follow and potentially replicate earlier work by collecting quantitative data on need fulfillment, affect and technology perception (as in [9, 23]).

Design

The questionnaire crossed two factors between participants: (1) quality of the reported experience (positive vs. negative) and (2) wording of questionnaire items (plain vs. negated). Regarding the first factor, approximately one half of the participants (N = 368) were required to report on a positive experience; the other half (N = 323) was required to report on a negative experience. Our aim was to allow comparison of positive and negative experiences, as previously done in a HCI by [23] and in the field of psychology by [31].

The second factor concerned the wording of questionnaire items on need fulfillment. When asking about negative experiences, Sheldon et al. [31] altered the wording of the items so that they became negatives. It is not clear, however, how answers are affected (Sheldon et al. did not analyze this). Thus, half of our participants answered the negated form of the need fulfillment question (both for positive and negative experiences); the other half answered the plainly formulated questions (as in [23]). As we do not have room to explore the effect of wording on response behavior, we use only the plain items for the analysis of need fulfillment.

Questions

Table 1 summarizes the 73 questions asked. The key item was an open-ended question that attempted to get a narrative description of an experience with an interactive product:

> Bring to mind a single outstanding positive experience you have had recently with interactive technology. Think of positive in whatever way makes sense to you. Please retell the experience as accurately and detailed as you remember, and try to be as concrete as possible. You can use as many words as you like, so that outsiders can easily understand your experience.

The negative group had the same description, expect that occurrences of "positive" in the question above were exchanged with "negative". We also made it clear to participants to describe the experience (rather than the technology) and gave them examples of such experiences (that did include examples of interactive technologies). This question provided the context for the remaining questions and participants were reminded throughout the questionnaire to answer all questions in conjunction to the reported experience.

We asked three questions on the **context of the experience**. These questions were taken from the Geneva Appraisal Questionnaire (GAQ, [29]) and concerned when, where, and with whom the experience took place.

Questions on **need fulfillment** were based on [9] and [31]. They relate to constructs of seven key psychological needs, such as competence, popularity, stimulation, and autonomy; each need was gauged with three questions. For five of these constructs, Hassenzahl et al. only used two questions for each need; we added the original formulations by Sheldon et al. [31]. As in the study by Hassenzahl et al., we did not ask about three further needs from the Sheldon paper (i.e., luxury, self-esteem and physical thriving).

We assessed **affect** using the Positive Affect Negative Affect Schedule (PANAS) questionnaire [33]. PANAS has been validated in many studies as a reliable predictor of affect. It contains 20 sentences each asking about a positive (e.g., active, alert, proud) or a negative adjective (e.g., hostile, irritable, jittery).

In addition, we asked questions about the **technologies** involved in the experience. In earlier studies [9] technology perception in the context of user experience has been assessed using the questionnaire AttrakDiff 2 [8]. As in [9], we used an abridged version of the questionnaire, gauging pragmatic and hedonics quality (each with 4 items) and overall evaluation of the product (with 2 items; see [10]).

We asked two groups of question based on earlier work, but they are only of minor relevance for the present paper. They were about the extent to which experiences are **attributed** to technology [9]. And we asked questions about participants' background and technology literacy.

Experience [9]
1 open question: "Bring to mind a single outstanding positive experience you have had recently"
Context [29]
3 questions with ordinal and nominal options: "How long ago did the experience occur?" ("some hours ago" – "some years ago"); "Where were you when you had the experience?" (e.g., "in my home", "a natural setting", "in the street or another public place"); "Who was present when you had this experience" (e.g., "Nobody. I was alone", "A partner or friend", "Several friends or acquaintances")
Need-fulfillment [9]
21 questions answered as "not at all" (1) to "extremely" (5): For instance, "During this experience I felt I was successfully completing difficult tasks and projects"; "During this experience I felt that I was 'becoming who I really am'"; "During this experience I felt free to do things my own way"
Affect [33]
20 questions, each about a negative or positive adjective, answered as "not at all" (1) to "extremely" (5): For instance, During this experience, to what extent did you feel excited"; "During this experience, to what extent did you feel afraid"
Technology [-]
1 open question about the technology used; 3 questions about frequency of use: "Which specific technology did you use when you had this experience?"; "For how long have you used this specific technology"; "How often do you use this specific technology"; "How much time do you spend on average when you're using this technology"
Technology perception [9]
10 questions, going from a negative to a positive endpoint (1 to 7): For instance, "I perceived the technology that I have used during the experience as confusing … dull"
Attribution [9]
1 question from "To a very small extent" (1) to "To a very large extent" (5); 1 open text question: "To what extent do you feel that the technology was the cause of the specific experience?"; "Why?"
Background [-]
7 questions about technology literacy; 5 demographic question: For instance, "How often do you use a computer?"; "Please indicate your level of education"

Table 1. Questions used in the survey.

Participants

Participants were recruited at Amazon Mechanical Turk through the intermediary company Crowdflower (which allows people without a US-credit card to post jobs to Mechanical Turk). A total of 938 participants started filling out the questionnaire. Of these, 706 completed the questionnaire and answered three verification questions at the end of the questionnaire. The verification questions were added after pilot testing and required participants to describe the purpose of the study without being able to go back and look at earlier questions or guidelines. We further reduced the number acceptable answers to 691 by excluding 15 participants for reporting vague experiences or for attempting to repeat earlier responses.

The majority of participants lived in the US (81%), though participants from 48 other countries were included (Canada, 2%, as the second most frequent country). Participants were between 15 and 71 years old (M = 32.7, Mdn = 30). 58% were women, 42% men. A majority of participants (87%) had attended or graduated from college and either worked full-time (37%), studied (20%), or were unemployed (15%).

Participants received 1.5 US dollars for completing the questionnaire, corresponding to an hourly salary of 5.1 dollars based on timings from a pilot study. The literature suggests that paying too much or too little may affect answer quality [5, 13]; the hourly salary was high compared to other studies but seemed reasonable.

Procedure

Participants were directed to a questionnaire hosted at surveygizmo.com. Participants could not go back to earlier questions. All questions except one on age were mandatory because we wanted a full data set. In pilot studies, completing the questionnaire took about 17.5 min.

ANALYSIS APPROACHES

We use three complementary ways to analyze the content of the narratives.

- We manually coded narrative structure [17, 18] and identified main themes of each structural element [1, 4]. Earlier work on user-generated descriptions of experiences [e.g., 9] has not made any assumptions about the type of description; narrative theory may help understand the content of narratives and provide a tested way of identifying structural elements.
- We used Linguistic Inquiry and Word Count (LIWC; [24, 25]). The LIWC is a fully automated tool to analyze the content of texts. It counts the occurrence of words of specific categories (such as Money, Religion, Family). LIWC has been used earlier in HCI [e.g., 16, 30]; its main use in this paper is to automatically compare the content of positive and negative narratives.
- We used a maximum entropy classifier [20] to distinguish positive and negative narratives. The classifier is a machine learning approach that finds words distinguishing positive and negative narratives. Its main use here is to find words and themes in narratives that LIWC does not capture.

Coding Narrative Structure and Identifying Themes

According to linguistic research, narratives of personal experiences have a common structure [7, 17, 18]. In a seminal paper, Labov and Waletzky [18] analyzed verbal narratives of "unsophisticated speakers" and found that such speakers follow a common narrative structure when reporting personal experiences. This common structure consists of six elements, which usually occur in a fixed order. In the following, we describe these elements and provide examples from our user narratives.

In the **abstract** the narrator briefly summarizes the story by announcing its main theme (e.g., "*Let me tell you the story, where I broke my new phone.*"). An abstract serves to attract the listener's interest. In our context, abstracts are expected to occur rarely as participants are unlikely to feel a need to attract our interests (as we asked them explicitly to narrate a personal experience).

In the **orientation** the narrator provides the context of the story by identifying time, place, persons, and their activities or the situation (e.g., "*I was on the computer with my son who lives in Las Vegas, and we were skyping for the first time. My grand daughter was born four days before…*").

In the **complication** the main event of the narrative is reported, usually something noteworthy or unexpected. It typically consists of a series of events that lead to a climax ("*I entered the name of the town and the road number, A508, as this was all I knew about the route. I started across country, recognizing some landmarks but was surprised at a roundabout to be told to turn right, but did as instructed. I followed the route but was taken on to a major road that I had wanted to avoid, and in to road works, and I began to feel really frustrated as it was not the route I had expected to be on.*").

In the **evaluation** the narrator comments on why the story is interesting or noteworthy. Such comments are likely to occur near the end, but also throughout the story (e.g., "*Google translate isn't a perfect translation but it does the trick. [...] It seems minor, but for me I can read just about anything I like anywhere on the web and I absolutely love that.*").

The **resolution** represents the outcome of the story, its conclusion. It is directly related to the complication (e.g., "*Eventually, I did end up on the right road but I had definitely not gone the most direct route...*").

Most stories end after the resolution, but sometimes the narrator adds a **coda** to signal the end of the story. The perspective of the story then returns to the present and its relevance for today is highlighted (e.g., "*I keep browsing through old files to refresh some more memories, drinking coffee and enjoying this time travel*").

Besides Labov and Waletzkys's structure, many other approaches exist. However, we think it fits well with our goal of systematically analyzing the content of short user experience reports as it is based on brief, topic-centered narratives (pp. 102 [26]).

Because this coding was very time consuming, only half of the narratives (the plain wording condition, N = 338) were coded. For the coding Labov's structural elements were slightly modified. Following the reasoning of Habermas et al. [7], we (1) added the structural element **attempt to solve the complication** and (2) did not code evaluation as an autonomous narrative sections, because evaluations may appear in any section of the narrative structure. In the "attempt to solve" element the narrator report on trying to revert the complication to normal (e.g., "*I then used the instruction manual to find other ways of programming the remote*."). Not all narratives contain this structural element.

Before beginning with the coding, a research assistant divided each narrative into propositions (as done in earlier work). We defined propositions in line with [6] as the smallest possible units of a text that can be believed, doubted, or denied or is either true or false. Propositions facilitate coding and allow for the calculation of the inter-rater reliability.

All narratives were coded by only one coder, except for a subset of 60 narratives, which have also been coded by the first author. Based on this subset inter-rater agreement levels were calculated. An agreement of 86% was achieved on dividing narratives into proposition. Moreover, Cohen's kappa indicated a substantial agreement for structural elements ($\kappa = .75$) and evaluation elements ($\kappa = .61$).

To identify the main themes in structural elements we used an adapted version of the affinity diagram technique [1, 4]. Within three sessions lasting each about two hours, two of the authors identified and classified the main themes within the orientation, complication, and resolution elements for each narrative. We did not do this for the other elements because they are used less frequently and mostly contain redundant content (abstract, coda). Moreover, the attempt to solve was merged with the complication and not analyzed separately. After each session the first author consolidated the grouping of the themes by rereading each narrative and checking the plausibility of the identified theme and subsequent grouping of the themes.

Linguistic Inquiry and Word Count (LIWC)

The frequency of word-use in narratives inter alia reflects the writer's psychological processes (e.g., affective, perceptual and social processes) and personal concerns (e.g., work, home, leisure and money). The development of the word categories of LIWC is based on a wide array of texts, including emails, speeches, poems, or transcribed daily speech. The LIWC is a well recognized within the field of psychology and has been applied in numerous studies linking daily word use to a broad array of real-world behaviors (for an overview, see [32]).

In the present study we used the LIWC2007 [24] to analyze the content of $N = 691$ narratives with regard to users' social/affective processes and personal concerns.

Machine Learning: Maximum Entropy Classifier

We used a maximum entropy classifier [20] to separate negative from positive narratives. We used a freely available implementation of such a classifier, the Stanford Classifier (http://nlp.stanford.edu/software/classifier.shtml, [14]). We used the classifier on default settings; the only advanced settings we used were SplitWordPairs and SplitWordShape. Earlier work suggests that classification on emotions and sentiments may be harder than more classic text classification tasks ([22] achieves 77%-83% in sentiment classification of reviews). Yet, if the classifier distinguishing positive from negative narratives achieves similar accuracy then its high-loading features may serve as a source of insight into the content of narratives.

Questions about the context	Negative (N=323)	Positive (N=368)
Where were you when you had this experience?		
In my own home	54.5%	44.6%
In the street or another public space	26.3%	25.8%
At work	9.0%	7.6%
In a public building or in a stranger's home	5.0%	10.3%
In a natural setting	3.1%	6.5%
In the home of friends or acquaintances	2.2%	5.2%
Who was present when you had this experience?		
Nobody. I was alone	51.7%	39.9%
A partner or friend	28.2%	26.9%
Several friends or acquaintances	8.4%	16.8%
Another person (acquaintance or colleague)	5.3%	6.3%
One or more persons unknown to me	2.5%	4.1%
A large crowd	4.0%	6.0%

Note. Percentages refer to the *N* indicated at the top of the corresponding row.

Table 2. Differences in the context of reported experiences.

RESULTS

In the first part of this section we analyze the ratings from the questionnaire by comparing positive and negative experiences. In the second part we examine the content of the narratives by means of the above-introduced approaches.

Questions on Context and Technology Usage

As can be seen from Table 2, positive and negative experiences differ in regard to the contextual factors *location* ($\chi^2 = 17.865$; $p < .05$) and *presence of other people* ($\chi^2 = 18.410$; $p < .05$). This means that negative experiences are more likely to occur at home (54.5 vs. 44.6%) and when being alone (51.7 vs. 39.9%), whereas positive experiences occur more frequently when other people are present (60.1 vs. 48.3%). Positive and negative experiences, however, did not differ on *how long ago* an experience occurred ($\chi^2 = 2.525$; $p = .64$).

Participants differ in how long they have been using a particular technology ($\chi^2 = 16.749$; $p < .05$). Positive experiences occur more frequently during the first 12 months of usage (49 vs. 35%), but when the technology is used for a longer period of time people tend to report more negative experiences (51 vs. 65%). After some time of use a products' ability to evoke positive experiences seems to fade. Moreover, there is a significant difference on the intensity of product use (i.e., how frequent has the product been used), $\chi^2 = 17.596$; $p < .05$. Participants in the positive condition reported a less frequent use of the technology: 58% in the positive vs. 71% in the negative condition indicate to use the reported technology at least once a week, whereas 36% in positive vs. 24% in the negative condition indicate to use the technology several times a month or less frequently.

Ratings on Affect, Need Fulfillment and Product Quality

All ratings on affect, need fulfillment, and product qualities of the positive and negative experience condition were compared by means of Mann–Whitney *U* tests. As all comparisons yielded a significant difference – except for the item 'determined' of the PANAS – we calculated effect sizes ($r = Z/\sqrt{N}$; [6], p. 550]) as an estimate of the relative importance of the items in differentiating between positive and negative experiences. Table 3 shows the statistics.

Most of our results on affect are in line with previous findings [23]: the positive items "enthusiastic", "proud", "inspired", "exited", "interested" and the negative items "upset", " irritable", "hostile", "distressed" achieve the largest differences between positive and negative experiences (for all items effect size $r \geq .50$; large effects [6]). In terms of saliency (i.e., achieved rating score), "interested", "attentive", "enthusiastic", and "exited" were the most salient items within positive experiences, whereas "upset", "distressed", and "irritable" were the most salient for negative experiences.

For need fulfillment our results differ from those of [23]: in their study only the need "self-esteem" was significant (which we did not assess, see method section), whereas in our study all seven needs were rated significantly higher on fulfillment in the positive than in the negative experience condition (all with at least $r > .30$; moderate effect). The largest difference was observed with the need "stimulation" ($r = .59$). Within positive experiences "competence", "relatedness", and "autonomy" were the most salient needs and "meaning" the least salient need. This mostly corresponds with previous findings on need fulfillment [9, 23, 31].

In terms of product quality the most pronounced difference was with "goodness" ($r = .66$) followed by "pragmatic quality" ($r = .53$), "beauty" ($r = .43$) and "hedonic quality" ($r = .41$). Within positive experiences "goodness" was regarded by far as the most salient product quality.

Scales	Negative (N=323)		Positive (N=368)			
	M	SD	M	SD	Z	r^1
PANAS positive						
Enthusiastic	1.82	1.13	4.19	1.04	-19.04	**0.72**
Proud	1.76	1.10	3.90	1.11	-17.94	**0.68**
Inspired	1.79	1.10	3.64	1.24	-16.21	**0.62**
Excited	2.31	1.30	4.15	1.04	-16.05	**0.61**
Interested	3.32	1.40	4.57	0.74	-13.07	**0.50**
Strong	2.46	1.13	3.48	1.06	-11.22	**0.43**
Attentive	3.69	1.24	4.29	0.90	-6.68	**0.25**
Active	2.99	1.23	3.59	1.14	-6.57	**0.25**
Alert	3.62	1.13	3.87	1.23	-3.75	**0.14**
Determined	3.72	1.27	3.90	1.03	-1.11	0.04
Total	2.75	0.79	3.96	0.66	-17.11	**0.65**
PANAS negative						
Upset	4.45	0.93	1.83	1.29	-19.59	**0.75**
Irritable	4.23	1.09	1.74	1.18	-19.20	**0.73**
Hostile	3.01	1.40	1.29	0.79	-16.45	**0.63**
Distressed	4.28	1.05	2.25	1.44	-16.40	**0.62**
Guilty	2.41	1.45	1.45	0.97	-9.57	**0.36**
Ashamed	2.20	1.39	1.33	0.79	-9.54	**0.36**
Nervous	3.28	1.51	2.32	1.40	-8.09	**0.31**
Jittery	2.84	1.45	1.97	1.25	-7.99	**0.30**
Scared	2.60	1.45	1.77	1.18	-7.89	**0.30**
Afraid	2.61	1.49	1.75	1.19	-7.77	**0.30**
Total	3.19	0.90	1.77	0.87	-16.77	**0.64**
Product qualities						
Goodness	4.15	1.95	6.55	0.90	-17.38	**0.66**
Pragmatic qual.	4.21	1.46	5.73	0.89	-13.85	**0.53**
Beauty	4.16	1.32	5.36	1.18	-11.38	**0.43**
Hedonic qual.	4.48	1.17	5.45	0.93	-10.87	**0.41**
Need fulfillment[2]						
Stimulation	2.16	0.92	3.54	0.98	-10.89	**0.59**
Meaning	1.92	0.92	3.04	1.03	-9.14	**0.49**
Competence	2.68	1.10	3.82	0.91	-9.02	**0.49**
Relatedness	2.39	1.22	3.72	1.26	-8.79	**0.47**
Security	2.40	0.96	3.36	0.85	-8.55	**0.46**
Autonomy	2.83	1.08	3.69	0.94	-7.26	**0.39**
Popularity	2.41	1.18	3.21	1.06	-6.21	**0.33**

Note. Values in bold are significant at the 5% level (two-tailed test), [1] effect size r for Mann-Whitney U test, [2] negative: n = 145, positive: n = 199.

Table 3. Differences between positive and negative user experiences in regard to affect, need fulfillment and product quality.

Structure and Main Themes of UX Narratives

On average the reported experiences consisted of 85 words (SD = 57); negative narratives tend to be slightly longer (90 vs. 81 words). In general the participants described their experiences quite richly. Besides the main action they also provided substantial information on the context of their experience. Furthermore, most reports contain an evaluation part, where participants highlighted why their experience is noteworthy and how they felt about it.

The majority of the reported experiences (96%) follows a basic narrative structure; beginning with an *orientation*, followed by a *complicating action* and ending with a *resolution* (OCR). Most of the narratives also included an *evaluation* part (85%). Other structural elements were used less frequently: *abstract* (12%) *attempt to solve the complication* (18%), and *coda* (16%). There was no difference in structure between positive and negative narratives, except for coda. Around 21% of the positive compared to 11% of the negative narratives ended with it ($p < .05$). Segmenting the UX reports according to Labov's [18] narrative structure worked well. Thus, people reported their experiences with technology in a 'natural' way, as they would narrate a personal experience.

As only few reports contained an abstract or coda we only analyzed orientation, complication and resolution for content (see Table 4). The most dominating themes within the **orientation** are navigation (e.g., driving around an unfamiliar area), communication (e.g., using Skype to stay in touch), devices (e.g., using a smart phone), obligations (e.g., having to finish a college project), and social relations (e.g., grandson). In the positive experience condition, communication, social relations, navigation (being lost somewhere) were more dominating themes; within negative experiences obligations, breakdowns, phones, navigation (GPS failure, non-specific orientation), and digital housekeeping (trying to fix, update, upgrade things) were mentioned more frequently. Here an example of a positive experience with the main theme social relations: "*One of the best experiences I have had has been the ability to video chat with my family on the other side of the state.*"

In the **complication** – where people describe the main action of their story – wayfinding (being helped by the GPS), communicating (overcoming distances and being able to see family & friends using video chatting), and creating (taking pictures or recording a video) were frequent themes within positive narratives. In negative narratives participants were more likely to write about digital housekeeping, getting lost (having issues with the GPS: getting strange or wrong instructions, driving in circles and ending up in the wrong place), and losing data.

When it comes to the **resolution** (the final action, the outcome of the story), people in the negative condition often conclude that the result of their experience was that they felt bad (mostly frustration), had lost something (e.g., data), or experienced some kind of failure (a problem remains, a goal could not be achieved). In contrast, the outcome of positive narratives is mostly about being socially related (being together, seeing each other, taking part or sharing special moments), being able to do stuff more efficiently or in an easier way, successful wayfinding (getting there) and about being saved by technology (e.g., being able to call someone in the case of emergency). These themes relate quite well to the most salient needs (i.e., relatedness, competence and autonomy)

Linguistic Inquiry and Word Count (LIWC) Analysis

The LIWC analysis shows that positive and negative narratives differ in 'psychological processes', but also 'personal concerns' (see Table 5). Note that percentages reflect the proportions of words matching a specific category relative to the entire text of a narrative.

Participants tend to use more words related to 'affective processes' when reporting a positive compared to a negative experience (4.4 vs. 3.9%). Within positive narratives, positive emotions are more extensively expressed (3.6 vs. 1.6%), whereas in negative narratives participants apply more negative emotional expressions (0.8 vs. 2.3%). Especially, words associated with 'anger' have high discriminatory power between positive and negative narratives. Anger is expressed very rarely in positive (less than 0.1%), but relatively more frequently in negative narratives (0.6%). These results match the PANAS ratings and the occurrence of themes related to affect in the resolution section of the manual content analyses (see Table 3 and 4, respectively). Apparently, affect plays a prominent role in UX. Further, the LIWC shows that social processes play a more prominent role in positive narratives (9.0 vs. 5.9%): words associated with family, friends or humans in general, are used more frequently within positive than negative narratives. This highlights the social aspects of positive UX, which corresponds to the findings of [9]. Finally, in terms of "personal concerns" positive narratives are more frequently referring to "leisure" and "home", but negative ones more to "work" and "achievement".

Machine Learning Analysis

A five-fold cross validation suggested an accuracy of 86.5% for the classifier (precision: 86.2%, recall: 86.6%). Compared to other emotion and sentiment classifications [22], this is high. Consequently, we proceed to examine the features that are critical in distinguishing positive from negative narratives. We screened and grouped the 400 best

	Main Themes	Neg. N=145	Pos. N=199
Orientation (context)	*Obligations*: appointments; meeting; working; school	12.4%	8.9%
	Breakdowns: something is broken; slow; non-optimal; breakdowns related to software	12.4%	0.5%
	Devices: phones	6.6%	3.6%
	Navigation: non-specific orientation; GPS failure	6.6%	2.6%
	Digital Housekeeping: trying to fix, update or upgrade things	7.3%	1.6%
	Communication: Skype; video chat; stay connected; being apart	4.4%	21.4%
	Social Relations: family members; partners; friends	1.5%	10.4%
	Navigation: being lost somewhere	2.2%	4.2%
	Devices: all other devices	5.8%	5.2%
	Navigation: specific location; meet friends	5.1%	5.2%
	Creating: taking pictures; recording a video	4.4%	4.7%
	Travel: Travel / Holiday	2.9%	3.1%
	Diverse: All remaining themes	28.5%	28.6%
Complication (main action)	*Digital Housekeeping*: trying to fix, update or upgrade things	16.1%	1.3%
	Getting lost: getting strange or wrong instructions; driving in circles; ending up in the wrong place	14.9%	0.4%
	Losing data: pictures; contacts; music; general data	9.9%	1.7%
	Disconnecting: Disconnecting	6.8%	0.4%
	Wayfinding: saved by the GPS; finding a specific place	-	18.9%
	Communicating: overcoming distances; talking face to face; getting in touch	1.2%	14.2%
	Creating: taking pictures; recording videos	1.9%	7.7%
	Information: accessing data; looking up things	1.2%	7.3%
	Enabling: learning; checking-in online; new possibilities through technology; being more efficient; searching for a job	1.2%	6.0%
	Money: saving money; earning money	0.6%	5.6%
	Diverse: All remaining themes	46.0%	36.5%
Resolution (outcome)	*Feel Bad*: frustrated; embarrassed; foolish; upset; angry; disappointed	15.1%	-
	Failure: problem remains; user gives up; goal not achieved	12.2%	-
	Loss: data: media: contacts	12.2%	-
	Inefficient: detour; wasting time; time consuming	8.6%	-
	Money: bad deal; losing money	2.2%	1.6%
	Social Relatedness: being together; seeing each other; sharing moments; find a partner or friend; keep in touch; give pleasure to someone	0.7%	26.2%
	To Be Enabled: being more efficient; new possibilities through technology; customizing	-	12.0%
	Wayfinding: getting there	4.3%	7.9%
	Saved by technology: emergency	1.4%	8.4%
	Feel good: fun; joy; emotion regulation; amazed; entertainment	-	8.4%
	Success: data recovered; goal achieved; problem solved	3.2%	5.3%
	Money: good deal, earning money	1.4%	4.7%
	Diverse: All remaining themes	38.5%	25.6%

Table 4. Main themes within the different narrative segments.

discriminating words. Table 6 summarizes the most relevant words for each group. Interestingly, "to be able" had the highest discriminatory power for positive narratives. This is in line with the finding from the content analysis where "to be enabled" was a main theme in the resolution of positive experiences. For negative experiences the word pair "have to" was most important. This corresponds well with "obligations", which is the main negative theme within the orientation (see Table 4). In general, the identified words fit well with the findings of the other approaches.

	Negative (N = 323)		Positive (N = 368)		Z	r^1
	M	SD	M	SD		
Psychological Processes						
Social processes	5.88	5.46	9.03	6.07	7.3	**.28**
Family	0.32	0.94	1.05	1.86	6.2	**.24**
Friends	0.32	0.87	0.63	1.17	4.4	**.17**
Humans	0.18	0.57	0.36	1.02	2.2	**.08**
Affective processes	3.89	2.66	4.40	3.08	2.3	**.09**
Positive emotion	1.63	1.76	3.63	2.96	10.6	**.40**
Negative emotion	2.25	2.11	0.76	1.19	11.3	**.43**
Anxiety	0.31	0.87	0.16	0.52	2.7	**.10**
Anger	0.56	1.00	0.07	0.36	9.1	**.35**
Sadness	0.62	1.15	0.31	0.73	3.8	**.15**
Personal Concerns						
Work	3.57	3.69	2.30	2.97	5.3	**.20**
Achievement	2.76	2.60	2.21	2.02	2.5	**.10**
Leisure	1.32	2.48	2.38	3.00	5.9	**.22**
Home	0.49	1.08	0.80	1.33	3.9	**.15**
Money	0.61	1.23	0.91	1.79	1.6	.06
Religion	0.11	0.51	0.08	0.43	0.1	.00
Death	0.08	0.43	0.03	0.20	1.2	.04

Note. Values in bold are significant at the 5% level (two-tailed test), [1]effect size *r* Mann-Whitney *U* test.

Table 5. Mean percentage of word use in positive and negative narratives arranged by LIWC categories.

DISCUSSION

Next, we discuss the main contributions with two questions: (1) what can we learn from the content of UX narratives and (2) what is the value of the analysis approach.

What may we learn from the content of narratives?

Whereas some recent small-sample studies have analyzed the content of narratives (e.g., [21, 23]), earlier work has not succeeded in doing so (e.g., [9]). One key contribution of the present paper is a systematical overview of the content of a large sample of positive as well as negative narratives. We see three groups of relevant insights.

First, our findings highlight the importance of emotion in UX. This is seen in the large effect size in the PANAS ratings, but also in the content of the narratives. All three content analysis approaches show that emotions are a prominent theme and have high discriminatory power between negative and positive narratives. The importance of emotion within UX is emphasized in numerous studies (for a review, see [3]). In contrast to earlier work, our study offers some deeper insights on the occurrence of emotion by combining the structural features of narratives with its content. We found that emotions are frequently reported in the resolution section, which means that emotions are the outcome of the experience. This goes especially for negative emotions such as anger and frustration.

Second, social aspects, such as being with friends and family, communicating with people or sharing important life events, play a prominent role in positive user experiences. Themes like being with friends and family, communicating with people or sharing important life events were frequently found in the narratives. For instance, participants wrote stories of how they saw their granddaughter for the first time by using Skype or how they shared pictures of their restaurant opening with family members. The importance of social aspects can be seen in the LIWC analysis, which reveals that words belonging to the categories of family and friends are more often used when reporting about positive experiences compared to negative ones. The same result is provided by the content analysis where "social relatedness" was the main theme in the resolution part of positive narratives. In addition, one of the questions on context revealed that positive experiences occur more frequently when other people are present. This contradicts some earlier findings that admitted the social component of positive UX, but also found that only a minority of experiences refers explicitly to social aspects [9, 23]. For instance, Partala et al. [23] concluded from their content analysis that "social aspects were […] mostly missing from the qualitative descriptions" (p. 31).

Third, positive experiences were often grounded in enabling people to do certain things (i.e., the possibility to do something new or more efficiently). The maximum entropy classifier as well as the manual content analysis showed that the words "to be able", respectively "to be enabled" are related to positive experiences. In contrast, negative experiences were more about being unable to do something: "failure" (manual content analysis), "could not" or "to try" (maximum entropy classifier). Together with the finding that positive experiences are more likely to occur at early stages of product usage, this emphasizes the importance of a product to be novel and to offer new possibilities to the user in order to create a good user experience [23].

What is the value of our analysis approach?

The manual analysis of user-generated stories suggested that they follow a canonical narrative structure, similar to that found in narrations about personal experiences [7, 17, 18]. This finding has implications for the collection of UX narratives, but also for their analysis. Structuring a questionnaire or interview by means of Labov's structural elements allows collecting information about a specific UX event or episode in a more systematic and complete way compared to a single open-ended question. For instance, users may be supported in remembering details of their experiences with a wizard-like questionnaire that guides them step-by-step trough the elements of a narrative. Such an approach may also enhance users' motivation to report an experience, since answering a series of targeted

Category	Experience: Negative	Experience: Positive
Activity	to have to, to try, could not, to work	to be able, to find, to allow, to help
Appraisal	frustrating, wrong, bad, limited, accidentally, extremely, slow	best, great, free, little, easy, good, first, easy to, happy, love
Content	text, work, message, destination, data	directions, picture, information, way home, money, face, movie
Context	troubles, problem, work, school, tasks	home, trip, town, car, life
Social	sister, son, parents, boyfriend	we, family, girlfriend, wife, together
Technology	GPS, Windows, screen, computer, phone, setup	Skype, video, app, digital, iPhone, internet, online

Table 6. Words discriminating between positive and negative experiences identified by the maximum entropy classifier.

questions is less demanding then writing a coherent text on a personal experience. In the context of interviews, using the canonical structure of a narrative as a guideline might prevent the interviewer from missing certain elements of the reported experiences. Regarding the analysis of narratives, identifying relevant themes within the different structural elements could help to better understand the meaning of themes. As each structural element has a specific function within a narrative (i.e., providing the context, describing the main action or highlighting the outcome), we can use this information to better interpret the meaning and relevance of a theme for the entire user experience. For instance, we argue that a theme occurring in the resolution part is of high relevance for the entire experience as it reflects the main outcome of an experience. In contrast, themes in the orientation highlight the context of the experience, but do not say much about the quality of an experience. In addition, the evaluation element of a narrative (i.e., why the story is interesting or noteworthy) is likely to occur within the resolution.

Another key contribution of the present paper is to combine manual analysis (using linguistic narrative structure) with automated analysis (LIWC, classifier), since large-scale studies are demanding to analyze manually. With our multi-method approach we show that word-based automated methods may help to characterize the content of positive and negative UX stories by generating insights, which corroborates those from the manual approach. In contrast to a fully manual approach, a multi-method approach is more efficient as only a subset of data has to be coded manually, since the picture can be completed by inferring from the outputs of the automated approaches performed on the entire data set. Moreover, the automated methods we used are generally accessible at low costs and are fairly easy to use (the LIWC in particular). This enables researchers or practitioners to apply these powerful tools to get insights into all kinds of text-based experience reports (e.g., user-generate product reviews) at low costs.

Finally, our approach is not limited to content exploration of UX narratives but could also be of used in other HCI domains. For instance, it could be applied to compare user narratives on two different versions of a company's product (or product of a competitor). In addition, one could investigate how users' perception of a technology has changed over time by comparing the content of user-generated product reviews over the past years.

Relation to earlier studies of Affect and Need Fulfillment

In addition to describing narratives, our findings speak to earlier work on affect and need fulfillment [9, 23]. Our study reaffirms earlier findings by revealing that affect is a key aspect in discriminating between positive and negative experiences. Thereby positive experiences are primarily associated with enthusiasm and excitement, whereas being upset and irritation are most salient for negative experiences. This corroborates the reasoning of [9, 23] that positive experiences "are more often related to personally meaningful aspects of user experience, e.g., stimulation and identification, while [negative experiences] are more often accompanied by more direct emotional responses, typically to pragmatic problems" (p. 31).

The findings on need fulfillment deviate from earlier studies regarding differences between positive and negative experiences. As we found a large difference for all needs, [9, 23] we observed only a difference for "self esteem". In terms of saliency we identified competence, relatedness and autonomy, as been the most important needs for positive experiences. This is mostly in line with earlier work [23, 28, 31]. In contrast to those studies, which primarily relied on rating data, we show on a large scale that the importance of need fulfillment within UX is also reflected in the content of user narratives. The themes in the resolution part of the narratives (e.g., social relatedness, to be enabled, wayfinding), as well as the results of the maximum entropy classifier (e.g., we, family, together, to be able, to find) and the LIWC (social processes), all relate to the most salient needs (i.e., relatedness, autonomy, competence).

Our study also illustrates what experiences related to need fulfillment and emotions look like. They are based on authentic experiences and are meaningful to users since they have been freely remembered. Such UX themes may help designers since they can serve as inspiration when designing for need fulfillment and emotion.

Limitations

Our research is limited in several ways. First, by using Amazon Mechanical Turk for recruiting and asking participants to remember a specific experience, we do not know if narratives are representative for general technology use. Moreover, our analysis is text-based and text is not a direct representation of reality. Hence, it is unclear as to how our analysis reflects actual experiences of users.

Second, collecting narratives by means of an online questionnaire might have influenced the way people narrate their experiences. Moreover, we were unable to probe more into interesting aspects of single experiences or to clarify ambiguous issues in a narrative. Interviews, in contrast, might have generated richer data on the single experiences. Nevertheless, the narratives had enough details and coherence to allow extraction of meaningful UX themes.

Finally, our approach provides an overview of emerging themes rather then an in-depth and rich understanding of the content of narratives. However, depending on the intentions of an investigator both can be useful.

CONCLUSION

User-generated descriptions of experiences provide valuable input to UX researchers who try to describe and model experiences. Whereas such descriptions are often tied to ratings of need fulfillment and affect, the actual descriptions are rarely analyzed. We studies 691 narratives of experiences with interactive products, analyzing their contents manually and automatically. The analysis illustrates the complexity of trying to understand experience through narratives and helps gain new insights about what distinguishes positive and negative experiences.

REFERENCES

1. Aronson, J. A pragmatic view of thematic analysis. *The Qualitative Report*, 2, 1 (1994), 1-3
2. Arrasvuori, J., Korhonen, H. and Väänänen-Vainio-Mattila, K. Exploring playfulness in user experience of personal mobile products. In *Proc. OZCHI 2010*. ACM (2010), 88-95.
3. Bargas-Avila, J. and Hornbæk, K. Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In *Proc. CHI 2011*. ACM Press (2011), 2689-2698.
4. Beyer, H. and Holtzblatt, K. Contextual design. *interactions*, 6, 1 (1999), 32-42.
5. Buhrmester, M., Kwang, T. and Gosling, S. D. Amazon's Mechanical Turk A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, 6, 1 (2011), 3-5.
6. Field, A. P. *Discovering statistics using SPSS*. SAGE publications Ltd, 2009.
7. Habermas, T. and Berger, N. Retelling everyday emotional events: Condensation, distancing, and closure. *Cognition and Emotion*, 25, 2 (2011), 206-219.
8. Hassenzahl, M. The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction*, 19, 4 (2004), 319-349.
9. Hassenzahl, M., Diefenbach, S. and Göritz, A. Needs, affect, and interactive products - Facets of user experience. *Interacting with Computers*, 22, 5 (2010), 353-362.
10. Hassenzahl, M. and Monk, A. The inference of perceived usability from beauty. *Human–Computer Interaction*, 25, 3 (2010), 235-260.
11. Hassenzahl, M. and Tractinsky, N. User experience - a research agenda. *Behaviour & Information Technology*, 25 2(2006), 91-99.
12. Karapanos, E., Zimmerman, J., Forlizzi, J. and Martens, J.-B. User experience over time: an initial framework. In *Proc. CHI 2009*. ACM (2009), 729-738.
13. Kittur, A., Chi, E. H. and Suh, B. Crowdsourcing user studies with Mechanical Turk. In *Proc. CHI 2008*. ACM (2008), 453-456.
14. Klein, D. and Manning, C. D. Optimization, Maxent Models, and Conditional Estimation without Magic. *Tutorial delivered at HLT-NAACL 2003*(2003).
15. Korhonen, H., Arrasvuori, J. and Väänänen-Vainio-Mattila, K. Let Users Tell the Story: Evaluating User Experience with Experience Reports. In *Proc. CHI 2010 Ext. Abs.* (2010), 4051-4056.
16. Kramer, A. D. I. and Rodden, K. Word usage and posting behaviors: modeling blogs with unobtrusive data collection methods. In *Proc. CHI 2008*. ACM (2008), 1125-1128.
17. Labov, W. The transformation of experience in narrative syntax. In, *Language in the Inner City: Studies in the Black Venacular*. University of Pennsylvania Press, Philadelphia, PA, 1972.
18. Labov, W. and Waletzky, J. Narrative analysis: Oral versions of personal experience. In J. Helm (ed.), *Essays on the Verbal and Visual Arts,* 12-44. Seattle, 1967.
19. McCarthy, J. and Wright, P. *Technology as Experience*. MIT Press, Cambridge, MA, 2004.
20. Nigam, K., Lafferty, J. and McCallum, A. Using maximum entropy for text classification. In *Proc. IJCAI 1999* (1999), 61-67.
21. Olsson, T. and Salo, M. Narratives of satisfying and unsatisying experiences of current mobile augmented reality applications. In *Proc. CHI 2012* (2012), 2779-2788.
22. Pang, B., Lee, L. and Vaithyanathan, S. Thumbs up?: sentiment classification using machine learning techniques. In *Proc. EMNLP 2002*. ACL (2002), 79-86.
23. Partala, T. and Kallinen, A. Understanding the most satisfying and unsatisfying user experiences: Emotions, psychological needs, and context. *Interacting with Computers*, 24(2012), 25-34.
24. Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A. and Booth, R. J. *The development and psychometric properties of LIWC2007*. www.liwc.net, 2007.
25. Pennebaker, J. W., Francis, M. E. and Booth, R. J. *Linguistic inquiry and word count: LIWC 2001*. Erlbaum Publishers, Mahwah, NJ, 2001.
26. Riessman, C. K. *Narrative methods for the human sciences*. SAGE, Los Angeles, 2008.
27. Rime, B., Mesquita, B., Philippot, P. and Boca, S. Beyond the Emotional Event - 6 Studies on the Social Sharing of Emotion. *Cognition Emotion*, 5, 5-6 (1991), 435-465.
28. Ryan, R. M. and Deci, E. L. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 1 (2000), 68-78.
29. Scherer, K. R. Appraisal considered as a process of multilevel sequential checking. *Appraisal processes in emotion: Theory, methods, research*, 92(2001), 120.
30. Scissors, L. E., Gill, A. J., Geraghty, K. and Gergle, D. In CMC we trust: the role of similarity. In *Proc. CHI 2009*. ACM (2009), 527-536.
31. Sheldon, K. M., Elliot, A. J., Kim, Y. and Kasser, T. What is satisfying about satisfying events? Testing 10 candidate psychological needs. *Journal of Personality and Social Psychology*, 80, 2 (2001), 325-339.
32. Tausczik, Y. R. and Pennebaker, J. W. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29, 1 (2010), 24-54.
33. Watson, D., Clark, L. A. and Tellegen, A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54, 6 (1988), 1063-1070.