Affordance-based Use Case Mining

Completed Research

Mostafa Mesgari
Love School of Business
Elon University
mmesgari@elon.edu

Abstract

The dominant function-based design perspective has been criticized for focusing more on the system and its functionality but less on the user and its interactivity. In line with the efforts to go beyond the function-based design and towards more relational design perspective, this study proposes to create use case models at the level of affordances rather than functions, so use cases focus on user interactivity of the system rather than its functions. Useful for redesigning and enhancing legacy systems, this study suggests a technique to mine the wealth of user log data to create use cases based on the sequence of affordance actualizations. It illustrates affordance-based use case mining by analyzing use cases of a real-world legacy system. The study contributes to the design research and practice by incorporating an affordance perspective into the popular use case modeling technique, and by proposing and demonstrating a technique to mine affordance-based use cases.

Keywords

Human computer interaction, use case design, affordance, data mining, sequence analysis.

Introduction

Redesigning and enhancing the legacy systems comprises major part of information technology investments. In fact, 67% of the software life cycle management cost involves software maintenance, including new functionality developments (Schach 1999). Glass (2001) proposes his rule of "60/60" that breaks down the costs of software development; it suggests about 40% to 80% (average of 60%) of the software development expenses involves the cost of software maintenance, and 60% of the software maintenance costs involves upgrading and adding new features to the software compared to just 17% for fixing the errors. This highlights the crucial role of the legacy system enhancement activities.

To redesign and enhance Information Systems (IS), the system development team needs to understand the target users and how they use and interact with the system to achieve their goal. For that purpose, they draw on a rich toolbox of design techniques. Use case modeling is proved to be one of the most popular tools in which actors and use cases are identified and their interactions are depicted as how different functions of the system respond to various needs of the actors (Jacobson 1996).

The dominant approach to create use cases is to follow the requirement analysis techniques from stakeholder identification to modeling the relationship between stakeholder and components of the system (Jacobson 2000). In the case of redesigning legacy systems, scholars suggest supplementing the requirement analysis with a data mining approach to mine the wealth of user log data for use cases (El-Ramly et al. 2002). While this may not replace the requirement analysis techniques, it surely provides useful insight to build robust use cases.

The current design theory and practice is mostly shaped around and dominated by the core concept of "function" (Ciavola and Gershenson 2016). However, there is growing consciousness that there is a number of limitations with the approach. First, the function-based approach limits the explanatory power of the design theories, meaning it is difficult to explain how the functions isolated from users play role in the consequences of the information systems. Second, the function-based design prioritizes system functionality and utility over its interactivity, so it leads to designer-led systems rather than user-led ones (Maier and Fadel 2009a).

To address the shortcomings of the function-based design, there are recent efforts to go beyond that and towards more relational understanding of system design, so our theory and tools are more susceptible to both users and the artifact (Maier and Fadel 2009a, 2009b). This study joins forces with the rest of the affordance-based design stream of research to incorporate affordance perspective into the use case model, one of the most popular design tools. Affordances are the action possibilities provided by the system to its users, so it is relational to both users and the system (Gibson 1986). The concept bridges the user and the artifact and makes it possible to focus the design efforts on the interaction between the two rather than the functions of the system.

This paper proposes to create use cases at the level of system affordances, rather than its functions. It provides more user-oriented understanding of the system and enhances design for user interactivity. Targeted at redesigning and improving legacy systems, this study lays out a technique to mine user log data to construct affordance-based use cases. Then it illustrates affordance-based use case mining by analyzing the case of Moodle usage in a North American business school.

The contribution of this study is twofold. First, it proposes to go beyond current function-based perspective to use case construction. It highlights the limitation of the function-based approach and the advantages of an affordance-based approach. Second, it suggests a technique to mine affordance-based use cases for legacy systems enhancements by analyzing affordance actualizations through user log data.

The paper is structured as per the following. First, it will discuss modeling user behavior with use cases. Then it lays the ground for affordance-based use case mining and proposes a technique for conducting data mining to identify patterns of affordance actualization sequences that could translate into use cases later. Next, the paper illustrates the affordance-based use case mining by analyzing a case of students using Moodle, the learning management system. It then concludes with highlighting the contribution and some limitations of the study.

Modeling user behavior with use cases

Understanding user needs, behavior, and interaction with the system and putting it at the center of all system design procedures has been the focus of User-Centered Design (UCD) for decades. Use case model is among the most popular techniques for understanding user needs and their interaction with the system (Bilow 1995). It is composed of two elements of actors and use cases. The actors represent various roles or type of users that would interact with the system. The use cases would be defined at various level; they may represent higher level goals of the actors that system responds to, or lower level needs of the actors, or even more specific functions of the system that responds to the interaction initiated by the actor. The use cases depict the core requirements of the system essential to respond to user needs and interactions. They define what the system should be able to do to be supportive of the target user group. They define the designers' mind and perspective of the system and guide them throughout the system design and development process.

System designers mostly use the conventional requirement analysis techniques to construct use case models, meaning they start with identifying stakeholders, then interviewing them and collecting various data about the business processes around them and then analyzing and translating them into use case models of actors and their interactions with various functions of the system. More recently, some suggest that the designers can use the wealth of user-system interaction information for constructing use case models for legacy systems already in place.

El-Ramly et al. (2002) use data mining techniques to explore patterns in user activities on the system to identify common use cases that could be used to redesign a legacy system and improve the user experience. They used sequence analysis techniques to identify the most popular sequences of pages (i.e. computer screens) that users followed through while using the system. However, analyzing sequence of actions (clicks on computer screen) may not provide meaningful use cases, because it neglects user intentions and why they are taking those actions. User behavior would not be well understood and modeled unless the actions are analyzed within their context and in light of the purpose they fulfill. Next, we outline an affordance-based use case mining that analyzes user actions in the context of the affordances they actualize.

Affordance-based use case mining

The dominant system design practices have long emerged around the paradigm of function-based design thinking. It defines the system as a combination of functions needed to be fulfilled for the system to reach its goals. While the function-based design has been highly fruitful for the design practitioners in so many ways, it is proved to be limited in the scope of design problems that it can address and the explanatory power of the design theories (Maier and Fadel 2009a). Moreover, the function-based design defines system mostly from designers' perspective and some time at the cost of diminishing user role, so it may prioritize system functionality over user interactivity.

To address the limitations of function-based design, a recent stream of design research is pushing for an affordance-based design paradigm that defines the system in terms of its affordances rather than its functions (Maier and Fadel 2009a, 2009b). It is incorporating the affordance perspective into design methods and procedures (Ciavola and Gershenson 2016). Affordance theory originates from ecological psychology, a branch of psychology that expands traditional notions of cognitive psychology to explain human perception in a way consistent with Darwin's evolutionary theory—perception as a process of mutual adjustment with the environment (Mesgari 2016). From the ecological point of view, perception is linked to action, and the intertwinement of the two facilitates the selection and adaptation processes of species in the environment. To explain this intertwinement and provide the link between perception and action, James Gibson (1977) coined the notion of an "affordance" to refer to the action-related perception of the environment, and to establish the foundation for the theory of affordances.

From an ecological psychology perspective, affordances are the action possibilities furnished by the environment; they are real relationships existing between the individual and the environment, rather than subjective meaning made cognitively in the mind (Gibson 1986).

For technology design scientists, affordance approach is the "psychology of materials"; it links the material design of technology to the consequences it brings about. Norman (1988, p. 9) believes that an affordance approach to design of everyday things can help designers to better understand human behavior, and to design things in a readily understandable manner. He uses the approach to bridge the material aspect of design to human perception of the designed material. Moreover, since the affordance concept refers both to attributes of the object and the actor, it would be a "powerful concept for thinking about technologies because it focuses on the interaction between technologies and the people who will use them" (Gaver 1991, pp. 79–80).

This study proposes to create use cases at the level of affordances users actualize. It comprises identifying major patterns in the sequence by which affordances are actualized by users. This paper uses a mixed-methods approach to analyze user interaction with the system at the level of affordances they actualize, and uses sequence analysis to explore the patterns in sequence of affordance actualization in user data. Our method starts with extracting affordances using in-depth interview with users, then continues with categorizing user action into the identified affordances, and lastly finishes with sequence analysis techniques to mine patterns in sequence by which affordances are actualized by users. While we do not construct the use cases, we discuss how the affordance-based sequences provide insights about creating affordance-based use cases.

The methodology to mine affordance-based use cases is as per the following steps:

1. Extracting affordances: interviews with experienced users to be conducted and analyzed using qualitative coding to come up with major affordances provided to users.
2. Identifying the user activities actualizing each affordance: card sorting to be conducted with experienced users as judges categorizing possible user actions into affordances extracted in the first step.
3. Analyzing the pattern in sequence of affordance actualizations: user action logs should be related to the affordance they actualize and then sequence analysis should be used to explore the pattern in the sequence of affordance actualizations by users.

To illustrate mining affordance-based use cases, next section reports on the analysis of use cases discovered by mining user interactions with Moodle, the learning management system.

Illustration: Moodle use cases

This section reports on the analysis of sequence of affordance actualizations in the case of students using Moodle in a North American business school. Moodle is the open-source learning management system developed and being maintained by Moodle HQ and large community of volunteer contributors. As of June 2017[1], more than 82,000 Moodle websites in 236 countries provide over 12,000 courses to more than 105,000 registered users. Moodle provides excellent setting for mining affordance-based use cases in multiple ways. First, it provides users with various affordances and some degree of freedom to choose and act on the affordances based on their learning style and preferences. Second, the detailed user log maintained by Moodle lends itself to data mining techniques to explore patterns in user transactions with the system.

This empirical illustration follows the three steps discussed in the above section. The first two steps are conducted in common with a distinct but somewhat related study (Mesgari et al. 2018), so it is described in more detail there; what follows is an abbreviated description of the procedure we followed and the results. The third step is unique to this study and extends the technique to get insights for building affordance-based use cases by looking into the sequence by which users actualize affordances.

For the first step, semi-structured interviews are conducted with 17 students who used Moodle for multiple courses and had extensive experience and understanding of how it can be used. The interviews took between 30 to 60 minutes and recorded and transcribed before analysis. They were then openly coded for action possibilities provided by the system and the purposes served that could guide us to the affordances perceived by users. Axial coding was used to group the open codes into higher level codes that explain user experience in terms of the affordances they perceived. The qualitative coding process resulted in five major affordances of Moodle:

- **Content Access:** Action possibilities enabling the students to access any course content that they need; these possibilities give the students read-only access to the course-related material.
- **Submission:** Action possibilities enabling the students to submit their work, answers, or ideas for part of their course grade, for which they might or might not receive subsequent feedback.
- **Communication:** Action possibilities enabling the students to communicate and share their ideas, opinion, and questions with the teacher, teaching assistants or fellow classmates; or to acquire awareness of what the teacher, teaching assistants or classmates communicated or shared; both parties have the chance to express themselves and engage in two-way interaction.
- **Practice:** Action possibilities enabling the students to practice what they have already learned about the course material.
- **Feedback:** Action possibilities enabling the students to get feedback on their learning, participation, submitted work or status or progress in the course.

For the second step, two rounds of card sorting exercise are conducted with fifteen different student judges for both rounds. In each round, experienced students are given a list of 53 specific actions extracted from Moodle log data and asked to sort them into the five extracted affordances. The first-round results have inter-rater reliability (Fleiss Kappa) of 0.78. The second round of card sorting is conducted with modified action definitions and new set of judges; it results in Kappa of 0.90 while each affordance acquires an agreement of over 0.83, that demonstrates substantial agreement among the student judges. It means that we can be highly confident of the actions categorized as actualizing each of the five affordances.

For the third round, Moodle users' log data are collected for 461 students of four sections of the same course taught by single instructor. The course is intentionally chosen because of its extensive use of various Moodle features including providing material, assignments, quizzes, grades, forums and others. The dataset includes over 346,000 records of user actions taken during a full semester.

Data analysis is conducted using R version 3.4.0 including its base package (R. Core Team 2015), cluster package (Maechler et al. 2015), reshape2 package (Wickham 2007), hyperSpec package (Beleites and Sergo 2015), and most important of all the three packages of TraMineR (Gabadinho et al. 2011) and

[1] https://moodle.net/stats/

TraMineRextras (Ritschard et al. 2013), Gmisc (Gordon 2016) that are used to run sequence analysis and visualize the results.

Extensive data wrangling is used to pre-process the dataset and create the state sequence matrix that could be used in sequence analysis. First, the user log data is cleaned to include time-stamped action data for all users. Then each action is identified with the affordance it actualizes. To come up with sequences of actions related to separate events, 30 minutes time difference in the time-stamp is used to separate consecutive events of the same user. This is because the system is set to log out after 30 minutes of inactivity, so the two actions apart for more than 30 minutes would be considered as two separate events. This process resulted in 37,086 separate sequences of affordance actualization with length range of one to 280. Figure 1 represents the percentage of events with various sequence lengths among all events in the dataset. 15% of the events include a single action, so they are removed from the analysis. Events with sequence length of seven and more each include less than 3% of the events, so they are removed from further analysis due to non-significant population size. The final analysis includes 21,339 events with length range of two to six activities.

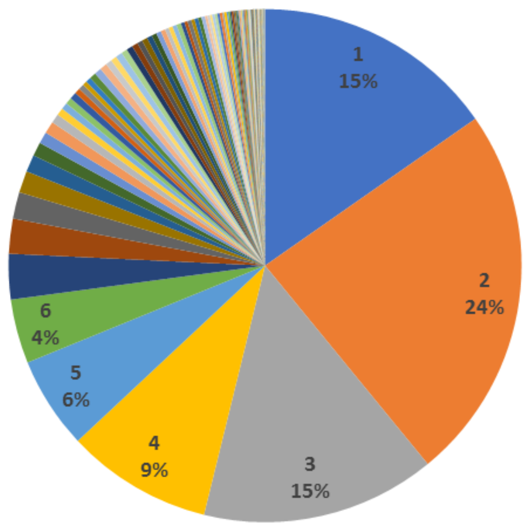

Figure 1. Frequency (%) of events with specific sequence length

Table 1 and Figure 2 represent the state frequency statistics that demonstrate the frequency (percentage) by which each affordance occurs at each place within the sequence. It provides useful insight about how the frequency of affordances change along the sequence of user activities. For instance, the first event in the sequence involves Content Access for 98% of the time, and Practice for 1% of the time. It means, user activity almost always starts with accessing content, and as users go further in each event, they actualize less of content access and more of feedback, practice, and communication.

Table 1. State frequency table*

Affordance	1	2	3	4	5	6	7
Communication	0.00	0.04	0.08	0.07	0.07	0.08	0.05
Content Access	0.98	0.82	0.62	0.66	0.65	0.59	0.53
Feedback	0.00	0.12	0.21	0.15	0.13	0.13	0.08
Practice	0.01	0.02	0.09	0.12	0.14	0.20	0.34
Submission	0.00	0.00	0.00	0.00	0.00	0.00	0.00

* State frequency numbers refer to the frequency (percentage) by which each affordance occurs at each place within the sequence.

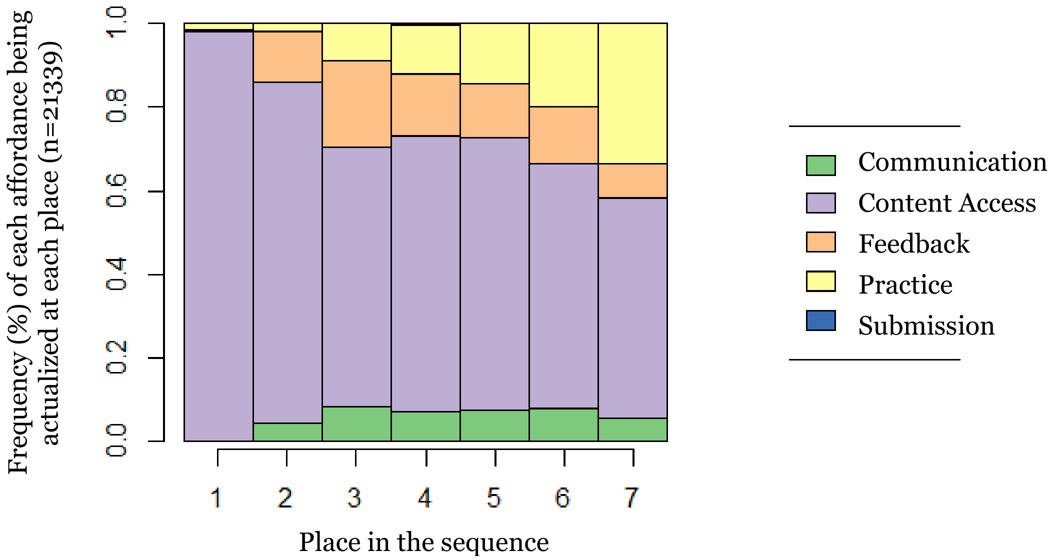

Figure 2. State frequency plot

Figure 3 represents the cumulative frequencies of the most frequent sequences of affordance actualization. The ten most frequent sequences represent over 70% of all the events. The first two most frequent sequences include consecutive actualizations of content access affordance. The next two most frequent sequences include one/two actualization(s) of content access affordance followed by actualization of feedback affordance. Another frequent sequence is when actualization of content access affordance is followed by actualization of communication affordance. The most frequent sequences provide insights that can directly guide the design of proper use cases.

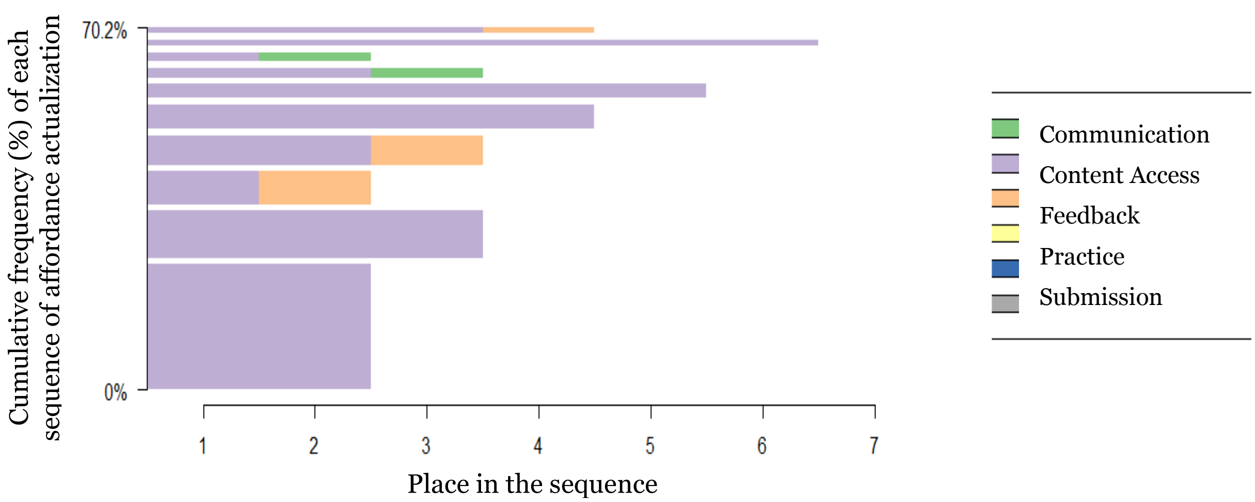

Figure 3. Cumulative frequencies of the most frequent sequences

Table 2 and Figure 4 represent the transition rates between the five states/affordances. They help understanding the most likely affordances that may be actualized after each specific affordance actualization. For instance, if any user is in the state of actualizing communication affordance, it is most probably going to transition to accessing content next, and somewhat likely to continue in communication or even less likely to get some feedback next. This analysis constructs the state transition diagram which normally comes with use case diagrams to better explain the system design and how it serves users to achieve their goal by going from one state to another.

Table 2. Transition rate between states

	[-> CMU]	[-> CNT]	[-> FDB]	[-> PRC]	[-> SBM]
[CMU ->]	0.33	0.55	0.12	0.00	0.00
[CNT ->]	0.06	0.76	0.16	0.02	0.00
[FDB ->]	0.01	0.59	0.35	0.04	0.00
[PRC ->]	0.00	0.08	0.02	0.90	0.00
[SBM ->]	0.00	0.25	0.06	0.00	0.69

CMU: Communication affordance
CNT: Content access affordance
FDB: Feedback affordance
PRC: Practice affordance
SBM: Submission affordance

Figure 4. State transition plot

Conclusion

This paper proposes to create use cases at the level of affordances rather than system functions. It is in line with the efforts to go beyond function-based design and towards more relational understanding of system design that fosters user interactivity. Moreover, to redesign and enhance the legacy systems, it suggests a technique to mine user log data to get useful insights on use case models based on the sequence of affordance actualizations by users. It illustrates the affordance-based use case mining by analyzing the case of Moodle use at a North American business school.

This study has multiple implications for design research and practice. For design researchers, it pushes for an affordance-based design and proposes to incorporate an affordance perspective into existing design tools like use case modeling. For practitioners, this study outlines a technique to mine wealth of user log data of legacy systems and explore useful insights for constructing affordance-based use case models.

The approach to discover use cases from user-interaction logs in general, and the affordance-based mining of use cases included, are limited in a number of ways. First, it does not apply to design of new systems for which there is no user log records available yet. However, analysis of the user log records for similar legacy systems may provide some insights to useful use cases. For instance, the insights from analysis of use cases for Moodle would be useful for designing a new learning management system. Second, the use cases derived from user-interaction log data is mostly representative of the existing system design and not the best fit for new feature introduction; so, it should be used and interpreted in the context of the current design specification and other data available. Third, the use cases created using this approach is limited to the user community involved with the system and somewhat representative of their goals and preferences. So, diversifying the user community in use case mining inquiries would be helpful in creating use cases of more generalizable nature.

References

Beleites, C., and Sergo, V. 2015. *HyperSpec: A Package to Handle Hyperspectral Data Sets in R*. (http://hyperspec.r-forge.r-project.org).

Bilow, S. C. 1995. "Defining and Developing User Interface Intensive Applications with Use Cases," *Report on Object Analysis and Design* (1:5), pp. 28–34.

Ciavola, B. T., and Gershenson, J. K. 2016. "Affordance Theory for Engineering Design," *Research in Engineering Design* (27:3), pp. 251–263. (https://doi.org/10.1007/s00163-016-0216-5).

El-Ramly, M., Stroulia, E., and Sorenson, P. 2002. "Mining System-User Interaction Traces for Use Case Models," in *Proceedings 10th International Workshop on Program Comprehension*, pp. 21–29. (https://doi.org/10.1109/WPC.2002.1021305).

Gabadinho, A., Ritschard, G., Mueller, N. S., and Studer, M. 2011. "Analyzing and Visualizing State Sequences in R with TraMineR," *Journal of Statistical Software* (40:4), pp. 1–37.

Gaver, W. W. 1991. *Technology Affordances*, pp. 79–84.

Gibson, J. J. 1977. "The Theory of Affordances," in *Perceiving, acting, and knowing: toward an ecological psychology*, J. B. Robert E Shaw (ed.), Hillsdale, N.J. : Lawrence Erlbaum Associates, Pp.67-82. (http://hal.archives-ouvertes.fr/hal-00692033).

Gibson, J. J. 1986. *The Ecological Approach to Visual Perception*, Lawrence Erlbaum.

Glass, R. L. 2001. "Frequently Forgotten Fundamental Facts about Software Engineering," *IEEE Software; Los Alamitos* (18:3), 112,110–111. (http://dx.doi.org.ezproxy.elon.edu/10.1109/MS.2001.922739).

Gordon, M. 2016. *Gmisc: Descriptive Statistics, Transition Plots, and More*. (https://cran.r-project.org/web/packages/Gmisc/index.html).

Jacobson, I. 1996. "Formalizing Use-Case Modeling," *Wisdom of the Gurus: A Vision for Object Technology* (8), p. 139.

Jacobson, I. 2000. "Basic Use-Case Modeling," *The Road to the Unified Software Development Process* (18), pp. 167–172.

Maechler, M., original), P. R. (Fortran, original), A. S. (S, original), M. H. (S, maintenance(1999-2000)), K. H. (port to R., Studer, M., and Roudier, P. 2015. *Cluster: "Finding Groups in Data": Cluster Analysis Extended Rousseeuw et Al.* (https://cran.r-project.org/web/packages/cluster/index.html).

Maier, J. R. A., and Fadel, G. M. 2009a. "Affordance Based Design: A Relational Theory for Design," *Research in Engineering Design* (20:1), pp. 13–27. (https://doi.org/10.1007/s00163-008-0060-3).

Maier, J. R. A., and Fadel, G. M. 2009b. "Affordance-Based Design Methods for Innovative Design, Redesign and Reverse Engineering," *Research in Engineering Design* (20:4), pp. 225–239. (https://doi.org/10.1007/s00163-009-0064-7).

Mesgari, M. 2016. "Essays on an Ecological Approach to User-Technology Interaction," PhD Thesis, PhD Thesis, Concordia University.

Mesgari, M., Okoli, C., and Guinea, A. O. de. 2018. "Creating Rich and Representative Personas by Discovering Affordances," *IEEE Transactions on Software Engineering*, pp. 1–1. (https://doi.org/10.1109/TSE.2018.2826537).

Norman, D. A. 1988. *The Psychology of Everyday Things*, Basic Books (AZ).

R. Core Team. 2015. *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria*. (https://www.r-project.org/).

Ritschard, G., Studer, M., Buergin, R., Gabadinho, A., Muller, N., and Rousset, P. 2013. *TraMineRextras: Extras for Use with the TraMineR Packages*, Genève: CRAN. (http://scholar.google.com/scholar?cluster=9526901776907303478&hl=en&oi=scholarr).

Schach, R. 1999. *Software Engineering*, (Fourth.), Boston, MA: McGraw-Hill. (https://courses.cs.vt.edu/csonline/SE/Lessons/LifeCycle/Lesson.html).

Wickham, H. 2007. "Reshaping Data with the Reshape Package | Wickham | Journal of Statistical Software," *Journal of Statistical Software* (21:12), pp. 1–20.