Mental models, performance and usability of a complex interactive system: The case of Twitter

Rodrigo Juárez
Department of Computer Science
Instituto Tecnológico Autónomo de México
Mexico City, Mexico
miguel.juarez@itam.mx

Victor M. González
Department of Computer Science
Instituto Tecnológico Autónomo de México
Mexico City, Mexico
victor.gonzalez@itam.mx

Abstract—This work presents the study of a complex interactive system (Twitter) and the analysis of the relation between people’s mental models, performance and usability perceptions. Participants were asked to perform a number of activities with Twitter and to draw graphical representations of their mental model about it. We identified three typical Mental Model Styles (MMS) used to describe Twitter. With these styles in mind, we conducted two analyses, one based only on the MMS and the other one based on the MMS and the expertise. We found that the level of expertise had a major impact on performance rather than the mental model style defining the understanding about the system. Furthermore, we found that usability perception was affected by both, the level of expertise and the mental model style.

Keywords—complex interactive system; Twitter; usability; mental models.

I. INTRODUCTION

In recent years a massive growth has occurred in the design, production and use of interactive technologies. This boom of technology produces the necessity of bearing in mind certain aspects that were not so important before to both, users and programmers. One of these aspects is the way in which people understand, interact, and adapt the system, i.e. the Human-Computer Interaction (HCI).

At first, the majority of final users were programmers themselves or people with a high technical knowledge. As years went by, the “final user” concept suffered a diversification process and today, the common usage scenario implies a person with few or null technical knowledge about the system he or she uses.

This diversification on the user base has profound implications for designers and programmers, because they have to consider how the system is understood, conceptualized and internalized in the mind of the users in order for it to be valued, used efficiently and achieving the purposes for which it was designed. Among the challenges, one should understand and consider that an interactive design has to depart from allowing users easily define a mental representation to communicate with programmers and designers’ aims and ideas and vice versa (the Gulf of Execution and the Gulf of Evaluation suggested by Donald Norman[1]).

Cognitive sciences had shown that it is quite difficult to understand the world in a direct way, instead we use a representation of the world, elaborated in our minds, to act like a medium [2]. The most used representations among researchers are mental models [3] and those are the lenses that we used to frame our research.

A. Mental Models (MM)

Johnson-Laird [4] defines mental models as a series of psychological representations of real, hypothetic or imaginary situations whose structure corresponds to the structure of the event they represent and that form a group of knowledge blocks that can be manipulated according to the needs of a person. They represent a collection of knowledge that helps people to build their understanding of the world and to solve the problems that emerge [5].

The association between MM and learning came along the creation of David Ausubel’s Meaningful Learning Theory [6] that can be defined as learning with understanding (when people are able to use the information they have acquired to do certain jobs). In order to achieve meaningful learning, the apprentice must generate a MM from a combination of previous and new knowledge that is tested through problems, and that is redefined by the experience and the appropriate feedback [6].

Complex systems are hard to understand because they are organized in a multilevel way that depends on local relations that, most of the time, are not obvious. Another reason that explains the difficulty of these systems is that, due to the huge amount of events and relations that have to be processed simultaneously, a huge load of data is stored in the memory [7]. To face these troubles, MM are useful resources because they can be seen as a small and independent piece of information, which resulted of a fragmentation process of the larger system [8].

The International Standard Organization (ISO) and the International Electronic Commission (IEC) [9], in their standard ISO/IEC 9241-11:1998, define usability as the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use. There is one problem with this definition, it gives performance criteria prominence and there
are three other aspects, significant to usability that cannot be measured easily: the process of action, the consequence of the process and the emotional attachment [10].

MM are important to usability because they provide an insight of those aspects, letting the designers, programmers, researchers, etc. understand better what is happening inside the users’ mind and, therefore, explain why users act the way they do, how they feel while using the system and how they understand it.

B. Online Social Networking Services (OSNS)

An OSNS can be defined as a platform that provides a private online space (user’s profile) along the tools that are needed to interact with other people on the Internet allowing them to find common interests and exchange experiences and multimedia resources [11].

One of the most used OSNS is Twitter¹. Twitter has established itself as the second largest OSNS in the world with over 500 million users². Due to the dynamic nature of Twitter – the information’s flow, its interactive and communication model, the new functions³, the constant GUI changes and that people tend to explain it according the way they use it⁴ – its complexity increases and optimal learning is more difficult to reach. We assume this compels the user to update his MM frequently in order to comprehend the way it works, though some users tend to stall this effort and new users are not used to it, this causes a gap of knowledge between users with different levels of expertise.

Those factors have resulted in Twitter operating under an interactive and communication model which has not easily matched with other OSNS technologies and which demands to understand operates which are not modeled by other forms of communication (e.g., e-mail, phone, snail mail).

The aim of this research is to identify differences among the MM that users and non-users use to understand Twitter and their impact in the participants’ performance (specifically success rate and completing time) and satisfaction with the system.

Our current work is based on previous analyses we conducted for the same research [12]. Whilst [12] focuses in the level of expertise as a factor of change in the performance and how people sees Twitter’s usability this paper focuses in the possibility of a major impact on performance due the differences in the MM styles the participants developed. Also, considering the extension of the researches there is a huge difference, as with our previous work [12] we only answer if there are differences on the mental models and if there are differences on the performance due to the variation of experience with Twitter. In contrast this research retakes the idea of different MM styles and tries to answer another question: does MM styles have an impact on performance?

With this in mind, we aim to contribute to the HCI literature by providing more elements to understand how users learn to interact with a complex interactive system and provide valuable insights for designers of OSNS and other forms of novel communication.

II. PREVIOUS WORK

One of the greatest problems of working with MM is that we only have indirect proofs of its existence, majority of the researchers infer their existence through studies involving the observation of the differences between expert and novice users or comparing the performance of users after being exposed, or not, to mental models [13].

Brandt and Uden [5] found that people without expertise could not articulate their knowledge behind their Internet search skills; when asked why sometimes the system responded in a particular way, they tended to guess.

Dixon and Johnson [14] found that one of the most important difference between novice and expert users is the capacity to comprehend the problem because novice users needed more time and effort to transfer the model made during the “understanding the problem” phase to a more familiar model used in the “solving” phase.

Kieras and Bovair [15] and Borgman [16] showed that people who were exposed to MM while teaching them how does a device worked and how to use it, took less time to understand the way it worked and completed the given task in a quicker and more efficient way than people that weren’t exposed to the MM.

Zhang [17] found that participants used four styles of MM to describe the Internet (Technical, Functional, Process and Connection) and that these MMMs where influenced by the personality and previous experience of the participants.

Hmelo-Silver and Pfeffer [7] and Thatcher and Greyling [18] conducted a series of studies about the MM people defined while using complex systems (aquatic systems and the Internet). They found that participants used different styles of MM and that participants with a higher level of expertise tended to include more concepts and elaborated relations on their MM than participants with lowers levels.

Studies about how people create MM to understand OSNS or about how the sense making is done are not vast. The studies related to OSNS focus on how people use them, which new functions they can have, how they affect (psychologically) the users or how to design new interfaces to certain user groups. Some of these studies include:

Liu, Chung and Lee [19] found that the content (information’s topics) and the technological (usability and mobility) satisfaction were more important than the social (social interaction) and the process (hang out) satisfaction in order to keep people using Twitter.

Hargittai and Litt [20] found that people with higher computational skills were most likely to use Twitter and, that the information gathered from a close group of people (family, friends, etc.) play a significant role in the adoption of a new technology.

¹ An information network whose content is updated immediately.
² http://semiocast.com/publications/2012_07_30_Twitter_reaches_hal_f_a_billion_accounts_140m_in_the_US
³ http://techbeat.com/2013/01/is-twitter-becoming-too-complicated/
⁴ http://www.rexblog.com/2010/07/19/46427
To summarize, there have been a vast number of studies about MM and their relation with performance in different areas, like the Internet or biology, in order to have a better insight on the mental processes that occurred during learning, unfortunately OSNS have been omitted, since the studies about them have focused on why and how people use it.

III. METHOD

With an experiment conducted with 30 college students (aged from 19-25 years old), we tried to identify and compare the MM the participants defined and how they affected their performance and the perception of usability of the system.

A. Participants

30 participants (14 men and 16 women) were chosen randomly within our university’s student community regardless their gender, semester and course. Their ages ranged between 19-25 years old (μ=22, σ=1.6). They were divided into two groups (each one with 15 participants) according to their expertise with Twitter, group “Twitter” was formed by participants with a year or more of expertise and group “no Twitter” was formed by participants who were not users of Twitter.

B. Data Gathering and Procedure

Each participant had a 2-hour session divided in four parts: a key skills test, six Twitter tasks, a drawing activity, and a SUS test.

To identify the knowledge participants had regarding communication technology and information management, we used the Key Skills 4 U Practice Test ICT: Level 2 Test A, designed in the United Kingdom by Learning and Skills Improvement Service, in order to quantify possible biases.

Six activities were considered as Twitter’s functional core because they represent what a user must know to be considered an active user within Twitter (writing a tweet, looking for and following a user, giving retweet, using mentions, using hashtags and sending direct messages). The participants, using a Microsoft Windows (Meebox) tablet with OS Windows 8, were given a common usage context, and using Twitter’s website interface, they tried to complete each task. For each one, the completing time was measured along whether the participants were successful or not.

After each activity was completed, the participants were asked to draw a MM to explain the activity they had completed and to describe it in a few sentences. After the six activities the participants were asked to draw a MM that joined all the concepts and relations regarding Twitter. For the analysis nine concepts were defined as necessary, which are: tweet, time line, follower, retweet, mentions, hashtags, trending topics, direct messages and user search.

To measure users’ usability perception we used one of the most used tests, the System Usability Scale (SUS). In order to be clearer while interpreting its results we used the method proposed by Bangor, Kortum and Miller [21], which associates letters to the SUS Score like schools do with grades, bearing this in mind the results were interpreted as: 90-100=A, 80-89=B, 70-79=C, 60-69=D, 0-59=F.

The analysis of the results had two phases. The first phase, was an analysis over the MM from the participants. The second phase was an analysis over the performance (Key Skills and SUS Test and the results from the Twitter Tasks). This second phase was divided in two types: the first one considered only the MM style as variable of study and the second one considered both, the MM style and the participants’ level of expertise. This was due to the fact that most of the previous studies considered the level of expertise as major factor in the comprehension of a complex interactive system and we wanted to have a better insight on the influence of the MM in the performance.

IV. RESULTS

A. MM Analysis

After we analyzed the drawings and based on the elements drawn, how the concepts were drawn and the drawing descriptions provided by the participants, we identified three Mental Model Styles:

1. Analogy MM (AMM): the participants of this style (26.67% of the “Twitter” and 20% of the “no Twitter” sample) relied on comparisons with other domain concepts to explain the system. The majority of the members focused on the social aspect (interaction between users), therefore they used analogies with human activities (tweet = talk, time line = blackboard in a classroom) whilst other participants imagined Twitter as a place where all the tweets, profiles, etc. were stored, therefore they used analogies with objects (profile = folder, tweet = files) (Fig. 1).

2. Technical MM (TMM): participants (53.33% of the “Twitter” and 33.33% of the “no Twitter” sample) focused on the technological/technical elements of the system. Some participants pictured Twitter as a huge database where everything was kept, others pictured it like a massive net of servers in which the users were computers and others participants imagined it as a huge interconnection between various devices (tablets, mobile phones, lap tops, etc.) and the “cloud” (representing an indeterminate place of storage and distribution) (Fig. 2).

3. Conceptual MM (CMM): Most of the participants (20% of the “Twitter” and 46.67% of the “no Twitter” sample) only wrote the concepts and their relations, other participants used graphic elements of the system (like textboxes) or screenshots to describe them (Fig. 3).

The analysis showed the majority of participants used a technical view (over half of TMM’s members did have a Twitter account), at the AMM there was no trend as half of its members belonged to the “Twitter” group and the other half to the “no Twitter” group, and the CMM was formed, mostly, by members from the “no Twitter group”, this trend among the MM styles showed that the lower the expertise the harder it was to construct an elaborated MM and to apprehend and transfer the concepts and relations among different domains.
Within the AMM and the CMM participants without Twitter expertise mentioned more of the pre-defined elements which meant they were more complete than the participants from the “Twitter” group and, TMM, was the only style in which the participants with a Twitter account used a more complete model.

On the SUS, the ANOVA test (F (2, 27) = 4.21, p<0.03) showed that there was a difference in the perception about the usability of the system among the MM styles, a subsequent Post-Hoc Test (Turkey-Kramer Method) showed that the difference was, specifically, among the TMM and the CMM (with a significance of 0.019), due to AMM and TMM’s participants graded the system with a B and the participants from CMM with a C.

a) Twitter Tasks
Both statistical tests, ANOVA and Chi-Square, showed no evidence of a difference among the MM styles in completing in time nor in success rate (Table I).

2) Variables of Analysis: MM styles and level of expertise
a) Key Skills and SUS
There wasn’t evidence of significant difference between the participants with different levels of expertise nor between MM styles in the Key Skills Test. Again, this ensures us bias-free results.

The analysis showed that participants from AMM and TMM didn’t have significant differences had a similar perception of the system, grading it with a B, but the participants from CMM did (t (8) = 3.81, p<=0.005). Participants with more experience with the system graded it with a B whilst participants without experience were more critical about it grading it with a D. Among the MM styles the Two-Way ANOVA Test (F (2, 24) = 1.95, p<=0.16) showed there was no difference either. Though a general remark was that it would be quite useful to have some tutorials to help new users understand the system.

<table>
<thead>
<tr>
<th>MM Style</th>
<th>T-Test</th>
<th>ANOVA TEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMM</td>
<td>t (3)=1.03, p&lt;=0.28</td>
<td>F (2,24)=1.95, p&lt;=0.164</td>
</tr>
<tr>
<td>TMM</td>
<td>t (10)=0.81, p&lt;=0.43</td>
<td></td>
</tr>
<tr>
<td>CMM</td>
<td>t (8)= 3.81, p&lt;=0.005**</td>
<td></td>
</tr>
</tbody>
</table>

a) Twitter Tasks
Considering the mean of the completing times, the results showed that in retweet (RT) and direct messages (DM) activities there was a difference between the participants. In the RT activity, the difference appeared among the participants from both, the TMM and CMM styles and, in the DM activity, only in the TMM style; even though the difference depended again on the levels of expertise rather than the MM styles (Table III).

Considering the success rate, we found that none of the activities showed evidence of a significant difference neither among the participants nor among the MM styles. The reason that explains why there was no difference at the success rate is that participants from the “no Twitter” group after, failing to complete it at their first try, tended to use a trial-error method.
that eventually led them to the solution but it affected their completing times.

TABLE II. RESULTS FROM THE TWITTER TASKS AFTER THE SECOND ANALYSIS

<table>
<thead>
<tr>
<th>Activity</th>
<th>MM Style</th>
<th>T-Test</th>
<th>Two-Way ANOVA Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Tweet</td>
<td>AMM</td>
<td>t (2)=−1.29, p&lt;0.32</td>
<td>F (2,24)=0.50, p=0.951</td>
</tr>
<tr>
<td></td>
<td>TMM</td>
<td>t (8)=−0.62, p&lt;0.55</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CMM</td>
<td>t (5)=−0.74, p&lt;0.49</td>
<td></td>
</tr>
<tr>
<td>2: Search &amp; Follow</td>
<td>AMM</td>
<td>t (2)=−0.91, p&lt;0.46</td>
<td>F (2,24)=0.250, p=0.78</td>
</tr>
<tr>
<td></td>
<td>TMM</td>
<td>t (5)=−1.94, p&lt;0.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CMM</td>
<td>t (8)=−0.71, p&lt;0.49</td>
<td></td>
</tr>
<tr>
<td>3: Retweet</td>
<td>AMM</td>
<td>t (2)=−1.29, p&lt;0.32</td>
<td>F (2,24)=0.108, p&lt;0.90</td>
</tr>
<tr>
<td></td>
<td>TMM</td>
<td>t (6)=−3.52, p&lt;0.01**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CMM</td>
<td>t (6)=−2.41, p&lt;0.05**</td>
<td></td>
</tr>
<tr>
<td>4: Mention</td>
<td>AMM</td>
<td>t (4)=−0.96, p&lt;0.39</td>
<td>F (2,24)=0.50, p&lt;0.61</td>
</tr>
<tr>
<td></td>
<td>TMM</td>
<td>t (10)=−0.96, p&lt;0.36</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CMM</td>
<td>t (8)=−2.14, p&lt;0.07</td>
<td></td>
</tr>
<tr>
<td>5: Hashtag</td>
<td>AMM</td>
<td>t (3)=−1.64, p&lt;0.20</td>
<td>F (2,24)=0.25, p&lt;0.78</td>
</tr>
<tr>
<td></td>
<td>TMM</td>
<td>t (8)=−0.75, p&lt;0.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CMM</td>
<td>t (6)=−0.95, p&lt;0.38</td>
<td></td>
</tr>
<tr>
<td>6: Direct Message</td>
<td>AMM</td>
<td>t (3)=−1.70, p&lt;0.187</td>
<td>F (2,24)=2.51, p&lt;0.102</td>
</tr>
<tr>
<td></td>
<td>TMM</td>
<td>t (11)=−2.79, p&lt;0.02**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CMM</td>
<td>t (8)=0.92, p&lt;0.39</td>
<td></td>
</tr>
</tbody>
</table>

V. DISCUSSION

The main goal of this research was to identify the differences among the mental representations (MM) that people created while they interacted with an interactive system and their effects on the participants’ performance and usability perception.

The analysis on the Drawing Activity showed that there were three styles of MM, though only one of these styles (Technical) appeared in the studies of Zhang [17] and Thatcher [18]. There was some tendencies on the MM styles, the TMM was mostly used by experienced users (eight out of thirteen) and the CMM was mostly used by non-experienced users (seven out of ten), even though we could see that there were some participants with experiences used a more elaborated MM and participants with experience used a simpler MM, like Thatcher [18] we concluded that time of usage is relevant to form better MM but no necessary.

After analyzing the results we can conclude that participants didn’t have different ICT knowledge, regardless the differences in expertise with Twitter or the MM style, this was unexpected due the different characteristics the participants had like career, age and gender; this contrasted with the findings by Hargittai [20] who found that people with high computer skills were more likely to use Twitter than people with lower skills.

Regarding the Twitter tasks, the analysis that considered only MM styles showed that there was no significant difference in the performance (neither completing time nor success rate). The second analysis, the one that considered MM styles and level of expertise, showed that only two activities (Retweet and Direct Messages) had significant differences but only among participants with different levels of expertise, as expected the participants with more experience needed less time to complete the tasks. The other four tasks didn’t showed differences, considering the conclusions of Kieras [15], Zhang [17] and Borgman [16], this was expected because they concluded that level of expertise didn’t had an impact on the performance. One possible explanation for this is the tasks’ complexity. The Retweet activity can be considered as complex because they involved more steps to be done, like searching the user and then the tweet, than the other tasks that had direct links or could be done in the home page, while the Direct Message activity had the difficulty of guessing the direct button (the iconography used caused confusion) and searching the user. An interesting finding was that the two tasks required searching for users, so we think that the most complex task is looking for users due the GUI problems or the confusion of searching for users or words. Also, there wasn’t a difference considering the success rate, although our experience was similar to Brandt’s [5] due the participants without experience with Twitter tended to use more the trial-error method.

This research shows that, although using MM does help understanding the theoretical part of an interactive complex system (along an explanation of how to use it is needed, including the GUI), it alone doesn’t guarantee a better performance as previous experience with similar systems prove to be also very influential.

Against the odds the SUS analysis presented the more interesting results. The first analysis showed that there was a significant difference among the MM styles. The AMM and TMM’s participants were more benevolent and graded the system with a B while CMM’s participants graded it with a C, nevertheless a Post-Hoc test showed that the difference was, actually, among only two MM styles, the TMM and the CMM. The second analysis showed that only the participants from CMM had a difference in how they saw usability in Twitter, while participants with previous experience graded it with a B, the participants without experience graded it quite low, with a D. These results are particularly interesting, if you have in mind the way the MM styles were formed, because in the CMM over 70% of the members didn’t have experience with Twitter, while 60% of the TMM members had more expertise. So we concluded that there were significant differences on how people see usability in Twitter, but these differences are due to the differences in the levels of expertise rather the MM styles participants defined.
VI. REACH AND FUTURE WORK

This research is intended to be a complement of a previous research[12] in order to gain a better insight in the process of sense making people has while using a complex interactive system, therefore it answers two questions (that limit the reach of the investigation): are there different MM styles? And do these MM styles affect the performance? It’s important to notice that these questions represent only earliest steps to get a full insight of the mental process of understanding a complex interactive system. Studies focused on how the MM styles affect performance and perception, how new users create and modify MM through time are needed or how can performance be affected by a combination of theoretical and practical explanation, even studies with other kinds of interactive complex systems (in this case another OSNS would be ideal to show if the MM created for one could be applied to others) are required.

But this kind of studies can be useful, not only under a research perspective, but also a practical one. Different methods of training for using a device or a system, that guarantee a much better comprehension, could be developed according to the MM style of the final user.

To conclude, we’re still far away from having a complete knowledge of the processes that take place on the users mind that compel them to act the way they act, even though the first steps have already been taken and everyone can be benefitted from them like researchers, designers, programmers and, of course, final users.

ACKNOWLEDGMENTS

We would like to thank all the participants (from the pre-tests to the final test) who gave us a couple of hours from their busy schedules. We thank Meebox S.A. de C.V. for facilitating the devices to conduct the study. This work has been supported by Asociación Mexicana de Cultura A.C.

REFERENCES


