Claims Require Evidence: An Essay on Arguments and Reasoning in HCI Papers

**Shi Yin**

University of Illinois at Chicago

Computer Science Department

syin8@uic.edu

# INTRODUCTION

Since Aristotle of Ancient Greece, logic has been studied in the disciplines of philosophy, mathematics, semantics, and computer science [7]. Reason is the capacity for consciously making sense of things, for establishing and verifying facts, and changing or justifying practices, institutions, and beliefs based on new or existing information [8]. But after evolving for thousands of years, human beings are still making mistakes (fallacy) when using logic to reason an argument. In this essay, we are going to go through and analyze three first-tier conference articles, one is considered very good in terms of using evidence to support claims, one considered average and one below average.

# A GOOD PAPER

## Background

In 2011, Lena Mamykina et al. submitted a paper entitled *Design Lessons from the Fastest Q&A Site in the West* to the annual CHI Conference [5]. In this paper, they analyzed a Question & Answer site for programmers, Stack Overflow (SO), found that over 92% of questions in this website were answered, and in a median time of 11 minutes. The major part of this paper is they using both quantitative analysis and qualitative interview to seek to the reason behind SO’s success and then arguing that this success attributes not only to the careful design considerations, but also to the high visibility and interactive involvement of the design team in the community [2].

## Claims and Evidence

This article listed some quantitative findings, such as ‘92.6% of questions are answered’, ‘first answers are in 11 minutes (medians)’, ‘most answer activity takes place in the first hours’ and so on. All these findings were directly based on the data analysis tool provided by SO and were illustrated very decently by figures. Since the number of SO users was as many as 300K and there were 2.2M answers (as of August 2010), this derivation process can be considered convincing. With this solid evidence in hand, the claim that SO is very successful is with high reliability.

In the following qualitative study section, the authors described interviews of 2 SO founders, 4 site design team members and 6 SO users. Their finding through this interview, that the success of Stack Overflow is due to the careful design considerations and the involvement of the design team within the users, is also somewhat convincing because of following reasons.

### Diversity among participants group

The interviewees consisted of project leaders, project members and project users. This diversity enabled a good understanding in all points of view. Along with the fact that the authors designed different questions for interviewees with different identities, it made the feedback from this interview very trustworthy.

Some may say that the authors should also interview users that quite inactive at SO to ensure the sample more complete. But I don’t think it is necessary, since inactive users may not have enough constructive opinion. Interviewing these users may let us know the aspects of SO that needs improvement, but this is not relevant to the ‘success’ of this site.

### Multiple quotations and facts

The authors provided lots of feedback from interviewees, all of which made the authors assumptions apparent. They also provided information from quantitative facts to support their assumptions (e.g. Figure 14 of this paper).

A common concern is that these quotations of interviewees may be partial. One reason is that the authors might only show what they’d like to hear so that they can draw a conclusion as they wanted. Another reason is the number of samples is too small. However, these defects are of qualitative research itself, not due to the authors’ mistake [1, 3].

## Conclusion

The authors provided solid evidence to show that Stack Overflow is indeed successful and explained this success by showing the feedback of an interview. Regardless some inherent shortcomings of qualitative research, this paper did a very good job supporting their claims. There might be another factors that contributed to the success of SO need to be found, but all the proposed reasons in this paper are of high creditability. Thus we consider this paper as of high quality.

# AN AVERAGE PAPER

## Background

Izabel Olson et al. submitted a paper called *“It’s just a toolbar!” Using Tangibles to Help Children Manage Conflict Around a Multi-Touch Tabletop* to TEI Conference in 2011. In this paper they presented a case study of children’s collaborative behavior around a multi-touch tabletop interface. They observed collaborative behaviors and conflict during the study and tried to solve this conflict by introducing a new tangible interface to control the toolbar. They argued that the tangible object “seemed to help the children resolve their conflict and to promote spontaneous turn taking behavior” [6].

## Claims and Evidence

This article described a 4-session study of 4 children collaborating on tabletop and desktop environments, with different interface to control the system. They then created a coding scheme to interpret the behaviors of the children, followed by a discussion of these behaviors. Generally speaking, the description is clear to understand, the coding scheme is reasonable and the result, coming with many transcripts of the video recorded collaborating process, is accurate.

However, according to their description and the result they showed, there might be alternative explanations to some phenomena.

### The first session is peace, the third one is war

The authors claimed that this is because the children found out that the toolbar could be dragged, but in fact it maybe just because during the first session, the children were new to the environment, thus performed with caution; after they got familiar with each other, with the researchers and with the research environment, they start to fight. Since the sessions were video recorded, it is with high possibility that the children wanted to be ‘good’ at first. The authors themselves said that ‘The children seemed to have a clash of expectations regarding appropriate social protocols in the tabletop sessions’. Thus this claim may be a ‘cum hoc ergo propter hoc’ fallacy.

### Behavior patterns in different sessions are different

The authors attributed this to the different types of interface the children had. But there were other variables that may result in different patterns. For example, the children played in different scenarios (wolf-sheep ecosystem, virus model, etc.). Maybe it is this difference in scenarios that result in the different behavior patterns.

Moreover, there are also some suspicious findings. In Figure 4 of this paper, we can see that ‘Grabbing’ in Session 4, in which the children used a tangible object, is not much fewer than in Session 3. It indicates that they were still not in real peace in Session 4. Also, the definitions of ‘Reaching Over’ and ‘Blocking’ are more likely related to a toolbar in the screen but not very relevant to an object held in hand. When a tangible object held in one child’s hand was necessary to activate the control toolbar, it’s useless to ‘reaching over the workspace to try to use the control’ and it’s unnecessary to ‘preventing another child from using the control using one’s body, arm, or a finger’. I doubt that there could be much of these two actions in the situation of Session 4. Thus the low appearance rates of ‘Reaching Over’ and ‘Blocking’ in Session 4 can hardly draw the conclusion.

## Possible Improvements

According to our discussion above, first we need better definitions the attempt of acquire or keep the possession in a better way. And if another group of children with different order of sessions could be introduced, the analysis might be more convincing.

## Conclusion

Because of the uncertainty brought by multiple inherent variables that were not handled well by the authors in this study, the conclusion of this paper that the tangible object helped children share controls and collaborate in a more effective way is somewhat but not very convincing. Since most of the defects mentioned above were not blamable to the authors, we consider this paper is in average level.

# A BELOW AVERAGE PAPER

## Background

Edge et al. published their research entitled *MicroMandarin: Mobile Language Learning in Context* in CHI 2011. In this paper they designed two versions of software for language study based on their knowledge in Cognitive Psychology and the theories from Second Language Acquisition (SLA), then conducted a study to claim that their design is of high value by providing much evidence [4].

## Claims and Evidence

We are not interested in how they expanded their findings from Cognitive Psychology through analysis of prior systems, their user research and theories from SLA to draw their new understanding that resulted in this new ‘contextual microlearning’ mobile application they presented in this paper. We will only argue why their evidence is far from adequate to show the value of contextual microlearning. Due to the pervasiveness of flaws in this paper, we’ll discuss in a different manner from the way in above two papers.

### Surveys are too subjective

One example is that the self-reported language levels are not specific enough to avoid personal bias. We have no clue whether the participants knew the definitions of these levels when selecting among them. If this was not the case, I’m afraid the level distribution is not valid, since different people would have different understanding of these vague levels. Another example is that what should be considered as frustrated is not clear in this paper. It is possible that some participants tend to not feel frustrated when they encountered difficulties while others very easily feel frustrated even with a small obstacle. Thus the answer distribution of surveys might not reflect the fact. There are more such examples in this paper.

### Hypotheses are not validly proved

The six hypotheses proposed in discussion section of this paper are not valid proved. First, H2 was not significant. Second, H1’s significance is not helping make any conclusion. Third, the author failed to provide adequate information to justify their interpretation of H5 and H6’s significance.

### Hypotheses are not relevant to the claim

Even if these hypotheses are justifiable, they still don’t show the value of the contextual microlearning. First, some hypotheses are not significant. Second, even though H3 and H4 are significant, the author failed to connect this significance to their claim. Third, maybe the significances of these hypotheses are adequate to indicate that the contexture version is relatively better than the frequency version, but through comparison of these two versions, our conclusion should stop here and not moving further, especially when the contextual version was compared with a version that was also not common and general.

## Possible Improvements

This paper has so many flaws that some minor improvements are not enough to build the validity of their claim. In general, the whole structure of this paper need to be improved, most importantly and essentially, is to make relevant assumptions and provide constructive evidence. The other advice is to organize material in a more efficient way.

## Conclusion

The authors of this paper failed to create a clear reasoning process. It is more likely piling up a series of facts and claim as the authors wanted, regardless whether they are relevant than a reasonable logical process. The authors provided much evidence for multiple hypotheses, but even of significance according to the t-tests, they have no proved correlation to the claim that the contextual microlearning is of high value. Thus we consider the reasoning chain in this paper as not valid.

# CONCLUSION

From above discussion, it is sufficient to draw that papers in top-tier conferences are of different qualities in term of reasoning. The flaws may be the result of the limit in current research condition, the complication of the research topic or the defect in the design of the study.

# REFERENCES

1. Advantages and disadvantages of qualitative research methods

<http://hive.library.uwa.edu.au/hive/hive.cgi/zip/218293/html/pop_advdis.html>

1. Chia, P.H., Chuang, J. Community-based Web Security: Complementary Roles of the Serious and Casual Contributors. In *Proc. CHI’11*. 1023 – 1032.
2. Disadvantage of Qualitative Research

<http://www.okstate.edu/ag/agedcm4h/academic/aged5980a/5980/qualrsch/QUALRSCH/sld010.htm>

1. Edge, D., Searle, E., Chiu, K., Zhao, J., Landay, J.A. MicroMandarin: Mobile Language Learning in Context. In *Proc. CHI’11*. 3169 – 3178.
2. Mamykina, L., Manoim, B., Mittal, M., Hripcsak, G., Harmann, B. Design Lessons from the Fastest Q&A Site in the West. In *Proc. CHI’11*. 2857 – 2866.
3. Olson, I.C., Leong, Z.A., Wilensky, U., Horn, M.S. “It’s just a toolbar!” Using Tangibles to Help Children Manage Conflict Around a Multi-Touch Tabletop. In *Proc. TEI’11*. 29-36.
4. Wikipedia Page of Logic <http://en.wikipedia.org/wiki/Logic>
5. Wikipedia Page of Reason <http://en.wikipedia.org/wiki/Reason>