Exploring the Role of Activity Trace Design on Evaluations of Online Worker Quality

Jennifer Marlow[1,2]
[1]FX Palo Alto Laboratory, Inc.
3174 Porter Drive
marlow@fxpal.com

Laura Dabbish[2,3]**, Jodi Forlizzi**[2]
[2]Human-Computer Interaction Institute, [3]Heinz College
Carnegie Mellon University
{dabbish, forlizzi}@cs.cmu.edu

ABSTRACT

Websites can record individual users' activities and display them in a variety of ways. There is a tradeoff between detail and abstraction in visualization, especially when the amount of content increases and becomes more difficult to process. We conducted an experiment on Mechanical Turk varying the quality, detail, and visual presentation of information about an individual's past work to see how these design features affected perceptions of the worker. We found that providing detail in the display through text increased processing time and led to less positive evaluations. Visually abstract displays required less processing time but decreased confidence in evaluation. This suggests that different design parameters may engender differing psychological processes that influence reactions towards an unknown person.

Author Keywords

Activity traces; Impression formation; Information presentation; Crowdsourcing

ACM Classification Keywords

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

INTRODUCTION

From social media sites to online work environments, technological platforms are increasingly recording and displaying individuals' activity histories. This is characteristic of a general trend in online social settings, where we are starting to see activity streams as a widely used form of information about other Internet users [5]. In the context of online peer production, access to individuals' past behavior and history can benefit collaborations and jump-start interactions by indicating who knows what or who has done what [7]. Visualizing this information can also influence how we feel about other people [8]. However, a variety of activity history design decisions

could influence the positivity of and confidence in initial impressions formed. Activity traces can be designed with varying levels of *abstraction* (the degree to which they use text versus visual abstractions) and *granularity* (the degree to which they provide detail versus aggregation).

One main challenge in presenting activity histories is managing information overload [4], and the choice between making information more accessible and easy to interpret (in the form of visual vs. textual presentation) is a primary consideration when displaying a large amount of information [12]. The amount of detail or granularity provided in an activity history has implications for the amount of mental effort required to process the information. A tradeoff exists between the level of detail of available information about a person's past behavior and the amount of time and cognitive effort required to fully interpret that information [11]. Careful information processing may come at the expense of efficiency [6].

To investigate these tradeoffs in activity history design, we conducted an experiment using an online peer review work task. We compared abstracted visual displays with detailed text displays of a participant's past work in an online crowdsourcing platform. We wanted to understand how the visual display of activity history affects the positivity of and confidence in evaluations of a worker in a lightweight peer production context.

The contribution of this work is insight into how the design parameters of granularity and aggregation influence psychological reactions to and perceptions of individuals. Our results inform the design of systems that record and display individual work traces.

RELATED WORK AND HYPOTHESES

Here we discuss how different visualizations of work history may affect psychological reactions to the worker.

Detailed textual displays

Text-based lists of activity information are detailed but require effort to interpret. A visual summary of the same information might be easier to process quickly but be less informative overall, as is the case with the London Underground map [1].

Assuming the content of activity is neutral (neither positive nor negative,) providing additional detailed information about past activities could help to reduce uncertainty about a person, which has been associated with increased liking

[3]. Prior work has also indicated that when people have a more detailed and concrete mental image of what a person is like, they feel more familiar and behave more pro-socially towards that person [11].

H1a: More granular activity traces will lead to more positive subsequent evaluations of the person through providing more concrete details about a person's past behaviors.

However, if expending mental processing effort leads to negative reactions [13], then an easy to process visual summary may induce positive impressions:

H1b: Less granular activity traces will be easier to process and increase positivity of impressions.

Processing fluency and abstracted visual displays
In addition to granularity (or amount of detail in the information), displays may be more or less abstract, which could affect effort of interpretation. Processing fluency can be defined as the subjective experience of ease with which people process information (or how effortful it is for them to do so) [2]. Fluency, or reduced processing effort, has been linked to positive interpersonal evaluations such as liking and perceived intelligence [2]. Therefore, it is possible that making things easier to process through presenting them in an abstract summary format will lead to more positive evaluations.

H1c: More abstract activity traces will lead to more positive evaluations of the person than less abstract traces by providing information in a way that is easier to interpret.

Impression confidence
Prior research suggests that people's behavior is influenced not just by what they think but by how confident they are in their likes and beliefs [13]. We were therefore interested not only in the evaluations made by people but by the degree to which they were confident in their attitudes.

Prior research on confidence of impressions suggests that increasing the amount of information available about a person increases confidence in in impressions of that person [13]. In addition, more details (even uninformative ones) increase the clarity of a perceiver's mental picture and in turn increase confidence [13]. Thus we expect:

H2a: More granular and detailed textual displays will lead to more confidence.

However, processing fluency can also increase confidence in evaluations [2]. This means detail may reduce confidence as a function of processing effort, and abstraction should enhance it.

H2b: More visually abstract representations of activity traces will lead to more confidence.

Hypothesized interaction effects
Finally, we might also expect that the effects of abstraction versus detail could depend on whether the work history

being generated is generally positive or generally negative. For example, minimizing evidence of bad work through abstraction could improve positivity of impressions of the worker by de-emphasizing mistakes.

H3: Visual abstraction will affect positivity of impressions for bad quality work but will not affect impressions for good quality work.

METHOD
In order to investigate the effects of level of detail versus visual aggregation, we designed three types of history visualizations. The designs (see Figure 1) were inspired by common formats of activity trace presentation on sites that use multiple formats to showcase individual work history in the context of software production, including GitHub, Stack Overflow and Masterbranch (see, e.g.[10])
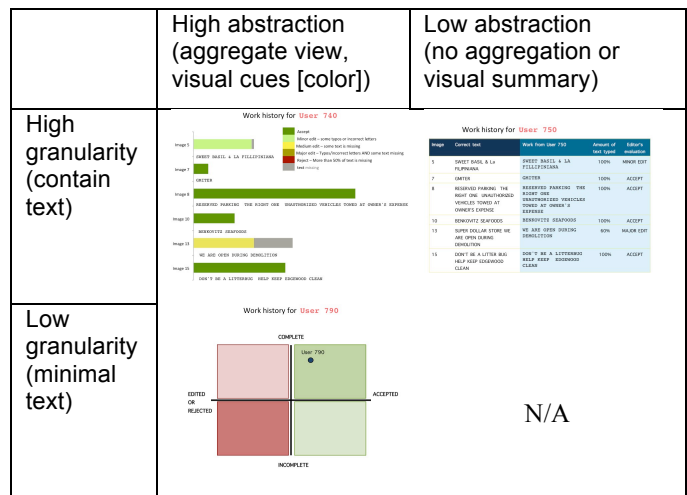
Figure 1. History visualizations (top row: graph (L), table (R); bottom row: quadrant). Examples shown here depict good work.

To separate the effects of granularity and abstraction, the visualizations varied from each other in the following ways: The *table* condition was high in granularity (with detailed text) and a low degree of abstraction. The table condition showed a grid filled with the text the participant had typed, the correct answer, the rating by the editor, and the percent of the work that had been completed. It showed a worker's work actions for six pieces of work.

The *graph* condition also had a high degree of granularity (textual information) but also featured some other visual abstraction cues in the form of color codes pertaining to the quality of the work. It contained the same information as the table, except the percent of the work completed and the rating by the editor were displayed in a visual format (through colored bar graphs.) It also showed work actions for six pieces of work. Finally, the *quadrant* condition was the most abstract, contained minimal text and plotted the participant on a grid showing the average completeness of their six pieces of work and the average evaluation of their work.

Work history quality

Finally, to test H3 we manipulated work history quality. For each of the conditions, there was a "good" quality history and a "bad" quality history. The information seeding the work activity history was taken from real performance on the task by Mechanical Turk workers. Overall, the good performance was on average 92% correct, while the bad performance was on average 50% correct.

Experimental setup

The experiment was set up as a crowdsourced task (HIT) in Amazon's Mechanical Turk. Participants were paid 20 cents to participate in a task billed as "Transcription Worker Evaluation." The description of the task was: "You will be shown some image transcription work done by another Turker and asked to evaluate its quality." Upon accepting the task, participants were shown an example of the task and told what they would be doing. The task involved transcribing the text found on images of various signs. Participants were shown examples of these signs.

At this point, participants were randomly assigned to one of the possible history visualization conditions. After viewing the history, they were asked to evaluate the person and the activity history display on the following criteria:

- Work quality ("How would you rate the overall quality of this person's work? Consider both the amount of text typed and the editor's evaluations.") This was on a sliding-point scale from 0 (poor) to 100 (excellent).

- Confidence in assessment of worker ("How confident are you in your response to the previous question?") (7 point Likert scale, 7=very confident)

RESULTS

184 participants completed the task, and they were roughly evenly distributed across the table, graph, and quadrant (good and bad) conditions. Table 1 shows the mean values for work quality rating, confidence, and time (in seconds) spent viewing the work history across the conditions.

	Granularity (Low=quadrant, High = graph + table)			Abstraction (Low= table, High=graph + quadrant)		
	Rating*	Conf.	Time*	Rating	Conf.*	Time*
Low	65.46^a	5.63^a	35.87^a	59.75^a	6.04^a	50.4^a
High	57.04^b	5.72^a	47.9^b	60.08^a	5.50^b	40.04^b

Table 1. Mean values for rating, confidence, and viewing time (in seconds) across conditions. Values not connected by same letter are significantly different at p <.05 (Wilcoxon test)

Manipulation checks

There was a significant main effect of history quality on impression positivity. That is, participants who saw the "good" history rated the work as significantly higher in quality (M=74.79) than participants who saw the "bad" history (M=44.98) (t[179]=37.01, p<.001). This confirmed that people were able to accurately interpret the quality of information portrayed in the visualizations, overall.

Additionally, to test the assumptions that detailed textual displays would require more mental effort to interpret than displays without text, we recorded the amount of time in seconds spent viewing the visualization page and compared that across the low and high granularity conditions. Because combining conditions led to uneven numbers in categories, we used a Wilcoxon non-parametric test to compare values across low and high-granularity visualizations and low and high-abstraction visualizations. As Table 1 shows, participants spent significantly more time viewing displays that were high in granularity. The same occurred for abstraction: Low abstraction involved more time spent on the display than high abstraction.

Effects on positivity of impressions

Hypothesis 1a predicted that a more granular visualization would lead to more positive impressions. Conversely, Hypothesis 1b predicted that a less granular, or aggregated visualization would lead to more positive impressions of worker quality. There was a significant main effect of granularity on rating quality (Z=2.09, p=.03), with the no-text (quadrant) condition providing more positive ratings than the granular, text-containing (graph and table) conditions. This supported Hypothesis 1b, and the notion that "less can be more" when learning about others [9]. The quadrant condition was also more accurate in their ratings of worker quality than the graph and table.

Hypothesis 1c, which posited that the more abstract visualization would lead to more positive impressions because it was easier to interpret, was not supported; abstraction did not have a significant effect on positivity of rating.

Effects on confidence

Hypothesis 2a predicted that more detailed textual displays would lead to greater confidence than less detailed ones, while Hypothesis 2b predicted that more abstracted visual displays would lead to greater confidence than less abstracted displays. There was no significant main effect of granularity on confidence but there was a significant main effect of abstraction on confidence (Z=3.79, p<.001). However, the opposite of Hypothesis 2b occurred: participants viewing the low abstraction format (table) were significantly more confident (M=6.04) than those who viewed the high abstraction graph or quadrant (M=5.50).

Interactions between work quality and visualization

We also were interested in further exploring the significant results to see if different visualizations had differing effects for good and bad quality. For example, minimizing details of bad work might lead to higher impressions than presenting bad work in a detailed format.

We conducted a 2x2 ANOVA comparing bad vs. good work and detailed vs. not-detailed presentations on

positivity. We used planned contrasts to compare the effect of exposure to bad work (detailed) and bad work (not detailed) as well as good work (detailed) and good work (not detailed). While the details did not affect positivity of impressions for good work, they did for bad work. Bad work presented in a textually detailed format (M=40.42) led to less positive ratings of quality than bad work presented in a format without text (M=53.67). This supported H3.

We conducted a similar 2x2 ANOVA comparing bad vs. good work and abstract vs. non-abstracted displays. In this case, abstraction did not affect confidence when the work was bad. However, when the work was good, lack of abstraction made people more confident (M=6.16) than when the work was presented in a visual abstraction (M=5.49).

DISCUSSION

In this work we saw how presenting activity history in a detailed, text-based versus visually abstract summary display had differing effects on various attitudinal outcomes such as positivity and confidence. Displays high in text-based detail fostered positivity in ratings of a worker but not confidence. On the other hand, displays that featured information in a visually abstracted format employing features such as color and aggregation decreased confidence in rating but did not influence positivity.

This could have implications for how people interact with one another or assess others' competence and ability in a peer production and peer review type setting (where typically little to no information about others' past work is provided). Increasing confidence in impressions may be detrimental if the initial impressions that people form are inaccurate. It could be that making activity history more detailed would increase confidence but discourage accuracy of processing by making observers less likely to consider all information [6]. A promising direction for future work will be to examine the effects of initial impressions (including positivity and confidence) on behavioral outcomes as well to see what the downstream implications of initial impressions are for a work product.

Our work contributes to research on impression formation online, especially by providing insight into how the presentation style and the quality of information interact. Designers of systems that employ activity traces should consider how display decisions can lead to confidence or positivity and adapt presentations accordingly to achieve desired goals. For example, if a system is able to automatically detect that a worker is performing poorly, it might display that information in a summary display in order to reduce negative opinions of that person (in accordance with findings supporting H3). However, if the goal of the system is to have workers improve each other's work, then providing bad work history in a detailed format could induce a negative impression that might make editors more sensitive to finding mistakes in the work.

REFERENCES

1. Ainsworth, S. The functions of multiple representations. Computers & Education 33, 2–3 (1999), 131–152.

2. Alter, A.L. and Oppenheimer, D.M. Uniting the tribes of fluency to form a metacognitive nation. Personality and Social Psychology Review 13, 3 (2009), 219–235.

3. Berger, C.R. and Calabrese, R.J. Some explorations in initial interaction and beyond: Toward a developmental theory of interpersonal communication. Human Communication Research 1, 2 (1975), 99–112.

4. Bontcheva, K. and Rout, D. Making sense of social media streams through semantics: a survey. Semantic Web Journal, (2012), 373-403.

5. Counts, S. and Fisher, K. Taking it all in? Visual attention in microblog consumption. In Proc. ICWSM (2011), 97-104.

6. Hullman, J., Adar, E., and Shah, P. Benefitting infovis with visual difficulties. *IEEE Visualization and Computer Graphics* 17, 12 (2011), 2213–2222.

7. Kim, D. and Shipman, F.M., III. Visualizing History to Improve Users' Location and Comprehension of Collaborative Work. In Proc. GROUP, (2012), 11–20.

8. Liao, Q.V. and Fu, W.-T. Expert voices in echo chambers: effects of source expertise indicators on exposure to diverse opinions. In *Proc. CHI*, (2014), 2745–2754.

9. Norton, M.I., Frost, J.H., and Ariely, D. Less is more: the lure of ambiguity, or why familiarity breeds contempt. *Journal of Personality and Social Psychology* 92, 1 (2007), 97-105.

10. Singer, L., Figueira Filho, F., Cleary, B., Treude, C., Storey, M.-A., & Schneider, K. (2013). Mutual assessment in the social programmer ecosystem: An empirical investigation of developer profile aggregators. In *Proc. CSCW*, (2013), 103–116.

11. Stephan, E., Liberman, N., and Trope, Y. The effects of time perspective and level of construal on social distance. *Journal of Experimental Social Psychology* 47(2), (2011), 397–402.

12. Thompson, R.S., Rantanen, E.M., Yurcik, W., and Bailey, B.P. Command line or pretty lines? Comparing textual and visual interfaces for intrusion detection. In *Proc. CHI*, (2007), 1205-1214.

13. Tsai, C.I. and McGill, A.L. No pain, no gain? How fluency and construal level affect consumer confidence. *Journal of Consumer Research* 37, 5 (2011), 807–821.