Identifying Knowledge Estimation Cues in Online Writing Workspaces

Arielle Elliott and William Horton

July 12, 2020

Identifying Knowledge Estimation Cues in Online Writing Workspaces

Arielle Elliott and William S. Horton

Department of Psychology

Northwestern University

Author Note

The authors declare that there no conflicts of interest with respect to this preprint.

Correspondence should be addressed to:

Arielle Elliott

Department of Psychology

2029 Sheridan Road

Evanston, IL 60208-2710

Email: arielleelliott2020@u.northwestern.edu

Abstract

The ability to assess another's epistemic state is important for successful collaboration. While many online contexts limit access to communicatively relevant audiovisual information (like facial expressions or gestures), platforms like Google Docs allow users to view the dynamics of written message production in real time. In this study, participants saw video screen captures of people typing in descriptions of familiar and unfamiliar objects. Visible typing speed and fluency reliably influenced epistemic judgments about the typist.

Keywords: written production, online collaboration, epistemic inferences

Identifying Knowledge Estimation Cues in Online Writing Workspaces

In face-to-face interactions, people rely on real-time feedback from their partner to adjust what they say depending on beliefs about what that person does or does not understand (Hinde, 1972; Clark & Marshall, 1981). These epistemic inferences about another's certainty and confidence are influenced not only by what that person says (message content), but also by how they say it (message delivery). The ability to accurately assess another person's understanding becomes especially important during collaborative tasks, as partners work together to achieve a common goal. While face-to-face contexts permit access to cues like gestures or speech disfluencies, text-based interfaces facilitate more subtle delivery-based cues, such as the relative timing of entire turns (Lea & Spears, 1992; Tidwell & Walther, 2002). In shared workspaces like Google Docs, though, users have real-time access to others' (written) communicative actions, permitting greater access to potentially informative cues. An open question is whether the dynamics of written message production in certain online environments are able to convey information about a partner's degree of certainty, similar to how speech pauses or disfluencies convey such information in spoken conversation.

As a first step toward examining the impact of real-time text-based cues on estimates of others' knowledge, we previously conducted a study intended to obtain basic information about the kinds of visible typing cues that might be associated with task uncertainty (Elliott & Horton, 2019). In a computer-based task, we asked participants to compose brief descriptions of sets of images. There were three categories of images – tangrams, photographs of objects, and facial caricatures – and half of the images in each category was either relatively difficult or easy to describe. Additionally, a subset of images repeated across multiple rounds of the task, thus becoming more familiar to participants. We used keylogger software to record the timing of each keypress (including backspaces and other special keys) and also collected video screen captures of the participants' typing behaviors. Analyses of the keypress data revealed that when participants were less familiar with the images, they displayed slower, less fluent typing patterns, including more pauses and more backspaces associated with error correction. This confirmed that message uncertainty can be reflected in the output visible onscreen, providing insight into the kinds of typing behaviors that would be most impacted by uncertainty.

Given these findings, a relevant question becomes whether *observers* of these behaviors would draw relevant inferences about user certainty. Here, we explore this question by presenting clips from the screen recordings collected as part of our previous study to a new set of naïve viewers, who were asked to make judgments about the level of confidence, familiarity, task effort, and task difficulty experienced by the original typists. In doing so, we are interested in how viewers watching a screen-recording of someone's real-time typing behavior would interpret these behaviors, and whether typing fluency and speed, in particular, would shape viewers' perception of typists' confidence and task familiarity.

Method

Participants

Four hundred and seven participants from Amazon's Mechanical Turk were recruited for this study via Turk Prime (www.cloudresearch.com). The Turk Prime interface allowed us to restrict participation based on first language learned and geographic location. Specifically, we restricted participation to native English speakers located within the United States, Canada, Australia, the Caribbean, and the United Kingdom. To further ensure participants were native English speakers, our survey included open ended questions about language experience.

Materials

Video materials for this study were selected from the computer screen captures collected as part of Elliott and Horton (2019). These screen captures recorded the key-by-key typing and mouse-movements of participants as they completed an image description task. From these screen recordings, we created separate video clips of typed descriptions for single items. Each selected clip began when the participant first visibly clicked the text description box to start typing and ended when they moved the mouse to click the arrow at the bottom of the screen to move to the next item.

The previous study employed six sets of images: tangrams, facial caricatures, and objects, with a set of images in each category selected to be "Easy" or "Difficult" to describe (48 unique images total). Each round of the task asked participants to type in, one-by-one, descriptions for individual items presented within sets of eight, and initial rounds involved homogenous category+difficulty sets (e.g., all eight "easy" tangrams). In later rounds, half of the images in each set repeated twice more, in mixed-difficulty sets (e.g., eight tangrams, half 'easy' and half 'difficult'). For the current study, we focused our selection of video clips on the 24 items (eight from each category) that were presented multiple times, as this gave more source descriptions from which to choose.

For each item, we selected multiple clips from the video captures, choosing clips that contained relatively fluent or disfluent typing. Specifically, each video clip was classified as either "Fast" or "Slow" in typing speed (measured as characters per second) and as either "Frequent" or "Infrequent" in backspacing (measured as proportion of backspace keypresses out of total keypresses for each description, a proxy for overall degree of error correction). These classifications were based on the measurements collected as part of Elliott and Horton (2019).

Together, they allowed us to examine whether certain typing behaviors would have more influence on viewers' inferences about the typist, as well how combinations of features might shift the viewer's interpretation. To reduce the degree to which participants would be influenced by the content of what was being typed, we attempted to select clips in which the description for a given item was similar in length across examples. We also tried as much as possible to ensure that the content of each clip mentioned similar aspects of the image. For instance, descriptions of ambiguous tangrams ranged from holistic interpretations (e.g. "a man running") to analytic shape-based formulations (e.g. "the triangle on top connected to a square"). In these cases, we selected clips for each disfluency category that included similar description types. Similarly, for descriptions of facial caricatures, we selected clips in which the participants referenced the same set of features.

These considerations resulted in a final set of 45 unique clips: 7 Slow speed + Frequent backspacing, 11 Fast speed + Frequent backspacing, 11 Slow speed + Infrequent backspacing, and 14 Fast speed + Infrequent backspacing. The Fast+Infrequent item category can be considered maximally "fluent" (i.e., the typing was relatively quick with minimal pauses, error corrections, or editing), while the other three combinations were each "disfluent" in different ways (either slow typing, frequent backspacing, or both).

The video clips were embedded into a Qualtrics questionnaire, which presented each respondent with only a single randomly chosen video out of the full set. After presenting the clip, an initial block of questions asked about the participants' perceptions of the typist. These included: *How confident do you think this person is about how to describe the image? How much effort do you think this person is putting into their description? How familiar is this person with the image they're describing? How much difficulty is this person having with their description?* Participants responded to each question on a 0-100 sliding scale.

A second block of questions prompted them to explain their answers to the previous block. That is, they were asked what aspects of the clip led to their initial judgments. For instance, one question stated "Do you feel like you were able to get a clear sense of what the typist was thinking as they carried out this task? Why or why not?" The final block of questions asked for Likert-scale ratings concerning whether participants noticed specific disfluency features in the clip. This was intended to assess what types of disfluency might be most salient to a viewer. All participants viewed the same set of questions, in the same order.

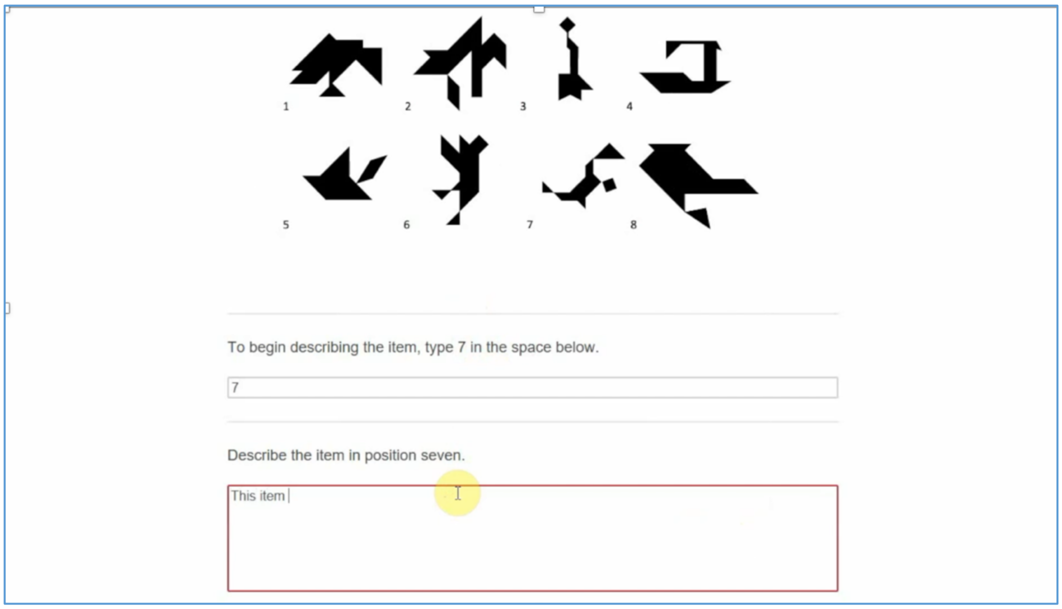
Following these questions asking for judgments of the video, the Qualtrics survey ended with several questions asking about the amount of time the participants spend online in general, their experiences with various types of online collaboration, and their rate of use of online communication.

Procedure

After accessing the survey through the Qualtrics link, participants read instructions informing them that they were being asked to carefully watch a short video in order to answer some questions about it. They were told that the video was a screen capture taken from a previous image description task, and that they would watch as someone typed in a description of a visual image, which would also be visible in the screen recording. Figure 1 presents a sample static screenshot of what participants saw. The video clip was embedded directly within the Qualtrics survey, and participants were asked to carefully watch the video, and were not allowed to click through to the next page until they remained on that screen long enough to view the entire video at least once. On the next screen, participants were presented with the first set of questions and were given the opportunity to watch (and re-watch, if necessary) the video again while responding. Participants then answered the remaining questions without viewing the clip again. Importantly, each participant was presented with only a single video clip and answered questions specifically about that one clip.

Figure 1

Screenshot from video capture of previous image description task.



Design

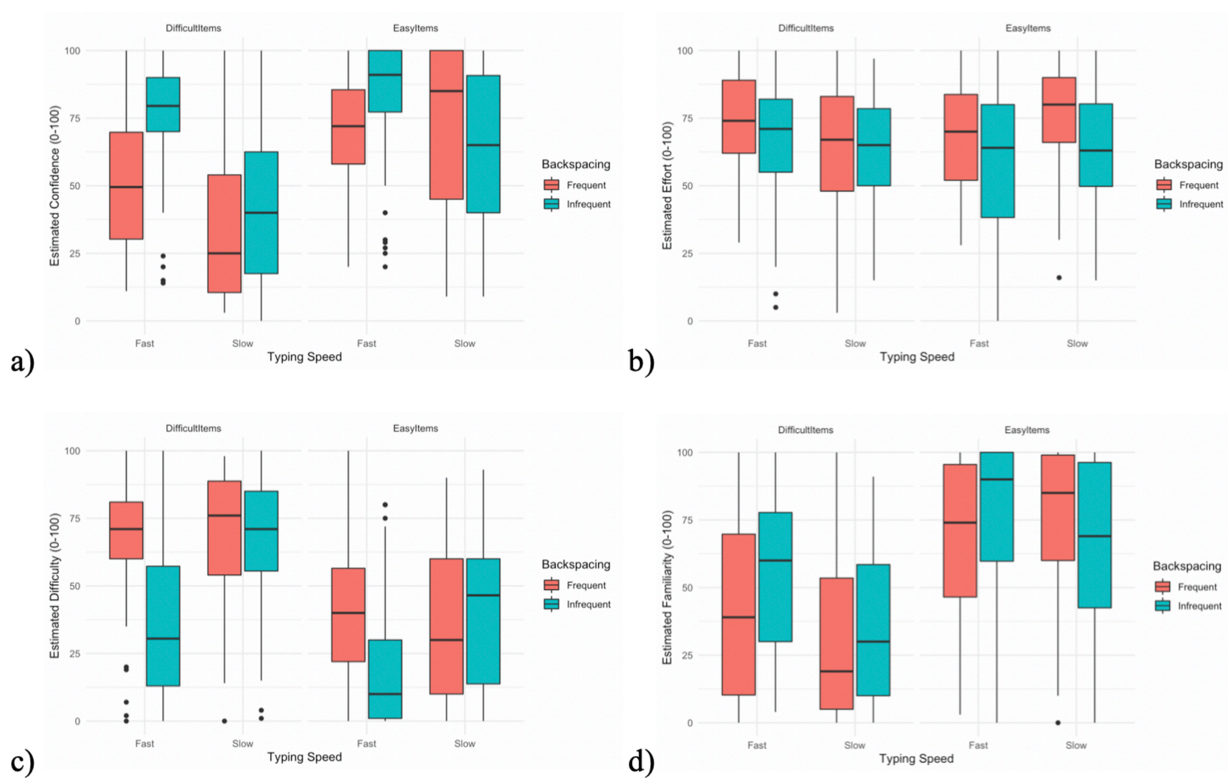
Each participant viewed and answered questions about one video clip showing a single item description. Across all 400 participants, from 4 to 16 (M=9.7) participants ended up answering questions about one of each of the 45 clips. The clips varied along four fully crossed independent variables: Item Category (tangrams, objects, faces), Item Difficulty (easy, hard), Typing Speed (fast, slow), and Backspace Frequency (frequent, infrequent). In the present analyses, we collapse across Item Category, considering instead only the effect of item difficulty, typing speed, and backspace frequency on participants' judgments.

Results

Here, we focus on the results of the first set of questions asking about participants' perceptions of the typist in terms of Confidence, Effort, Familiarity, and Difficulty. Figures 2a-d present boxplots for the responses to each of these scales, by item difficulty and typing characteristics. These responses for each scale were submitted to a separate mixed effect model, with Item Difficulty, Typing Speed, Backspace Frequency and their full interactions as contrast-coded fixed effects, and intercept terms only for subjects and video clips as random effects.

Figure 2

Boxplots of ratings of a) typist confidence, b) effort, c) task difficulty, d) familiarity, by item difficulty, typing speed, and backspace frequency. The dark bar within each box represents the median (50th percentile) and the upper/lower boundaries represent the interquartile range.



Confidence. Figure 2a presents the judgments of typist confidence. Descriptions for easy items were judged as more confident (M=71.99) than descriptions for difficult items (M=53.78), b=-0.62, SE=0.13, $p < .001$. Descriptions with fast typing were also judged as more confident (M=71.11) than descriptions with slow typing (M=51.69), b=-0.61, SE=0.13, $p<.001$), and descriptions with infrequent backspacing (M=66.15) were judged as more confident than descriptions with frequent backspacing (M=57.96), b=-0.38, SE=0.13, $p<.001$. There was also a significant Speed X Backspacing interaction, b=0.60, SE=0.27, $p<.02$. For descriptions with slower typing, backspace frequency did not affect confidence judgments, while for descriptions with faster typing, frequent backspacing was associated with perceptions of less confidence.

Effort. Figure 2b presents the judgments of the degree of effort made by the typists in producing their descriptions. Here, descriptions with more frequent backspacing were seen as more effortful (M=69.42) than descriptions with infrequent backspacing (M=62.68), b=0.28, SE=0.12, $p<.02$.

Difficulty. Figure 2c presents the judgments of how much difficulty the typists had in formulating their descriptions. Descriptions for easy items were judged as involving less difficulty (M=32.29) than descriptions for difficult items (M=56.94), b=0.80, SE=0.13, $p<.001$. Descriptions with fast typing were also judged as less difficult (M=38.10) than descriptions with slower typing (M=53.43), b=0.46, SE=0.14, $p<.001$, and descriptions with infrequent backspacing (M=66.15) were judged as involving less difficulty (M=40.11) than descriptions with frequent backspacing (M=51.39), b=0.45, SE=0.13, $p<.001$. There was also a significant Speed X Backspacing interaction, b=-0.82, SE=0.27, $p<.002$. For descriptions with slower typing, backspace frequency did not affect difficulty judgments, while for descriptions with faster typing, frequent backspacing was associated with perceptions of greater difficulty.

Familiarity. Figure 2d presents the judgments of how familiar the typists seemed with what they were describing. Here, descriptions for easy items were judged as involving greater familiarity (M=70.38) than descriptions for difficult items (M=42.19), b=-0.86, SE=0.15, $p<.001$. There was also an effect of typing speed, with fast typing associated with judgments of greater familiarity (M=60.98) than slower typing (M=49.95), b=-0.30, SE=0.15, $p<.04$.

Discussion

Our results suggest that viewers are sensitive to differences in typing patterns and can make judgments about a person's degree of proficiency with a written task based in part on how they type. In particular, both speed and error correction (backspacing) contributed to judgments of typist confidence and difficulty, while backspace frequency influenced judgments of effort and speed influenced judgments of familiarity. Overall, judgments about the typist were influenced by characteristics of the "delivery" of the visible descriptions. This suggests that when access to other cues is limited, as they often are in online collaborative contexts, people may be able to adapt by attending to subtler sources of information to draw inferences about how their partner is managing the task.

As remote online platforms become increasingly prevalent in collaborative work, it is important to develop a better understanding of the mechanisms that shape communication within these contexts. To adapt concepts that have been well addressed in the literature on face-to-face collaboration, we need a clearer view of the factors that influence communicative behaviors when many of the typical features of face-to-face interaction are missing. Without access to gesture, facial expression, tone of voice, or physical surroundings, users must rely on other sorts of cues to facilitate collaboration. Here, we have begun to examine how typing patterns can inform representations of an online partner. This sets the stage for exploring how similar behaviors may shape interactions in real-time collaborative writing contexts.

References

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10-63). Cambridge, England: Cambridge University Press.

Elliott, A., & Horton, W. S. (2019, November). *Typing fluency as a visible indicator of uncertainty in written production*. Poster presented at the 60th annual meeting of the Psychonomic Society. Montreal, CA.

Hinde, R. A. (Ed.). (1972). *Non-verbal communication*. Cambridge University Press.

Lea, M., & Spears, R. (1992). Paralanguage and social perception in computer-mediated communication. *Journal of Organizational Computing*, *2*, 321–341.

Tidwell, L. C., & Walther, J. B. (2002). Computer-mediated communication effects on disclosure, impressions, and interpersonal evaluations: Getting to know one another a bit at a time. *Human Communication Research*, *28*, 317-348.