

Eliciting Gestures for Novel Note-taking Interactions

Katy Ilonka Gero*
katy@cs.columbia.edu
Columbia University
New York City, NY, USA

Chris Melancon
cmelancon@google.com
Google Research
Portland, Oregon, USA

Lydia B. Chilton
chilton@cs.columbia.edu
Columbia University
New York City, NY, USA

Mike Cleron
mcleron@google.com
Google Research
Mountain View, California, USA

ABSTRACT

Handwriting recognition is improving in leaps and bounds, and this opens up new opportunities for stylus-based interactions. In particular, note-taking applications can become a more intelligent user interface, incorporating new features like autocomplete and integrated search. In this work we ran a gesture elicitation study, asking 21 participants to imagine how they would interact with an imaginary, intelligent note-taking application. Participants were prompted to produce gestures for common actions such as select and delete, as well as less common actions (for gesture interaction) such as autocomplete accept/reject, ‘hide’, and search. We report agreement on the elicited gestures, finding that while existing interactions are prevalent (like double taps and long presses) a number of more novel interactions (like dragging selected items to hotspots or using annotations) were also well-represented. We discuss the mental models participants drew on when explaining their gestures and what kind of feedback users might need to move to more stylus-centric interactions.

CCS CONCEPTS

• **Human-centered computing** → **Touch screens; Empirical studies in HCI.**

KEYWORDS

elicitation study, note-taking applications, intelligent user interfaces

ACM Reference Format:

Katy Ilonka Gero, Lydia B. Chilton, Chris Melancon, and Mike Cleron. 2018. Eliciting Gestures for Novel Note-taking Interactions. *ACM Trans. Graph.*

*Research performed while an intern at Google Research.

Authors’ addresses: Katy Ilonka Gero, katy@cs.columbia.edu, Columbia University, New York City, NY, USA; Lydia B. Chilton, chilton@cs.columbia.edu, Columbia University, New York City, NY, USA; Chris Melancon, cmelancon@google.com, Google Research, Portland, Oregon, USA; Mike Cleron, mcleron@google.com, Google Research, Mountain View, California, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](https://permissions.acm.org).

© 2018 Association for Computing Machinery.

0730-0301/2018/8-ART111 \$15.00

<https://doi.org/10.1145/1122445.1122456>

37, 4, Article 111 (August 2018), 10 pages. <https://doi.org/10.1145/1122445.1122456>

1 INTRODUCTION

As handwriting and hand-drawn figure recognition improves [2, 12, 26], we see new opportunities for stylus-based interactions. Features like autocomplete and integrated search, previously confined to text-based computing environments, may soon be implemented in handwriting-based environments. People report that using a pen still affords the most high precision, “in the flow” interactions, even in the face of the variety of computing mediums available (including digital pens) [13], and so in this work we consider how note-taking applications can become a more intelligent user interface, automatically responding to users’ pen actions without requiring as many modes or menus. We envision future note-taking applications to be more like the word processor in terms of how it departed from the typewriter: future note-taking applications will make interacting with handwritten, nonlinear notes – not just words, but also diagrams and sketches – as easy as interacting with typed text is currently in modern word processors.

While studying gestures for surface computing has a long history in human computer interaction [6, 21, 27], less attention has been paid to gestural interactions for general, stylus-based note-taking activity. Current work tends to focus on specific applications that are complementary to keyboard-centric digital activities, like handwritten annotations on typed documents [14, 16, 23, 24], or drawing and note-taking as part of data analysis [5, 22]. But general purpose, handwritten notes remain a mainstay in our lives, despite all the computing mediums available [13].

Our work focuses on entirely handwritten note-taking as a highly impactful use case of surface computing. Handwritten notes, and the physical embodiment that they afford, have been shown to improve the retention of lecture material [8, 10], be a critical component of the design process [15], as well as allow for the benefits of nonlinear thinking [7]. Making handwritten notes more interactive, by giving them the editing capabilities that computers allow, may allow handwritten notes to take on the power currently only seen when working with 3D, physical objects, which has been shown to increase people’s causal reasoning [18] and improve their ability to solve insight problems [17].

In this work, we seek to understand how users envision interacting with an intelligent note-taking application where handwritten notes are a first-order object and users are able to interact with

‘digital ink’ without always having to revert to menu options or a keyboard. We ran an elicitation study, asking 21 participants to imagine how they would perform eight unique actions. We found that while some actions have a very high level of agreement, others are quite distributed, with participants drawing on a variety of mental models to imagine how to perform the novel action.

Our contributions are:

- elicited gestures for eight note-taking actions, including autocomplete accept/reject, and their level of agreement;
- a description of three mental models that participants drew on, including the legacy word processor model; and
- a discussion of how feedback might guide users’ understanding of digital ink affordances.

2 RELATED WORKS

2.1 Gesture Elicitation Studies

Gestures for surface computing have a long history in HCI, with foundational issues like motor control complexity, visual and auditory feedback, and memorability being well studied [27]. Elicitation studies for gestures, or ‘user-defined gestures’, are often used to understand how people currently think about surfacing computing while simultaneously aiding in the design of gestures for novel interactions [16, 21]. Nacenta et al. [11] found that user-defined gestures are easier to remember and that participants prefer them, though in commercial products gestures tend to be pre-defined (though informed by much user testing, and users may have options for customization).

One critique of the gesture elicitation study methodology is often termed ‘legacy bias’, where participants’ proposed gestures are heavily biased by their experience with existing computing interactions. Several methods have been proposed to reduce legacy bias, including having participants produce more than one gesture per referent [9]. Though this technique has seen mixed results in reducing legacy bias [20], we implement a version of it in this study that also results in qualitative data about participants’ mental models.

2.2 Digital Note-taking

As the use of touchscreen devices increases, primarily smartphones but increasingly larger tablets, surface computing interactions have become much more common, opening up research for new applications like collaborative document reviewing [24, 25] and creating data visualizations [4, 5]. Many of these areas use elicitation studies to understand how users want and expect to perform new actions.

Simultaneously, improvements in not just handwriting recognition [2, 12, 26] but also handwriting generation [1, 3] have opened up new opportunities for handwriting-based surface computing. Handwritten notes are an incredibly varied and important medium. Riche et al. [13] did an extensive diary study of analog pen use, and reported 9 affordances of pens, including externalizing thoughts (often to remember things), producing high-fidelity marks (indicating the precision and fluidity of expression pens provide), and automatic usage (people often feel like writing with a pen keeps them “in the flow”). All of these speak to the importance people still find in pen-based expression over all the computing mediums available to them.

Very recently, some tablet-based note-taking applications have begun to support pen gestures for a few actions. While many note-taking applications support circling to select items, striking through to highlight (when in highlighting mode), and various taps to bring up menu options, both Nebo and MS Journal have recently introduced:¹

- crossing out or scribbling to delete text (MS Journal, Nebo)
- introducing a space between text by drawing a vertical line (MS Journal, Nebo)
- joining words by drawing a vertical line upwards (Nebo) or a curve between two words (MS Journal)
- inserting a new line by drawing a backwards ‘L’ (MS Journal)

Though the adoption of styluses and their related applications is quite recent, we expect to see an increase in their usage as technology improves and their affordances approach those found in analog pens. For instance, features previously only seen in text-based environments, like autocomplete and integrated search, can now be implemented in handwriting-based environments. In this work we study how users imagine interacting with such an intelligent user interface, in order to aid in the improvements of these technologies.

3 METHODOLOGY

3.1 Study Design

We recruited 21 participants for the study from the USA, and the entire study was run in English. All participants had to have used a tablet for at least 6 months and own a stylus. We recruited participants from a variety of age ranges, as well as a variety of tablet types and operating systems. The demographics of our participants can be found in Figure 1.²

Due to the covid-19 pandemic, our study took place remotely. Participants joined a video conference from a computer and from their tablet. They shared the screen of their tablet, and pointed the webcam of their computer at their tablet such that we were able to see their hands when they interacted with the tablet. An example screenshot of the study setup, with all personally identifiable information removed, can be found in Figure 2.

Participants were told that the study was about how people wanted to interact with tablet-based note-taking applications, and was not a test of their knowledge of existing interactions. They were told that they would be shown a series of still images that they should imagine are part of a note-taking application – any text in these images should be assumed to have been handwritten. For each image they would be asked how they would take some action on the screen, and they would be asked to show both a stylus and a finger interaction. They were then taken through an example elicitation exercise, where they were shown some words on a screen and asked how they would move these words downward.

¹This is as of April 2022, and we refer to documentation at <https://developer.myscript.com/docs/concepts/editing-gestures/> for Nebo and <https://support.microsoft.com/en-us/office/edit-your-document-with-natural-gestures-7edbcf8e-0004-484d-9b62-501a31c23ee9> for MS Journal.

²We also asked participants about their experience with different note-taking apps, which was varied and contained subtle differences, e.g. a participant might use a note-taking application primarily for taking notes on pdfs or primarily for keeping to-do lists, and either could be done with or without a stylus. While most used one or more note-taking apps, few used them primarily with their stylus, and none mentioned using Nebo or MS Journal.

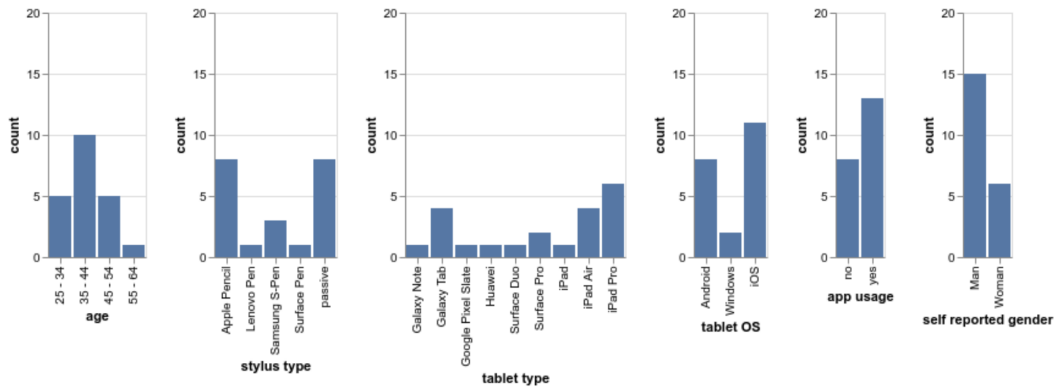


Figure 1: Histograms of all the demographic data we collected about the 21 participants in our study. We are able to recruit a wide range of tablet types and stylus types. Note that ‘app usage’ refers to whether or not a participant reported using a stylus-based note taking app.

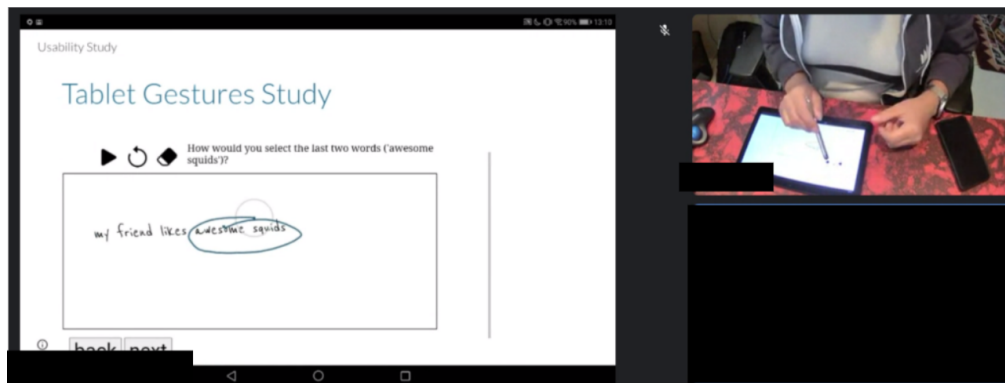


Figure 2: A screenshot of the study set up as it appeared to the facilitator. All personally identifiable information has been redacted. On the left the participant is screen sharing from their tablet, and on the right we had a camera pointing at their tablet such that we could see their hands.

When participants interacted with the image, touch events were recorded by leaving a trail – in Figure 2 you can see that as the participant circles the words, a mark of where their stylus has been is shown. The beginning of a touch event is indicated by a grey circle. Participants were told that this was part of the data collection process, and did not indicate that they were drawing on screen. The additional webcam pointed at their tablet allowed us to determine if they were using a stylus or a finger.

Participants were then led through the actions listed in Table 1 by the facilitator. Four of these actions (‘select two words’, ‘select from phrase’, ‘copy’, and ‘delete’) were chosen as basic actions currently supported in most stylus-based note-taking applications.³ The remaining four were selected as potential new features that improvements in handwriting recognition/generation and tablet-based software may be able to support in the future.

Participants were encouraged to think aloud, which most participants did successfully. When a participant didn’t explain or

verbally think through a gesture that the facilitator found confusing or particularly unique, the facilitator would ask the participant to explain their reasoning or what they were thinking of when they did that gesture. Also if a participant exclusively indicated they would use a menu option or a keyboard, the facilitator would prompt them to provide an interaction that did not include menu options or keyboards. This enabled us to avoid issues in elicitation studies where participants rely solely on existing mental models, by first letting participants respond in their most intuitive way, and then asking them to continue thinking of other interactions. This methodology has been suggested by Morris et al. [9]. Sometimes participants were unable to think of alternate interactions, even when prompted and given time, in which case the facilitator would allow them to move on. For these reasons, our data is less clean than in some elicitation studies – for instance not all participants did the same number gestures. However, we found that this more conversational approach encouraged participants to think outside of menu interactions, and allowed us to gather rich qualitative data about participants’ mental models.

³In particular we considered GoodNotes, OneNote, Squid, and Notability.

action	prompt
select two words	How would you select these two words?
select from phrase	How would you select the last two words ('awesome squids')?
delete	How would you delete these two words?
hide	How would you 'hide' or 'collapse' these bullet points?
autocomplete accept	How would you accept the suggested text completion?
autocomplete reject	How would you reject the suggested text completion?
copy	How you could copy 'research squids' so you could paste it elsewhere?
search	How would you search google for the last two words, 'sea snails'?

Table 1: Actions and the text prompt provided to participants. Each action had a still image associated with it – ‘select two words’ showed just two words on the screen, while ‘select from phrase’ showed those two words embedded in a short phrase. ‘hide’ showed a phrase with some bullet points beneath it, ‘copy’ showed a list of short phrases, and ‘search’ showed a phrase.

action	num codes	annotator agreement
select two words	8	.87
select from phrase	8	.85
delete	9	.84
copy	10	.84
hide	8	.94
search	7	.78
autocomplete accept	11	.83
autocomplete reject	11	.75

Table 2: The number of codes in each codebook, where there was one codebook per action, and the inter-annotator agreement reported as Cohen’s Kappa. All prompts had a very high level of agreement.

3.2 Data Coding

The raw result of our study was 21 videos of participants responding to the action prompts listed in Table 1. These videos were then split into individual action responses such that each clip contains a single participant responding to a single action prompt. The clips were then arranged into ‘supercuts’ such that a researcher could view all of the responses to a single action prompt. We then proceeded to code the participants’ gestures.

Some of the gestures also included a select action. For instance, for the ‘copy’ action, participants often first selected the text and then did a gesture for copy. For this reason, some actions are also coded with a select code, in addition to the main code for that action. For each gesture we also code the pointer used: ‘stylus’ ‘finger’, or ‘both’.

Two researchers coded each action. First they each watched the relevant supercut alone, and took note of common gestures. Then they came together and developed a codebook which would allow them to code the most common gestures. Then they each watched the supercut again and coded each gesture. Finally an inter-annotator agreement score was calculated and disagreements were resolved with a third party. The inter-annotator agreement scores (Cohen’s kappa) and number of gestures in each codebook can be found in Table 2. Additionally, if an action also included

a select gesture, all researchers used same ‘select’ codebook for consistency.

Because different researchers created the codebooks for different actions, the codebooks had different levels of specificity. For instance, ‘copy’ had three different codes that included some kind of tap gesture (‘long press’, ‘double tap’, ‘triple tap’) while search had none. In order to align the codes, two researchers created an ‘umbrella’ codebook, where each code from all the codebooks was assigned a high-level ‘umbrella’ code. This allowed us to compare all the actions while having more fine-grained coding data (the original codebooks and codes) to refer back to.

4 RESULTS

Figure 3 shows the prevalence of all gestures coded for each action. In particular, this figure shows the prevalence of the first non-menu gesture performed by participants. We report eight umbrella codes – ‘annotation’, ‘automatic’, ‘circle’, ‘drag’, ‘line’, ‘mark’, ‘tap’, ‘write’, plus an ‘other’ category. Examples for each code can be found in Figure 4a, and an example of the most prevalent gesture for each action can be found in Figure 4b.

In Figure 3 we also report the level of agreement, A , for each action. We calculate the agreement level as

$$A_a = \sum_i \left(\frac{|P_i|}{|P_a|} \right)^2$$

where a is a given action we studied, P_a is the set of all gestures performed for this action, and P_i is a set of gestures with identical codes. In the case of gestures in the ‘other’ category, we count each gesture as in its own category.⁴ This is in line with Wobbrock et al. [21]. We calculate the agreement based on the umbrella codes. Vatavu [19] notes that this agreement score is a function of subjective decisions by the researchers about how to group gestures, and we found that to be the case in our study as well. For instance, as reported the agreement score for ‘copy’ is 0.36 – relatively high. But if we break the ‘tap’ gestures into ‘long press’, ‘double tap’, ‘triple tap’, the score drops to 0.165. For this reason, we report the agreement in the context of the full prevalence data – i.e. the histograms in Figure 3 – as well as more in-depth detail in the text of the paper. We also report on how the select gesture correlates

⁴e.g. If there are 20 gestures, and 12 are ‘line’, 5 are ‘circle’, and 3 are ‘other’, then the agreement would be calculated as $(12/20)^2 + (5/20)^2 + (1/20)^2 + (1/20)^2 + (1/20)^2$.

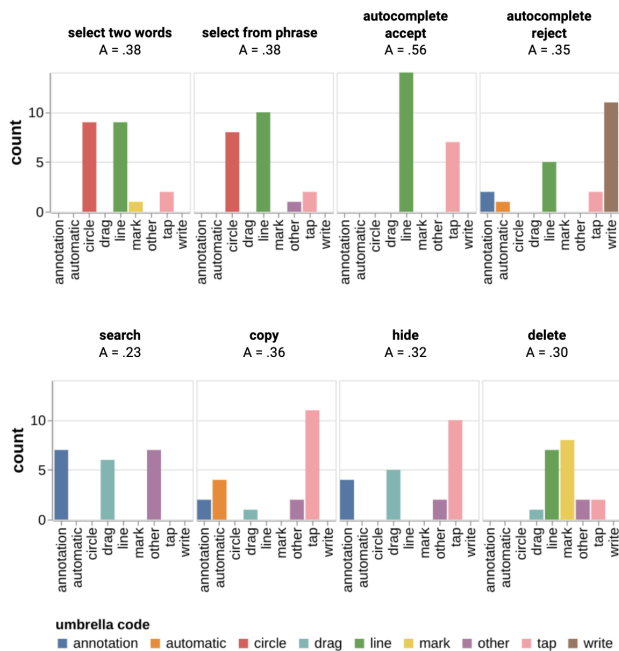


Figure 3: Here we show, for each action, the prevalence of the gesture types the participants performed, along with the level of agreement calculated for that action. While some actions have a clear consensus on a single gesture, for instance ‘line’ for ‘autocomplete accept’, many actions had two primary gestures, or for instance in the case of ‘search’, an even distribution across three gestures.

with different participant attributes, as well as qualitative data on the mental models that participants drew on when planning and explaining their gestures.

4.1 Elicited Gestures

4.1.1 Select. The select gestures saw a relatively high level of an agreement, with most participants either selecting by circling the text or by swiping over the text in a straight line. Sometimes participants swiped through the text, for instance as one might expect if highlighting or striking through, while others swiped at a location that may be considered more like an underline. Some participants noted that when text was small it could be difficult for them to accurately strike through it, so underlines may have intended to be more like highlights. Since the difference between these was often ambiguous, and for this reason both as coded as ‘line’. While there were other gestures – like tapping on words, or drawing the extents of an imagined selection box – overall ‘select’ saw a very high level of agreement, albeit across two different gestures. In subsection 4.2 we report on the select gestures used as part of other actions, and how consistent participants were in these gestures.

4.1.2 Autocomplete Accept and Reject. Autocomplete accept had the highest level of agreement of all the actions we studied, with the majority of participants drawing a line from left to right through

the suggested text to indicate accept and minority of participants tapping on the suggested text to indicate accept. Participants talked about familiarity with this action, for instance when typing text messages or emails, and referred to the ways they would perform this in other applications.

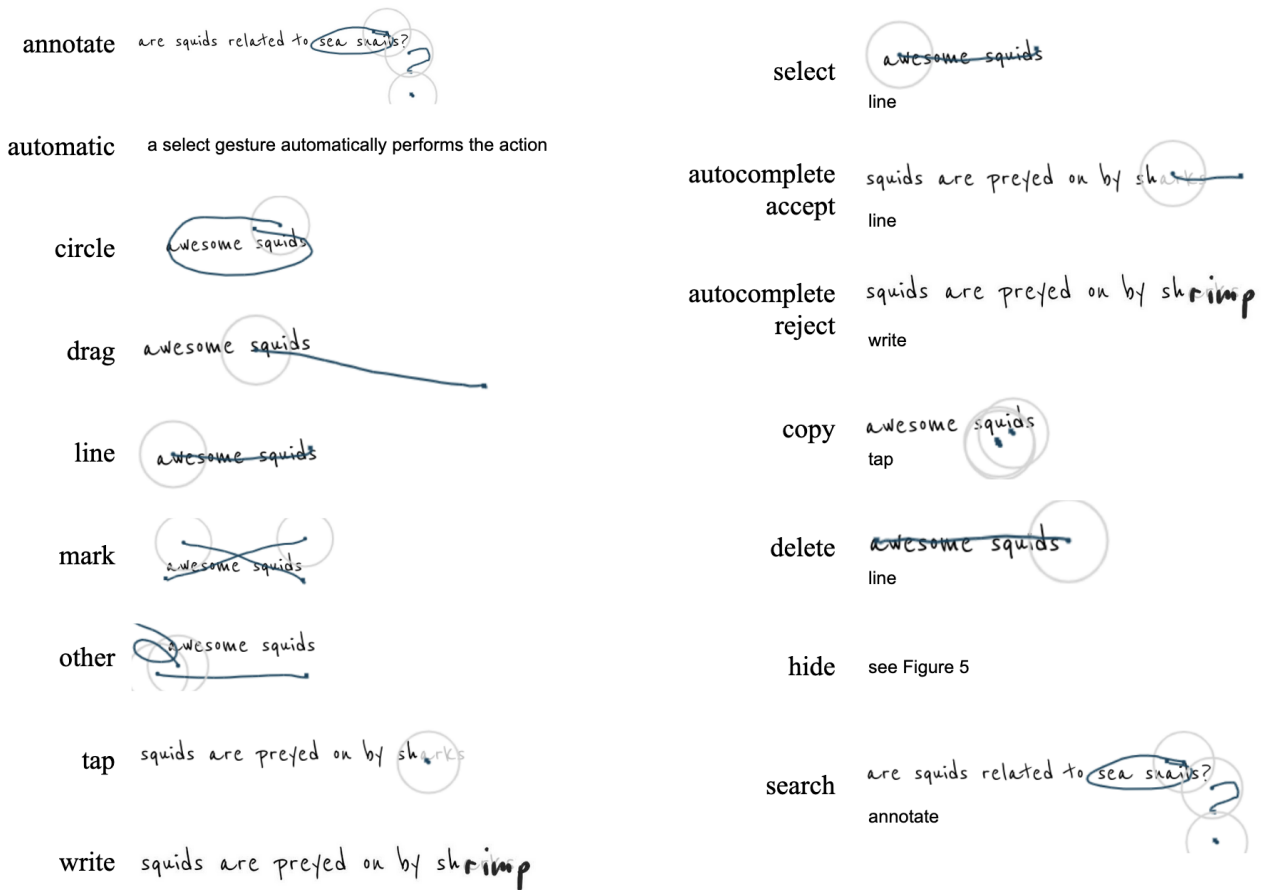
Autocomplete reject had a reasonable consensus around ‘write’, in which they would write over the suggested text. Again, participants referred to their experience with autocomplete in other contexts, where if they simply continued typing then the suggestion would disappear. However, we did see some participants use a line gesture, where the line would be drawn from right to left, which was also very common in the delete gestures.

4.1.3 Search. Search, in which participants were asked how they would perform a web search for text they had written from within the application, saw an even distribution across three different gesture categories, one of which was ‘other’. For this reason we consider search to have a fairly low level of consensus. The two common gestures that we did see were ‘annotation’, where participants would draw something like a question mark or a ‘g’ (for Google) to initiate a search on selected text, and ‘drag’, where the participant would drag selected text to a special region on the page. In the ‘other’ category we saw gestures such as shaking the selected text or highlighting text with a special web search marker. But overall participants talked about how, while they were familiar with searching for text from various applications, they were unsure how to do this without moving to a browser.

4.1.4 Copy. The most prevalent gesture for performing a copy was some kind of tap gesture. Participants did a variety of different kinds of tap gestures, whether it be a double or triple tap, or a long press. Participants who used a long press noted that they may expect different actions to be performed at different lengths of time, for instance a short press indicating select while a one or two second press indicating copy, and an even longer press could perform some other actions. Several participants mentioned how they were aware that the various tap gestures were often overloaded, in that tap gestures already had a lot of actions associated with them. Another prevalent copy gesture was ‘automatic’, in which any select gesture automatically resulted in a copy. This, combined with the prevalence of tap gestures, suggests that participants considered copy to be a common action that requires a simple, common gesture.

4.1.5 Hide. For this action, participants were asked how they would hide or collapse a set of bullet points underneath a title. Figure 5 shows this prompt as it was presented to participants. In pilot testing we found the some people had some difficulty understanding this concept, but in the final study participants seemed very confident in understanding this action, referring to a variety of different situations that had similar actions, like expanding and collapsing text in a wiki or in the note-taking application Notion, opening and closing ‘drawers’, or hiding columns or rows in spreadsheets.

Many participants first responded by saying that for this kind of action that there must be some kind of icon they could tap to hide or show the content, for instance a plus icon or a caret. In this case, we prompted participants to consider how they might define what content they wanted to hide.



(a) Illustration of each code we used to code all of the actions studied.

(b) Most prevalent gesture performed by participants for each action studied.

Figure 4: Illustrations of various gestures. These illustrations are recreations of what we saw participants perform. Traces of a touch event are in the blue ink, and the gray circles represent the beginning of a touch event.

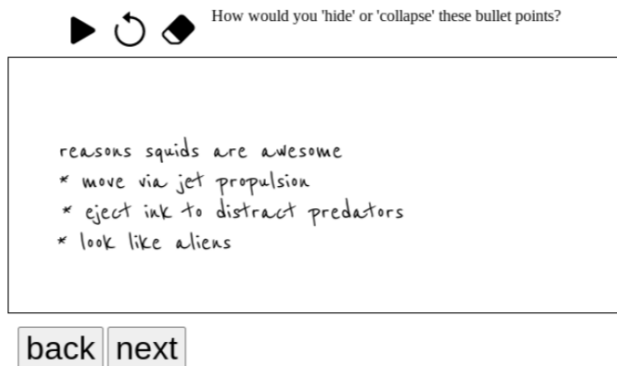


Figure 5: Prompt for the 'hide' action as it was shown to participants. Participants often tapped on the top line ('header') or bullet points ('content') in order to perform the hide action.

The most prevalent gesture for this action was some kind of tapping gesture, though some participants tapped on the title and others tapped on the content they were trying to hide. Another common gesture was to drag the content onto the title, which several participants mentioned that the way the content was created would be key to allowing them to hide it later. We come back to this idea of the implications object creation and representation in the discussion section.

4.1.6 Delete. Delete did not have much consensus about the gesture to be used. Many participants make some kind of obscuring mark over the text, whether it be scribbling over the text or drawing an X or something similar. Figure 6 shows a variety of marks participants made to indicate delete. Many participants wanted to strike through the text from right to left, and noted that the directionality was important as they expected swiping from left to right would indicate select. These participants often mentioned how striking from right to left evoked the way a cursor would move from right to left when they hit the delete button on a keyboard.



Figure 6: Examples of various ways participants used a ‘mark’ to indicate they wanted to delete the text.

Though there was much variation in the delete gestures, there was clearly a common theme of somehow obscuring the text.

4.2 Select Gesture Correlates

Several of the actions we studied could reasonably require a select gesture prior to a gesture specific to the action. For instance, to copy text participants often first selected the text with one gesture and then performed another gesture to initiate a copy. This allowed us to look at the consistency of participants’ select gestures across different actions, and investigate if participants’ select gestures had any correlation with the demographic information we collected.

Figure 7a shows the prevalence of the different select gestures used broken down by participant. Participants are grouped by the type of select gesture they used. Five participants consistently used the circle gesture to select, or used no select gesture at all (‘none’). (A participant might not use a select gesture, for instance, when deleting.) Six participants consistently used a swipe (or ‘line’) gesture, and the remaining were not consistent in the select gesture they used.

We investigated if any of the demographic data we collected had any correlation with the gesture a participant was likely to use. Only previous experience with a note-taking application showed any difference. Figure 7b shows this difference: participants who didn’t report using stylus-based note-taking applications were far more likely to use the swipe gesture for select. This doesn’t mean that participants who use a note-taking app never used a swipe gesture, but rather that they also relied heavily on circling. We investigate potential reasons for this in the next section.

4.3 Mental Models

One value of running this study with a facilitator was that we were able to collect qualitative data about how participants were thinking about their gestures. This allowed us to investigate the mental models and other priors participants were bringing to the elicitation study.

4.3.1 Word processor. Many participants were very attached to the idea of a cursor and keyboard, even when reminded to imagine that all text had been handwritten. For instance, P10 said, “I would long press and then select. You have to take the cursor from the front to the back.” and P4 said, “I would put the cursor here. And then if there was a gesture for a return or something.” These quotes are emblematic of a larger theme where participants expected even handwritten text to have the affordances of a word processor.

This possibly explains participants’ preferences for a swipe select gesture versus a circling one as seen in Figure 7b. Participants with less experience with stylus-based note-taking apps may be drawing more heavily on a word processor mental model, where swiping over text is an analogous to dragging a cursor over text. We come

back to implications of the word processor mental model in the discussion.

4.3.2 Annotations as commands. Some gestures involve direct manipulation, and are analogous to interacting with physical objects, like dragging or tapping on text. But there were many gestures that were more like annotations, where one might imagine an assistant interpreting and acting on the annotations. Figure 8 shows a variety of examples of gestures that participants’ performed that were coded as ‘annotate’.

Interpreting annotations requires distinguishing between types of ink: ink that is drawn as part of a note, and ink meant to be an instruction. We might consider this ‘permanent’ versus ‘transient’ ink. Distinguishing between these types of ink may prove difficult because it relies on understanding the users’ intention. Consider a scribbling gesture for delete. How does an intelligent user interface differentiate between a user wanting to remove ink from the screen by scribbling over it versus a user who wants to scribble over ink on the screen but doesn’t actually want the ink to disappear?

4.3.3 Creation priors. Some participants were sensitive to how ink was created. For instance, P1 used the example of a mindmap: if they drew a connecting line between two words, at that point they would know if the line was ‘attached’ to the words or just stray ink. This would change how they would expect the ink to behave in later actions. Similarly, when considering select gestures, P7 said “I might double tap on it because sometimes that’ll make a selection depending on how the word was created.”

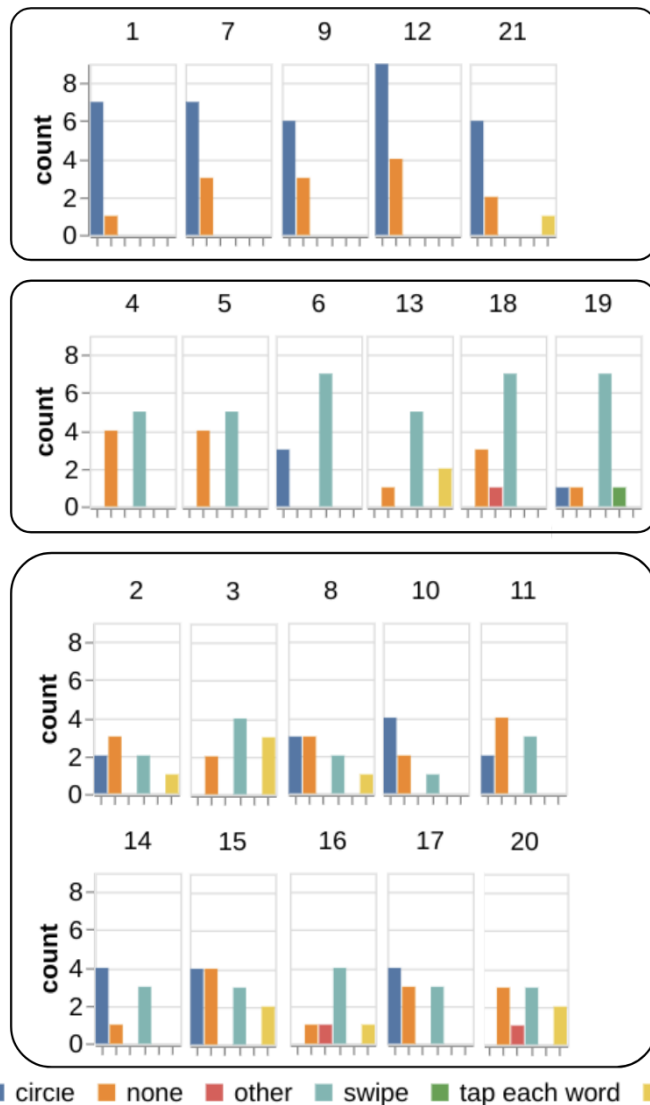
These examples show us how participants expect ink on the screen to have different affordances depending on what occurred when that ink was created. Given that in our study participants were shown ink on the screen instead of being asked to produce this ink themselves, this led to some uncertainty in the participants about how they might be able to perform different actions. We investigate the kind of feedback the participants might have been expecting in order to understand the affordances the ink would have in the discussion section.

5 DISCUSSION

5.1 The Word Processor Legacy, and Other Legacies to Consider

It is unclear if the reliance on the word processor mental model was somehow an artifact of our study, despite our efforts to have participants imagine a non-word processor environment, or if the word processor mental model is simply an extremely prevalent prior. It could be due to the actions we studied, most of which have clear, word processor counterparts. But it seems equally likely that the word processor mental model has pervaded our sense of what computing is. Much computing is simply a more static version of a word processor – typed text laid out linearly that can be highlighted and copied.

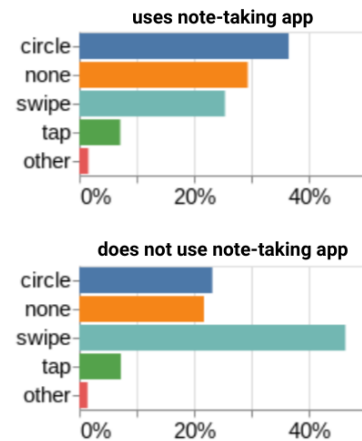
We envision a future of computing that is much more dynamic. Paper has many affordances that the traditional word processor does not, and we believe that future computing interfaces will have some of those affordances. Moving to the more freeform nature of pen and paper will require users to let go of the word processor model, in order to allow for more non-linear affordances. The use



circlers

swipers

varied



% of all select gestures done by those participants

(b) Breakdown of select gestures by whether or not a participant reported using a stylus-native note-taking application. Participants who did not report using such an application were much more likely to swipe to select as compared to their counterparts.

(a) Histograms for each participant's use of select gestures, grouped manually by their observed preferences. While some participants had clear preferences for a 'circle' gesture, or a 'swipe', most participants used a variety of gestures throughout the study to select ink.

Figure 7: Because some of the actions we used to prompt participants could involve a select gesture, we were able to look at participants preferences for select gestures across all actions. Participants' preferences for select gestures may be explained by whether they had experience with existing stylus-native note-taking applications.

of swiping gestures to navigate mobile devices may be a precursor (no pun intended) to this.

One participant mentioned his experience as a proofreader many years ago as guiding some of his responses. And we can see in the introduction of ink gestures in Nebo and MS Journal (as documented in the Related Works section) a similar reference to proofreading marks. Though there is no one definitive set of proofreading marks,⁵

we may see a resurgence of them as digital mediums shift more towards pen interactions. While proofreading marks are typically about editing text, in this study we were looking at new, digitally native interactions like search and copy. We did see, for instance in search, the use of proofreading-esque marks, suggesting the possibility that such marks may be able to transition into digitally-native interactions.

⁵For instance there are slight differences between those documented in the Chicago Manual of Style (see <https://www.chicagomanualofstyle.org/help-tools/proofreading->

https://www.wiedenhof.nl/ul/odwmtr/proofreaders'_marks.pdf) and those in Merriam Webster (see https://www.wiedenhof.nl/ul/odwmtr/proofreaders'_marks.pdf) and they vary even more across languages, as to be expected.

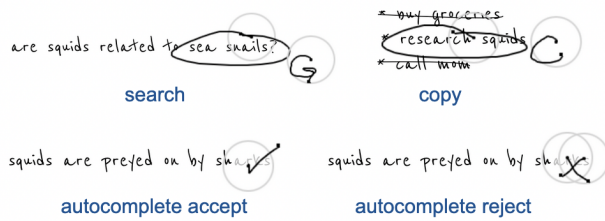


Figure 8: Examples of the various ways participants used annotations to indicate what action they wanted to perform.

5.2 Feedback at Time of Ink Creation

Several participants mentioned the kind of feedback they'd expect from the application, to confirm that it would perform their intended action. For instance, P13, when explaining their gesture for 'copy', said "Maybe use the stylus to underline or go across the words and then it shows a lift off the screen and I can just move it." Others commented that "depending on how it was created" they would expect different gestures to be available. While this is likely a mental model that participants are drawing on from existing applications, it is an important one to consider. When, for instance, is the application correctly interpreting their handwriting? Many participants commented that existing handwriting recognition technology does not work for their handwriting or that even though they have not tried handwriting recognition they imagine that because their handwriting is so poor it would likely not work for them. (On this note, we may note that almost all participants commented on the poor quality of their handwriting, which perhaps suggests that everyone believes their handwriting is below the mean.)

Investigating the kind of feedback users require at the time of ink creation will be an important part of designing intelligent note taking applications. For some actions the feedback is clear, for instance participants will know immediately if their 'delete' gesture deletes what they intended, because the ink will or will not disappear. Similarly with 'search', participants will either see search results appear or not. However for other actions like 'copy', or actions we didn't study but expect to be part of intelligent user interfaces of the future like creating freeform diagrams and mindmaps, it is less clear what kind of feedback will provide users with the necessary information but not distract them from the paper-like experience we expect these applications to want to mimic.

5.3 Limitations and Future Work

In this study, participants were asked to imagine how to do something, sometimes against their instincts – even if they wanted to use a menu option, they were prompted to imagine another way to perform the action. Sometimes they noted a given gesture might not be the best, but it was all they could think of. Some studies, like Wobbrock et al. [21], ask participants to rank how well a gesture fits with its action, and how easy it is to perform a gesture. Others, like Nacenta et al. [11], look at the memorability of new gestures. Future work should investigate the fit and ease of novel gestures, as how likely users are to remember these gestures after being introduced to them.

Additionally, it is important to consider the context in which this elicitation study occurs. The affordances styluses have and the state of note-taking applications are constantly changing. This study represents a moment in time, and a particular, American- and Anglo-centric cultural context. This is not to devalue elicitation studies, nor the elicitation study we present in this paper. Rather, we acknowledge the limits of the study as being extremely contextual. Elicitation studies should be replicated and re-run in order to understand how people's priors vary across cultures and time.

6 CONCLUSION

In this work we studied the gestures users envision performing to enact new actions for an imaginary, intelligent note-taking application. We studied both baseline actions like select and copy as well as novel actions like accepting a handwritten autocomplete and performing an integrated search on handwritten text. Through an elicitation study we found that participants had high agreement for the select and autocomplete accept/reject actions, and much lower agreement on the actions search, copy, hide, and delete. We found that participants with experience using digital note-taking apps are more likely to select by circling text (rather than 'highlighting' it) and that many participants still draw on a word processor mental model even when interacting with handwritten digital notes. We discuss how we envision a future of computing that moves away from the word processor mental model and towards the freeform nature of analog pen and paper, where stylus-centric interactions can help people quickly and easily trigger rich writing features.

REFERENCES

- [1] Emre Aksan, Fabrizio Pece, and Otmar Hilliges. 2018. Deepwriting: Making digital ink editable via deep generative modeling. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [2] Kartik Dutta, Praveen Krishnan, Minesh Mathew, and CV Jawahar. 2018. Improving cnn-rnn hybrid networks for handwriting recognition. In *2018 16th international conference on frontiers in handwriting recognition (ICFHR)*. IEEE, 80–85.
- [3] Sharon Fogel, Hadar Averbuch-Elor, Sarel Cohen, Shai Mazor, and Roei Litman. 2020. Scrabblegan: Semi-supervised varying length handwritten text generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4324–4333.
- [4] Nam Wook Kim, Nathalie Henry Riche, Benjamin Bach, Guanpeng Xu, Matthew Brehmer, Ken Hinckley, Michel Pahud, Haijun Xia, Michael J McGuffin, and Hanspeter Pfister. 2019. Datatoon: Drawing dynamic network comics with pen-touch interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.
- [5] Yea-Seul Kim, Nathalie Henry Riche, Bongshin Lee, Matthew Brehmer, Michel Pahud, Ken Hinckley, and Jessica Hullman. 2019. Inking your insights: Investigating digital externalization behaviors during data analysis. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*. 255–267.
- [6] Chunyuan Liao, Francois Guimbretiere, and Ken Hinckley. 2005. PapierCraft: a command system for interactive paper. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*. 241–244.
- [7] Tamas Makany, Jonathan Kemp, and Itiel E Dror. 2009. Optimising the use of note-taking as an external cognitive aid for increasing learning. *British Journal of Educational Technology* 40, 4 (2009), 619–635.
- [8] Kayla Morehead, John Dunlosky, and Katherine A Rawson. 2019. How much mightier is the pen than the keyboard for note-taking? A replication and extension of Mueller and Oppenheimer. *Educational Psychology Review* 31, 3 (2019), 753–780.
- [9] Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, MC Schraefel, and Jacob O Wobbrock. 2014. Reducing legacy bias in gesture elicitation studies. *interactions* 21, 3 (2014), 40–45.
- [10] Pam A Mueller and Daniel M Oppenheimer. 2014. The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. *Psychological science* 25, 6 (2014), 1159–1168.
- [11] Miguel A Nacenta, Yemliha Kamber, Yizhou Qiang, and Per Ola Kristensson. 2013. Memorability of pre-designed and user-defined gesture sets. In *Proceedings of the*

- SIGCHI conference on human factors in computing systems*. 1099–1108.
- [12] Raymond Ptucha, Felipe Petroski Such, Suhas Pillai, Frank Brockler, Vatsala Singh, and Paul Hutkowsky. 2019. Intelligent character recognition using fully convolutional neural networks. *Pattern recognition* 88 (2019), 604–613.
- [13] Yann Riche, Nathalie Henry Riche, Ken Hinckley, Sheri Panabaker, Sarah Fuelling, and Sarah Williams. 2017. As We May Ink?: Learning from Everyday Analog Pen Use to Improve Digital Ink Experiences.. In *CHI*. 3241–3253.
- [14] Hugo Romat, Emmanuel Pietriga, Nathalie Henry-Riche, Ken Hinckley, and Caroline Appert. 2019. SpaceInk: Making Space for In-Context Annotations. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 871–882.
- [15] Masaki Suwa and Barbara Tversky. 2002. External representations contribute to the dynamic construction of ideas. In *International Conference on Theory and Application of Diagrams*. Springer, 341–343.
- [16] Poorna Talkad Sukumar, Anqing Liu, and Ronald Metoyer. 2018. Replicating User-defined Gestures for Text Editing. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces*. 97–106.
- [17] Frédéric Vallée-Tourangeau, Sune Vork Steffensen, Gaëlle Vallée-Tourangeau, and Miroslav Sirota. 2016. Insight with hands and things. *Acta psychologica* 170 (2016), 195–205.
- [18] Gaëlle Vallée-Tourangeau, Marlène Abadie, and Frédéric Vallée-Tourangeau. 2015. Interactivity fosters Bayesian reasoning without instruction. *Journal of Experimental Psychology: General* 144, 3 (2015), 581.
- [19] Radu-Daniel Vatavu. 2019. The dissimilarity-consensus approach to agreement analysis in gesture elicitation studies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [20] Adam S Williams, Jason Garcia, Fernando De Zayas, Fidel Hernandez, Julia Sharp, and Francisco R Ortega. 2020. The Cost of Production in Elicitation Studies and the Legacy Bias-Consensus Trade off. *Multimodal Technologies and Interaction* 4, 4 (2020), 88.
- [21] Jacob O Wobbrock, Meredith Ringel Morris, and Andrew D Wilson. 2009. User-defined gestures for surface computing. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1083–1092.
- [22] Haijun Xia, Nathalie Henry Riche, Fanny Chevalier, Bruno De Araujo, and Daniel Wigdor. 2018. DataInk: Direct and creative data-oriented drawing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [23] Dongwook Yoon, Nicholas Chen, and François Guimbretière. 2013. TextTearing: opening white space for digital ink annotation. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. 107–112.
- [24] Dongwook Yoon, Nicholas Chen, François Guimbretière, and Abigail Sellen. 2014. RichReview: blending ink, speech, and gesture to support collaborative document review. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. 481–490.
- [25] Dongwook Yoon, Nicholas Chen, Bernie Randles, Amy Cheatle, Corinna E Löckenhoff, Steven J Jackson, Abigail Sellen, and François Guimbretière. 2016. RichReview++ Deployment of a Collaborative Multi-modal Annotation System for Instructor Feedback and Peer Discussion. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 195–205.
- [26] Mohamed Yousef, Khaled F Hussain, and Usama S Mohammed. 2020. Accurate, data-efficient, unconstrained text recognition with convolutional neural networks. *Pattern Recognition* 108 (2020), 107482.
- [27] Shumin Zhai, Per Kristensson, Caroline Appert, Tue Andersen, and Xiang Cao. 2012. Foundational Issues in Touch-Surface Stroke Gesture Design: An Integrative Review. *Found. Trends Hum.-Comput. Interact.* 5, 2 (Feb. 2012), 97–205. <https://doi.org/10.1561/11000000012>