Using Affinity Diagrams to Evaluate Interactive Prototypes

Andrés Lucero[(✉)]

Mads Clausen Institute, University of Southern Denmark, Kolding, Denmark
`lucero@acm.org`

Abstract. Affinity diagramming is a technique used to externalize, make sense of, and organize large amounts of unstructured, far-ranging, and seemingly dissimilar qualitative data. HCI and interaction design practitioners have adopted and used affinity diagrams for different purposes. This paper discusses our particular use of affinity diagramming in prototype evaluations. We reflect on a decade's experience using affinity diagramming across a number of projects, both in industry and academia. Our affinity diagramming process in interaction design has been tailored and consists of four stages: *creating notes, clustering notes, walking the wall,* and *documentation*. We draw examples from eight projects to illustrate our particular practices along these four stages, as well as to ground the discussion.

Keywords: Interaction design · KJ method · Evaluation · Analysis

1 Introduction

Researchers have recently been looking into and studying different (design) methods and reflecting on how they are used in practice by the human-computer interaction (HCI) and interaction design communities. Most of these studies have looked into methods that were originally conceived within and are closely related to design practice, such as probes [22, 32], workbooks [6], and mood boards [24]. First introduced in the 1960s, affinity diagramming (or the KJ method) [2, 13, 16] has its origins in anthropology and social science and has widely been accepted within HCI research. Affinity diagramming is a technique used to externalize, make sense of, and organize large amounts of unstructured, far-ranging, and seemingly dissimilar qualitative data [12]. Common uses of affinity diagramming include analyzing contextual inquiry data [2, 13] clustering user attributes into profiles [21] or requirements [1], problem framing and idea generation [7, 30] and prioritizing issues in usability tests [9].

In this paper, we reflect on a decade's experience using affinity diagramming to evaluate interactive prototypes. Our affinity teams usually consist of two researchers who collect data from 10 to 24 participants (i.e., observations of use during a task, and semi-structured interviews), independently write affinity notes (i.e., 500 to 2500 notes), and jointly analyze the data (i.e., build an affinity diagram) over a period of two to three weeks. To better suit small to medium interaction design projects in industrial and academic contexts, we have tailored and scaled down Beyer and Holtzblatt's six stages of contextual design [2, 13] to four stages. First, when *creating notes*, we embrace the

affordances of paper [18, 28] and build affinity diagrams with physical paper by producing handwritten sticky notes. Second, when *clustering notes*, we invite team members to go through each other's notes in sequence, to avoid ownership issues and to create a better understanding of the context when an observation of use is made. We also avoid using interview questions to structure the data, letting overarching topics naturally emerge [4] from it. Third, in *walking the wall*, we take advantage of color-coded sticky notes (i.e., one color per participant) to check at a glance whether enough people have raised an issue. We also discuss practices related to pruning the wall, which include merging, arranging, and removing note clusters. Finally, in *documentation*, we pick relevant user quotes and count notes to communicate and quantify our main findings. The main contributions of this paper include: a systematic analysis of affinity diagramming use for prototype evaluations in HCI and interaction design over an extended period; an adaptation of earlier affinity diagramming techniques such as the ones described by Beyer and Holtzblatt [2, 13] which have been tailored to suit small to medium projects; and a discussion on practices that are relevant for general affinity diagramming.

This paper is structured as follows. We begin by discussing related work on affinity diagrams, how they are used in HCI and interaction design, and existing support tools. We then take real-life examples from eight industrial research projects to illustrate the four stages of our particular use of affinity diagramming for interactive prototype evaluations. We close by discussing what we have learnt by adapting affinity diagramming to our own practices, followed by conclusions.

2   Related Work

2.1   Affinity Diagramming and the KJ Method

Affinity diagramming is a technique used to externalize, make sense of, and organize large amounts of unstructured, far-ranging, and seemingly dissimilar qualitative data [12]. Japanese anthropologist Jiro Kawakita devised the KJ method [16] (on which affinity diagramming is based on) as a tool for use in anthropology to synthesize idiosyncratic observations of raw data obtained through fieldwork to find new hypotheses. In Japan, the KJ method has become popular as a systematic approach to problem solving in fields such as research, invention, planning, and education.

Affinity diagrams (or KJ method charts) are a wall-sized paper-based hierarchical representation of data. The KJ method consists of four basic steps [16]. In *label making*, the main facts or issues in relation to the data are captured onto separate pieces of paper (or sticky notes). Rather than grouping notes in predefined categories, affinity diagrams are built from the bottom up. Therefore, in *label grouping*, individual notes are shuffled and spread out on a table, and then read several times. After interpreting and considering their underlying significance [9], individual notes are put up on a large empty table or blank wall one at a time, forming teams of labels (or clusters) that are iteratively rearranged [29]. 'Lone wolves', or notes that do not seem to fit in the existing clusters, are left aside for later use. Clusters are then given titles (or named) and, if working on a table with pieces of paper, all notes are put together in a pile with the title

clipped on top. Cluster names and 'lone wolves' are read and grouped into more abstract groups giving rise to general and overarching themes [20]. In *chart making*, the resulting clusters are spatially arranged and transferred to a large sheet of paper, where they are annotated using symbols and signs (i.e., connection, cause and effect, inter-dependence, contradiction) to show the relationships between groups. Finally, in *explanation*, the resulting chart is first verbally explained and then described in writing.

   Although conceived for individual use, the KJ method is well suited for collaborative data analysis, supporting parallel work and creation of a shared interpretation of the data. The greater-than-human-sized space often used allows people to simultaneously view, discuss, and modify the artifact [18]. The simple skills needed to fill a table with pieces of paper, or a wall with sticky notes, and move notes around to suggest new associations make affinity diagramming (or the KJ method) a tangible, easy, and approachable way to look into complex data.

2.2   Different Uses of Affinity Diagrams

Contextual Inquiry. Beyer and Holtzblatt [2, 13] adapted the original KJ method to analyze observational and interview data gathered during contextual inquiries [5, 9, 20, 21]. Affinity diagramming is often used as a starting point for design [20], helping keep design teams grounded in data as they design [9]. The contextual design process begins with *contextual interviews*, where field data from between four and six different work sites is gathered in an attempt to understand work practice across all customers. Second, *interpretation sessions and work modeling* allow every team member to experience the interviews and capture key points (affinity notes), after which five models of the user's work are created (i.e., flow, cultural, sequence, physical and artifact models). Third, *consolidation and affinity diagram building* consist of merging the data from the five models and the affinity notes to represent the work of the targeted population. Fourth, in *visioning* the team reviews the consolidated data by walking the data and then running a visioning session of how the user's work will be streamlined and transformed. Fifth, *storyboarding* consists of fleshing out the vision in detail using hand drawn pictures and text. Finally, *paper prototypes and mockups interviews* consist of designing user interfaces on paper and testing them with users.

   While Beyer and Holtzblatt's use of affinity diagrams is aimed at the early stages of the design process, our work looks at how affinity diagrams can provide support to analyze interactive prototype evaluations in the later stages of the design process. Moreover, stages such as *contextual interviews, work modeling, storyboarding,* or *paper prototypes and mockups interviews* are no longer relevant once interactive prototypes are in place and ready to be evaluated, as those activities should have happened earlier in the process.

User Profiles and Requirements. At the start of developing a new software or product, user profiles (or personas) are constructed to provide a direction to the whole process [21]. Team members create a list of audience attributes on sticky notes, thus bringing their own views on the user. These attributes are then collectively clustered into three to eight user profiles. Notes that do not fit in existing clusters create new

ones. Clusters are discussed and notes are moved around until everyone agrees on them. Sometimes it is the customers themselves (i.e., experts in a given field) who create an affinity diagram to specify requirements for a new system [1, 4]. Starting from the design brief, experts write down fairly succinct statements (i.e., requirements, needs, wishes and hopes) on sticky notes capturing the features, properties and expected behaviors of parts of the new system. Requirements are clustered into common themes, while near-duplicate ones are discarded.

Problem Framing and Idea Generation. In design practice, affinity diagrams are used to analyze a design problem or to create first design solutions [4, 7]. With a given design problem in mind, designers write down words, short sentences, or create small sketches on sticky notes to stimulate diversity in ideation phases. Ideas are then clustered to identify common issues and potential solutions, ultimately helping to frame the design problem. Used to achieve consensus and understanding among participants in a discussion [30], a radial affinity diagram is usually built on a table. A key problem or theme is first placed in the center of the base sheet. Participants then take turns in placing their cards on the table, aligning similar cards in a spoke-like fashion. Visual connections between similar cards can be created using lines of paper clips. After discussion, participants fix the card positions and links. In web design, affinity diagramming is used as a form of collaborative sketching [18] to create sitemaps with the structure of a website. Information architects and visual designers first collect ideas about what should be in a website onto sticky notes and then arrange them on the wall into categories, usually on a whiteboard. Sitemaps can grow fast and end up including 200 to 300 notes. Visual designers sketch page designs directly on empty spaces of the whiteboard.

Usability Tests. In usability testing, affinity diagrams are also used to help teams prioritize which issues will be fixed and retested [4, 9]. At the start of a usability test session, the team assigns a sticky note color to each participant and watches as they perform tasks from an observation room. Observations and quotes are captured on the notes, which are put up on a wall. Common interface issues and problems will emerge. Several colored notes in one issue will indicate that many people experienced a similar problem and should probably be addressed first.

In HCI and interaction design, affinity diagrams have also been used to analyze (post-task) interview data from interactive prototype studies [3, 31]. In this paper, the novel and particular use of affinity diagrams in prototype evaluations to analyze both observations of use while participants perform a task and (post-task) interview data is discussed [17, 23, 25, 26].

2.3   Affinity Diagramming Support Tools

Software tools have been available to create affinity diagrams. CDTools[1] was a software package to support the Contextual Design process offered by InContext Design. PathMaker[2] is a split screen interface that allows recording, dragging, and grouping

---

ideas into affinity sets. StickySorter[3] allows working visually with large collections of notes. Koh and Su [19] created an affinity diagram authoring environment that provides an infinitely large workspace to post and organize notes, dynamic group layout and repositioning, meaningful zoom levels for navigation (i.e., to fit note, group of notes or selection area), ways to save, restore and export diagrams (JPEG), as well as ways to search for notes and groups (i.e., by text matching and related words). The main disadvantages of these tools include a lack of support for collaboration (i.e., single user) and having to do things in a certain structured way [10].

Support for affinity diagramming has also been available in prototype form. The Designers Outpost [18] combines the affordances of paper and large physical workspaces with the advantages of digital media. People write on sticky notes with a normal pen, which are then added to an electronic whiteboard used as a canvas. Using computer vision, the system tracks the position of the notes, which can be physically moved around or removed from the board. Physical notes can be structured, by drawing lines from one note to another, and annotated with digital pens. AffinityTable [8] replicates the benefits of physical affinity diagramming (i.e., copying, clustering, piling and collecting) and enhances current practices (i.e., highlighting, focusing, searching, retrieving images) by combining one vertical and one horizontal display, plus digital pen and paper. Multi-touch gestures and physical tokens provide input to the interactive table. The GKJ system [30] provides a way to digitize affinity diagrams and the process of building them. Using wireless Anoto-based pens, the system records annotations on cards and on a base sheet. Pen gestures are used to determine the position and orientation of a card, as well as to group/ungroup clusters of cards. Time stamped gestures allow visiting the history of the affinity diagram process, and also serves as an undo function. A PC editor allows further editing the virtual diagram. Harboe et al. [11] proposed a distributed system of digital devices (i.e., mobile phones, tablets, stationary camera, and projector) to augment affinity diagramming while aiming to support existing paper-based practices. Using a Magic Lens metaphor, paper notes tagged with unique QR codes can be tracked with the camera of a mobile phone or tablet. Once recognized, the note can be augmented with additional metadata information, which is projected on top of the physical note. The prototypes discussed here aim to augment affinity diagramming, however Klemmer et al. [18] report that some digital features have the potential to interrupt the designers' creative flow and can be considered distracting (i.e., too many things flashing).

Despite the pervasiveness of new technologies, paper remains a critical feature of work and collaboration [27, 28]. Luff et al. [28] discuss some of the affordances of paper that seem critical to human conduct. Paper is *mobile* as it can easily be relocated and juxtaposed with other artifacts, and *micro-mobile* as it can be positioned in delicate ways to support mutual access and collaboration. Paper can be *annotated* in ad hoc ways, allowing people to track the development of the annotations and recognize who has done what. Paper is *persistent* [18], retaining its form and the character of the artwork produced on its surface. In addition, paper allows people to simultaneously see its contents from different reading angles, and it can become the focus of gestures and

---

remarks [27]. The affordances of paper have played a key role in our practices with affinity diagrams, including our preference to use physical paper to digital alternatives, as well as to manually write notes on sticky notes.

3   Eight Affinity Diagrams

We reflect on a decade's experience using affinity diagramming to evaluate interactive prototypes, both in industry and academia. Real-life examples from eight industrial research projects where the technique has been used (see Table 1 for an overview) will help illustrate how affinity diagrams are used in practice, as well as to ground the discussion.

Table 1. Overview of eight affinity diagramming cases.

Prototype	Evaluation type	Evaluation participants	Note takers	Note technique	Affinity notes	Affinity team size	Notes on wall	Notes discarded
A *MindMap*	Group (2 × 3)	6	2	Mix	258	2	232	10 %
B *Pass-Them-Around*	Group (5 × 4)	20	2	Mix	811	3	740	9 %
C *MobiComics*	Group (3 × 9)	27	2	Printed	2433	5	2243	8 %
D *EasyGroups*	Group (6 × 4)	24	2	Mix	1096	2	992	9 %
E *FlexiGroups*	Group (4 × 6)	24	2	Mix	715	2	673	6 %
F *Image Space*	Individual	10	2	Handwritten	505	2	391	23 %
G *Twisting Touch*	Individual	24	2	Handwritten	1037	2	630	39 %
H *NotifEye*	Individual	13	2	Handwritten	1276	2	946	26 %

Prototypes A to E [25] were related to groupware, exploring the use of mobile phones for collaborative interactions in different physical and social use contexts (i.e., office work, media consumption at home, public expression in a pub, and general group formation). These prototype evaluations were conducted with different numbers of participants (i.e., between six and 27 people) in groups of varying sizes (i.e., between three and nine people per group). Prototypes F, G, and H, on the other hand, were evaluated individually (between 10 and 24 participants per evaluation). Prototype F [23] was a social network service built around sharing personal photos. Prototype G [17] allowed using a flexible handheld interface to provide input for interaction. Finally, Prototype H [26] explored the use of interactive glasses to provide notifications on the go. Most of these evaluations were conducted in controlled lab environments, except for prototypes C, F, and H, which took place in public spaces.

In all prototype evaluations, two researchers independently made notes as they watched videos of participants performing an interaction task and a semi-structured interview. Handwritten notes were created for all prototypes but one, i.e. C, which used digital notes printed on label templates only. In addition, digital notes printed on sticky notes or on paper (plus removable tape) were produced for four prototypes (i.e., A, B, D, and E), resulting in a mix of handwritten and digitally created notes. The size of the team that built the affinity diagram ranged from two to five people, and always included

the same two researchers who made the notes in the first place. The resulting affinity walls differed in size, ranging from 232 to 2243 notes. Around 15 % of the notes were discarded, except for two cases, prototypes G (39 %) and H (26 %).

4 Affinity Diagramming Process

We outline stages and properties of our particular use of affinity diagramming for interactive prototype evaluations. As was mentioned earlier, Beyer and Holtzblatt's [2, 13] use of affinity diagrams is intended for the early stages of the design process, and some of their stages are not relevant for prototype evaluations. We have tailored their process to better suit small to medium interaction design projects in industrial and academic contexts. More specifically, we have combined their first two stages (i.e., *contextual interviews* and *interpretation sessions and work modeling*) into *creating notes* by placing prototype evaluations at the core, and removing the work modeling activity. Their third and fourth stages (i.e., *consolidation and affinity diagram building* and *visioning*) are closely related to *clustering notes* and *walking the wall*, respectively. Finally, their last two stages (i.e., *storyboarding* and *paper prototypes and mockup interviews*) have been replaced by *documentation.* Our process consists of four stages: *creating notes, clustering notes, walking the wall,* and *documentation*.

4.1 Creating Affinity Notes

This first stage of the process starts with the actual evaluation of the prototype. After carefully planning the main research questions, internal ethics committees and privacy reviews, consent forms, the introduction, the task, coffee breaks, the semi-structured interview questions, the debriefing, and rewards, we collect data over a period of one or two weeks (depending on the number of participants). Each evaluation session usually lasts between one and two hours. As a result, we end up with 6 to 24 h of video to analyze, both from observations of use during a task, and from the semi-structured interviews. Interpretation sessions are then conducted within 48 h after the prototype evaluations. Two researchers with mixed background (i.e., a designer and a psychologist, or a designer and a computer scientist) independently make notes as they watch videos of an interaction task and a semi-structured interview. Affinity notes typically include handwritten text, but can also comprise drawings and annotations (Fig. 5). The number of affinity notes can vary between 500 and 2500, depending on the number of interviews, their duration and the level of detail captured. It usually takes us twice as much time to write affinity notes as the length of the videos.

Handwritten Sticky Notes. We begin by placing a stack of sheets of A3 paper on our desk (Fig. 1a). These sheets are used as a canvas onto which to stick notes. A3 is a comfortable format to work with while writing affinity notes; it is large enough to put several sticky notes on it, and small enough to fit on a desk in front of a computer. Traditionally, standard 3×3 or 5×3 inch (7.6×12.7 cm) Post-it notes have been used to write affinity notes. However, we have found those sizes to unnecessarily increase the overall size of the affinity wall and thus the amount of walking for the affinity team

members. Moreover, larger affinity notes invite more verbose expressions from the note takers. Therefore, we use the smaller 2 × 2 or 2 × 1.5 inch (5.1 × 3.8 cm) sticky notes, which are available in different brands, colors, and slightly different sizes (Fig. 1a). These smaller sticky notes fit in a grid of 8 by 6 notes, totaling 48 notes per sheet.
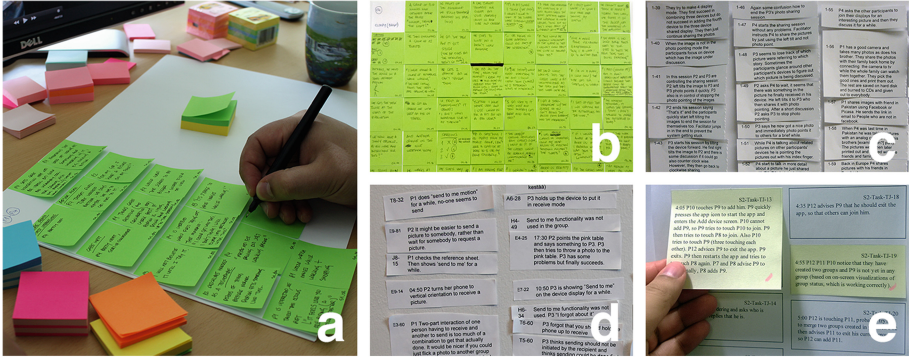
Fig. 1. Creating affinity notes. (a) Writing notes on sheets of A3 paper, (b) handwritten sticky notes, (c) digital notes printed on paper, cut with scissors and attached with tape, (d) digital notes printed on labels, and (e) manually printed sticky notes.

We assign a sticky note color for each participant (Fig. 1b). In case of group evaluations (e.g., prototypes A to E on Table 1), we define one color for each group of participants. With smaller groups of people (e.g., three participants), we try to treat them as individuals as much as possible and thus assign three sticky note colors, one for each. Using different note colors for each participant allows us to tell how many people raise a certain issue by glancing at each category.

Blank sticky notes are arranged (usually in columns) and are given a unique identifier consisting of a participant or session number, followed by a running sequence number (e.g., P1_01 or S1_01), which is handwritten in the lower-right corner. Note takers do not need to identify themselves on each note as their handwriting provides a quick way to know the author of a given note. We draw a single thick black line to separate task from interview data (Fig. 1b). We place two or three blank sticky notes on the A3 canvas before starting the video to take notes. To optimize our use of time, subsequent notes are then added and identified in parallel during natural transitions or silent moments in the video. However, some parts of the video do require us to explicitly pause and rewind the video to capture a note in more detail. Those breaks also provide opportunities to add new blank notes to the canvas. Once a sheet of A3 paper is filled with 48 affinity notes, that sheet is moved to the back of the stack. Working with a stack of sheets allows the note taker to quickly consult whether something has been missed or captured earlier. Once all videos have been analyzed, the sheets of A3 paper filled with affinity notes can now be easily transported to the affinity room.

Digital Notes Printed on Paper, Labels or Sticky Notes. We have also experimented with using different digital note types in our evaluations of prototypes A to E (Table 1).

First, digital affinity notes are typed on personal computers, then printed on paper, and each note is individually cut with scissors [10] (Fig. 1c). In an attempt to provide a similar function and feel as sticky notes, removable tape, blue painter's tape, or yellow masking tape are used so the notes can be easily attached and moved around the affinity wall. The extensive manual work needed to produce such notes can easily delay the start of building the affinity wall by a couple of hours. Second, label templates (e.g., Avery®) are another alternative to printing digital affinity notes (Fig. 1d). Since the entire surface of the note then becomes adhesive, it produces mixed results as we have had difficulties removing the notes from the wall depending on the wall finishing materials (e.g., paint, wallpaper, wood, glass, etc.). Third, affinity notes can be laser printed on sheets of Post-it® notes. However, priced at almost US$1 per note and available in one-size (3 × 3 inches), this alternative is both expensive and impractical. We have come up with a cheaper way to print digital affinity notes on standard sticky notes (Fig. 1e). *MixedNotes* [14] is a software tool that imports text from any text editor, recognizes blank lines as separates notes, adds a unique sequential identifier, optimizes the size of the note to be printed based on the amount of text, and shows an on-screen preview of how notes will be printed on paper. *MixedNotes* first prints a background sheet, a template onto which sticky notes manually cut in three different sizes are attached, and then the same sheet must be put back into the printer for the digital notes to be printed on the sticky notes. Despite claims that digital notes could improve current practices (i.e., faster note-taking, searching), we have not found this to be the case (e.g., search option only used once).

4.2 Clustering Notes

Selecting and Preparing a Room. Holtzblatt et al. [13] stress the importance of getting a dedicated team room for the duration of the project to avoid wasting time finding another room, packing up materials, and relocating half way through the process. However, not all organizations have a room that can be blocked for one or two weeks. We have used different strategies to secure such a dedicated space for a couple of weeks both in companies and universities. For example, we have reserved an internal meeting room for two weeks (Fig. 2c), used our own shared office (Fig. 2d), used a common innovation space outside our premises, temporarily repurposed a usability lab (Fig. 2a), and used an office space that was emptied before major renovations (Fig. 2e). Room sizes have ranged from a 2.3 × 3.6 m meeting room (Fig. 2c) to a 6 × 8.4 m usability lab (Fig. 2a). The room should provide plenty of wall space for the affinity to spread out [13].

We begin by mounting the sheets of A3 paper containing the affinity notes on the wall (between 1.2-2 m from the floor) using masking tape (Fig. 2b). We then line up the remaining wall space with white flipchart sheets, butcher paper, or statically charged polypropylene film onto which the note clusters will form. In addition, we sometimes put white cardboard panels on tables (Fig. 2d). When using smaller rooms, we have even used the windows and the door (Fig. 2c). Our ideal room (Fig. 2a) has moveable whiteboards that we line up with white flip chart sheets of paper held up by magnets. Such a space allows us to rearrange parts of the wall, be it entire whiteboard panels or
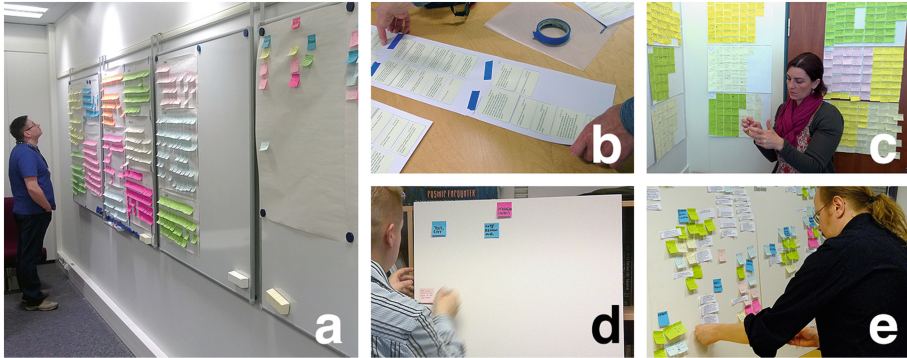
Fig. 2. Clustering notes. (a) A typical room at the start of clustering with notes on the left and right, and empty space in the middle, (b) preparing the room by putting up sheets of affinity notes on the wall, (c) going through notes individually, (d) forming first clusters on a white cardboard panel, and (e) clusters have initial names and a few notes below them.

just a couple of sheets. In such a space, we mount affinity notes towards the left and right wall edges, leaving an empty space in the middle for the affinity wall.

Affinity Diagramming for Prototype Evaluations. When building an affinity wall to analyze prototype evaluation data, the two note takers who created the affinity notes are already familiar with the data since they have seen the videos for all sessions. In cases where the affinity team has included more than two persons (e.g., prototypes B and C), although these people have not created affinity notes, they have been present earlier as observers during the actual prototype evaluations. Therefore, unlike Holtzblatt et al. [13], at the start of building the affinity we do not break the notes up into piles of 20 (i.e., to make it less intimidating). Instead, we invite people to start by going through each other's notes sequentially to avoid potential ownership issues. Our data also more heavily depends on the details of the context when an observation of use was made, therefore we do not mix up the notes (i.e., mix up users). For example, when going through the first part of the evaluation (i.e., the task), it is important for us to be able to identify whether, e.g., a certain function was successfully triggered or a group was having problems on the first, second or third try. An isolated note saying, *"they are having difficulties completing the task,"* will be interpreted differently whether this is happening at the start or by the end of the task.

Despite the seemingly structured way of analyzing the interaction observations, when going through the second part of the data (i.e., the semi-structured interview questions) we avoid as much as possible using the interview questions to structure the data. Instead, we let overarching topics naturally emerge from the data.

Building the Affinity. At the start of building the affinity, team members start by reading each other's notes in silence (Fig. 2c). People will pick notes that raise important issues in relation to the prototype and begin forming rough clusters (Fig. 2d). Once a couple of clusters have been created with a few notes, people will begin verbally coordinating where certain notes are being clustered. Questions such as, *"where are you putting the notes related to this issue?"* will begin to emerge. Thus we

alternate between moments of silence and moments of discussion, the former near the start of the process, the latter as the affinity progresses. Clusters with a few notes below them are initially named and labeled with a blue note (Fig. 2e). These clusters are in turn grouped into more abstract groups labeled with pink notes.

4.3   Walking the Wall

Discussing and Pruning the Wall. After the team has completed a first round reading all notes, roughly between a third and half of the affinity notes will have been moved from the A3 sheets of paper to the affinity wall, thus forming note clusters (Fig. 3a). Early rounds of discussing the wall will then concentrate on communicating the emerging clusters with the team, checking if these clusters fail to cover some important general topics, and identifying overlapping clusters that could potentially be merged. As a result of these discussions, the teams will agree on an initial set of clusters and (blue and pink) labels. Drawing arcs [12] and sticking stripes of tape [15] between clusters can be used to visually show related parts of the affinity diagram. A note that belongs in two clusters can be duplicated, or split by ripping the paper and adding tape.

The team may also decide at this point to define a tentative cluster hierarchy. On one hand, fixing a hierarchy this early on in the process tends to limit the bottom-up nature of the process. On the other hand, shifting panels around late into the process to define a hierarchy can have a detrimental effect. Social and spatial awareness of the affinity diagram is an important part of building a cognitive model of the data [4, 10, 16]. Moving parts of the affinity wall can create confusion and lead to a waste of time when people are trying to find where to place a given note.

Fig. 3. Walking the wall. (a) Affinity notes have been moved the A3 sheets of paper to form the affinity wall, (b) people slowly start discussing the contents of each category, (c) moving several notes at a time when there are enough categories, (d) verbalizing the act of moving a note to the wall, and (e) user statements in first person are written on larger yellow sticky notes (Color figure online).

In later rounds of discussing the wall, the team will more closely inspect the contents of each cluster (Fig. 3b). Specific notes that are unrelated to a cluster may be put on a different cluster, set aside to potentially create a new cluster, or even back in its original location on the A3 sheets of affinity notes. Similarly, clusters may be merged, moved to a different location, or can altogether disappear. Pruning the wall thus includes merging clusters, arranging the cluster hierarchy, and removing notes (and clusters) from the affinity wall.

Adding Notes to Existing Clusters. As was mentioned earlier, the affinity team will switch back to reading affinity notes from the A3 sheets of papers in between rounds of discussing and pruning the wall. As the affinity wall progresses, team members will be more familiar with the existing clusters and have a better understanding of where a given note could be put up on the wall. As a result, people will evolve from moving one note at a time to the wall, to several simultaneously (Fig. 3c). Each finger can represent a cluster, and several notes can be placed on each finger, and thus ten notes can be moved to the wall at the same time. Another phenomenon that we have observed at these later stages of the process is that team members are more open to verbalize the act of moving a given note up to the wall (Fig. 3d). –*"Listen to this note [reads the note aloud]. It goes here."* –*"Yes."* Such actions are performed to confirm that existing categories remain valid also after new bits of data are added to the wall.

Each round of discussing and pruning the wall, and adding notes to existing clusters often takes between 60 and 160 min. Besides alternating between these two different types of activities, it is important to have 30 to 60 min breaks in between rounds. Affinity diagramming can be a mentally demanding activity [13], especially when it is carried out for three to five days in a row. Natural breaks such as lunchtime or taking a coffee, but also an unrelated meeting can provide a much-needed time for the mind to rest and think about something else, other than the ongoing analysis.

Finalizing the Wall. Once there are no more useful notes left on the wall, typically the last 15 % of the notes (Table 1), we read each category note by note and make sure each note belongs to that category. We also make sure there are enough notes and more than two people raising an issue in each category. We then check that every blue and pink note still makes sense within the hierarchy. For each blue label, a succinct user statement describing the issue that holds all individual notes together is written on a large sticky note [2] (Fig. 3e).

4.4   Documentation

Creating a Digital Record of the Affinity Wall. Keeping a digital record of the finalized affinity wall (Fig. 4b) allows sharing the results across sites. Miura et al. [30] indicate that in the typical KJ method, there is one participant who digitizes the outcome of the diagram by inputting all card content and arranging the card structure using a mouse and keyboard. Similarly, we assign a scribe [18], a member of the affinity team who saves the resulting info. Using PowerPoint or Word, a document with the hierarchical category structure and its descriptions is created (Fig. 4a). Depending on the number of affinity notes and detail of the documentation, roughly one to three

PowerPoint slides are made for every pink label. Each slide contains the name of the pink and blue labels, the user statement written in first person that describes all notes under a blue label, plus a selection of the most representative user quotes on that particular issue (typically between 1 and 4 user quotes per blue label). Picking relevant user quotes, ones that capture the essence of what participants tried to tell us, will play an important role in communicating the main (positive and negative) findings of the study to different stakeholders, improving existing designs, and further disseminating the end results (e.g., publication at an academic conference).
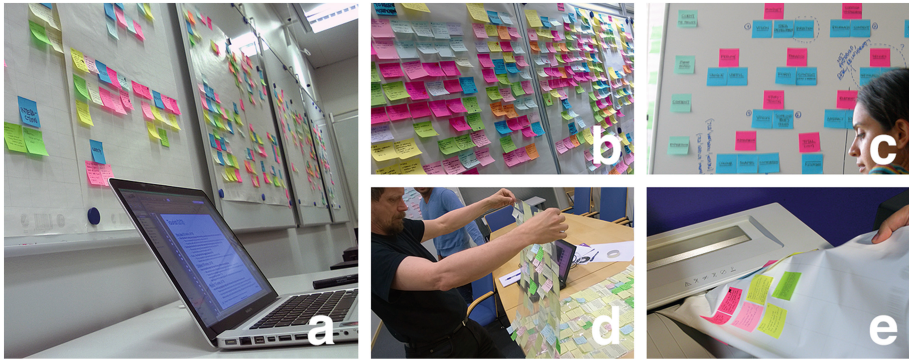
Fig. 4. Documenting the wall. (a) Digitally recording the wall using a laptop, (b) a final affinity wall with pink and blue labels, (c) building the final category structure on a whiteboard, (d) cleaning the room to its original state, and (e) shredding the affinity wall (Color figure online).

In addition to making a digital version of the affinity diagram, we take (high-resolution) digital photographs of each panel. Each photo usually contains one pink note, two to five blue notes, plus their corresponding affinity notes that make up those categories. We sometimes also take close-up panel shots (e.g., if we need to leave the room and save the results to a digital document at a later time). Proper lighting conditions and (color) contrast between the text and note should be considered when photographing panels [19]. To visually assist the wall documentation process, we sometimes build the final hierarchical structure of categories on a whiteboard (Fig. 4c).

Quantifying Observations of Use and Raised Issues. Another particular use of affinity walls for interaction design that we have developed over the years is to count the total number of notes and the number of people that raised an issue. Counting the total number of notes allows us to check how frequently the participants mentioned an issue or topic. When creating the final affinity wall hierarchy, the overall note numbers for each pink and blue label provide us with an additional way to prioritize one topic over another. Special care should be taken to identify if a category with a large number of notes consists of one or two people mentioning the same issue repeatedly.

Similar as for usability testing [9], we also count the number of people that raised an issue. This is where using different note colors comes in handy as we can glance at a category and quickly get a sense of how many different people mention a certain issue

(Fig. 4b). By doing this, we are able to quantify our (mostly) qualitative findings. An opening statement such as *"most participants (16/20) explicitly said the prototype was easy to use"* will usually accompany a qualitative finding. Such statements allow us to shed light on and better ground our qualitative findings.

Cleaning Up and Discarding the Affinity Wall. After the affinity diagram has been duly documented, the room must be cleaned up. Besides taking the affinity wall and the (mostly) empty sheets of A3 paper down from the wall (Fig. 4d), the room needs to be arranged back to its original state (i.e., moving tables and chairs, and packing materials). Lining up the wall space with large sheets of paper at the start of the process for note clusters to form can greatly speed up the clean up process. Removing large sheets of paper is easier done and faster than manually removing 500 to 2500 notes one by one. Rolling up the affinity diagram to temporarily store it should be avoided, as the notes will tend to bend and might altogether fall [19]. Once the affinity wall has been taken down, and depending on internal practices regarding data handling and privacy, it is time to shred and discard the data (Fig. 4e). Depending on the final number of notes, cleaning up and discarding the wall can take up to two hours.

5   Discussion

5.1   Return on Investment and Impact

A recurring question within organizations when using affinity diagrams in interaction design evaluations is that of resources. Beyer and Holtzblatt [2] first collect data from 15–20 participants, producing 50 to 100 notes for each two-hour interview, for a total of 1500 notes. They then recommend having one person per 100 notes to build the affinity diagram in one day, for a total of 15 people. Due to the large number of participants (i.e., 10 to 24) and the resulting notes involved (i.e., 500 to 2500), our affinity diagramming process takes a team of two researchers between two and three weeks to collect the data (i.e., one week) and complete the analysis (i.e., one to two weeks). It takes us two to five days to individually write affinity notes, two to five days to build the affinity wall with the affinity team, and one or two days for an assigned scribe to document the wall. Holtzblatt et al. [13] describe two-person projects where it can take two to three weeks to gather requirements for participant numbers of less than ten. Therefore, we feel that our two-person team being able to analyze data from up to 24 people in the same amount of time is a good success indicator in terms of resources. Thanks to the mix of qualitative and quantitative analysis, plus the level of detail in our findings for each prototype evaluation, we have been able to easily justify assigning two people full-time to work on an affinity diagram.

Regarding impact, by systematically introducing affinity diagramming to our industrial research projects we were able to not only find existing UX and usability issues with our interaction designs, but also to identify and define new lines of research, requirements that had to be integrated to our designs, and new ideas that were filed as invention reports. While most resulting affinity diagram clusters covered issues with the interactive prototype that was currently being evaluated, for every project we had two

additional panels, one labeled 'Future Research Areas/Other Topics' and another one 'Ideas/IPR'.

5.2 Number of Notes for Observations of Use and Interview

When creating affinity notes for interaction design studies, it is difficult to estimate the ratio of observation of use notes during the task to semi-structured interview notes. For a 45-min video consisting of 15 min of data for the task and 30 min for the semi-structured interview, one would perhaps expect a 1:2 ratio in the final number of notes as can be seen from Fig. 1b, where 20 and 40 notes were made respectively. However, we have come across unusually large numbers of notes from the observation part, sometimes even reaching 50 % of the total number of notes. The evaluation of Prototype E (Table 1) was related to groupware and explored general group formation using mobile phones for collaborative interactions. The fact that there were six participants trying out different strategies to form a group made it increasingly difficult to keep track of the overall situation. Tasks were taking place in parallel and there were many micro interactions happening. Although the task videos in this case were relatively short (i.e., 15 min), the complexity of the interaction data had a big impact on the final number of notes. Similarly, we roughly estimate that for a 45-min video it would take 90 min (or double the time) to generate affinity notes. However, in the case we are currently discussing, the amount of details in the interaction also had an effect on the overall note-taking time.

Our suggestion is for the note takers to start the interpretation sessions, go through the videos for the same two or three participants, and check how many notes each has made for the task and the semi-structured interview parts for each participant. Although there can be differences in the level of detail that each note taker captures, this coordination provides a reasonable estimate of whether someone is focusing on too much detail or being too general before all sessions have been interpreted.

5.3 Note Types and Leftover Notes

The most common affinity note types include observations of use (Fig. 5a), good participant quotes for use in publications (Fig. 5b), and design issues or ideas. As stated earlier, roughly 15 % of the notes are discarded (i.e., notes that are not included on the final affinity wall). Some of these notes are used to perform general counting (e.g., how many pedestrians and bicycles a participant encountered during the task) (Fig. 5c), to mark certain parts of the evaluation (e.g. *"Task given by [the] instructor"*) (Fig. 5d), to draw an arm posture (Fig. 5e), to personally indicate how well a task was performed (Fig. 5f), to draw different group formation strategies (Fig. 5g), or to record contextual information (e.g., *"Police drives slowly past:)"*) (Fig. 5h). While note types 5c-5 h may at first seem superfluous, we argue for their importance from a holistic interaction design perspective as they help researchers thoroughly analyze micro-interactions, social interactions, and other contextual factors that may have an effect on the overall results.
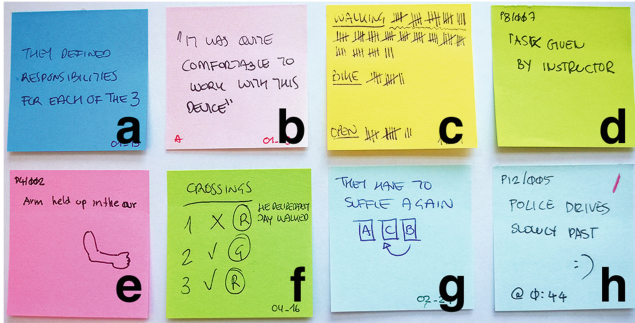
Fig. 5. Leftover affinity notes. (a) Observations of use, (b) participant quote, (c) general counting, (d) personal separator, (e) drawing of body posture, (f) personal counting, (g) formation strategies, and (h) personal contextual observation.

Prototypes F, G and H (Table 1) have unusually large numbers of unused notes (23 %, 39 %, and 26 %). Based on the total number of affinity notes, these three projects could be considered average in size at 505, 1037, and 1276 notes. These three evaluations were conducted individually, which may cause note takers to make and write down more disconnected, random, and anecdotal observations. However, we do not believe this to be the sole source for the large number of discarded notes. We also ran out of time to go in more depth and further remove notes from the A3 sheets of paper to transfer them to the final affinity wall, as we spent three or four days to build these affinity walls.

6   Conclusion

By reflecting on a decade's experience using affinity diagramming across a number of projects, we have discussed how we have tailored the process for use in HCI and interaction design evaluations to four stages: *creating notes, clustering notes, walking the wall,* and *documentation*. Digital affinity diagrams can be especially convenient for situations when data are available in digital format (e.g., tweets, Facebook or YouTube comments), allowing single users to perform data coding, clustering, counting, and statistics in Excel, and are easy to transport. Despite existing attempts to augment affinity diagramming by making parts of the process digital, we have found that traditional paper affinity diagrams are better suited for collaborative analysis, they support building a cognitive model of the data by social and spatial awareness, and allow people to quickly transition from directly moving data around to having the full overview of the wall by simply taking a few steps away from or towards the wall. Providing such a flexible access to data in digital format, and in a way that is suitable for collaborative analysis would require the use of very large displays. In addition, we have embraced the affordances of paper by producing handwritten sticky notes, as we have not found alternative digital note types to offer clear advantages (i.e., speed, searchability). In tailoring Beyer and Holtzblatt's affinity diagramming process to provide better support when analyzing interactive prototype evaluations, we

have been able to better understand the context of use by checking the data sequentially, perform a micro-interaction analysis by creating and looking into detailed notes that might otherwise be discarded (e.g., counting, body postures, strategies), and to quantify qualitative findings at a glance using color-coded sticky notes.

References

1. Benyon, D.: Designing Interactive Systems: A Comprehensive Guide to HCI, UX and Interaction Design. Pearson Education, Harlow (2013)
2. Beyer, H., Holtzblatt, K.: Contextual Design: Defining Customer-Centered Systems. Morgan Kaufmann, San Francisco (1998)
3. Buur, J., Soendergaard, A.: Video card game: an augmented environment for user centred design discussions. In: DARE 2000, pp. 63–69. ACM (2000)
4. Cox, D., Greenberg, S.: Supporting collaborative interpretation in distributed Groupware. In: CSCW 2000, pp. 289–298. ACM (2000)
5. Curtis, P., Heiserman, T., Jobusch, D., Notess, M., Webb, J.: Customer-focused design data in a large, multi-site organization. In: CHI 1999, pp. 608–615. ACM (1999)
6. Gaver, W.: Making spaces: how design workbooks work. In: CHI 2011, pp. 1551–1560. ACM (2011)
7. Geyer, F., Pfeil, U., Höchtl, A., Budzinski, J., Reiterer, H.: Designing reality-based interfaces for creative group work. In: C&C 2011, pp. 165–174. ACM (2011)
8. Geyer, F., Pfeil, U., Budzinski, J., Höchtl, A., Reiterer, H.: AffinityTable - a hybrid surface for supporting affinity diagramming. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part III. LNCS, vol. 6948, pp. 477–484. Springer, Heidelberg (2011)
9. Hanington, B.: Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions. Rockport Publishers, Beverly (2012)
10. Harboe, G., Minke, J., Ilea, I., Huang, E.M.: Computer support for collaborative data analysis: augmenting paper affinity diagrams. In: CSCW 2012, pp. 1179–1182. ACM (2012)
11. Harboe, G., Doksam, G., Keller, L., Huang, E.M.: Two thousand points of interaction: augmenting paper notes for a distributed user experience. In: Lozano, M.D., Gallud, J.A., Tesoriero, R., Penichet, V.M.R. (eds.) Distributed User Interfaces: Usability and Collaboration, pp. 141–149. Springer, London (2013)
12. Hartson, R., Pyla, P.S.: The UX Book: Process and Guidelines for Ensuring a Quality User Experience. Morgan Kaufmann, Amsterdam (2012)
13. Holtzblatt, K., Wendell, J.B., Wood, S.: Rapid Contextual Design. Morgan Kaufmann, San Francisco (2005)
14. Jokela, T., Lucero, A.: MixedNotes: a digital tool to prepare physical notes for affinity diagramming. In: AcademicMindTrek 2014, pp. 3–6. ACM (2014)
15. Judge, T.K., Pyla, P.S., McCrickard, D.S., Harrison, S., Hartson, H.R.: Studying group decision making in affinity diagramming. Technical report TR-08-16, Computer Science, Virginia Tech (2008)

16. Kawakita, J.: The original KJ method. Kawakita Research Institute, Tokyo (1991)
17. Kildal, J., Lucero, A., Boberg, M.: Twisting touch: combining deformation and touch as input within the same interaction cycle on handheld devices. In: MobileHCI 2013, pp. 237–246. ACM (2013)
18. Klemmer, S.R., Newman, M.W., Farrell, R., Bilezikjian, M., Landay, J.A.: The designers' outpost: a tangible interface for collaborative web site. In: UIST 2001, pp. 1–10. ACM (2001)
19. Koh, B., Su, R.: Interactions for addressing shortcomings encountered when physically creating and manipulating large affinity diagrams. http://www.benjaminkoh.com/AffinityTool.html
20. Koskinen, I., Zimmerman, J., Binder, T., Redström, J., Wensveen, S.: Design Research Through Practice: From the Lab, Field, and Showroom. Morgan Kaufmann, San Francisco (2011)
21. Kuniavsky, M.: Observing the User Experience: A Practitioner's Guide to User Research. Morgan Kaufmann, San Francisco (2003)
22. Lucero, A., Lashina, T., Diederiks, E. and Mattelmäki, T.: How probes inform and influence the design process. In: DPPI 2007, pp. 377–391. ACM (2007)
23. Lucero, A., Boberg, M., Uusitalo, S.: Image space: capturing, sharing and contextualizing personal pictures in a simple and playful way. In: ACE 2009, pp. 215–222. ACM (2009)
24. Lucero, A.: Framing, aligning, paradoxing, abstracting, and directing: how design mood boards work. In: DIS 2012, pp. 438–447. ACM (2012)
25. Lucero, A., Jones, M., Jokela, T., Robinson, S.: Mobile collocated interactions: taking an offline break together. Interactions 20(2), 26–32 (2013)
26. Lucero, A., Vetek, A.: NotifEye: using interactive glasses to deal with notifications while walking in public. In: ACE 2014, Article 17, 10 p. ACM (2014)
27. Luff, P., Heath, C.: Mobility in collaboration. In: CSCW 1998, pp. 305–314. ACM (1998)
28. Luff, P., Heath, C., Norrie, M., Signer, B., Herdman, P.: Only touching the surface: creating affinities between digital content and paper. In: CSCW 2004, pp. 523–532. ACM (2004)
29. Mayhew, D.J.: The Usability Engineering Lifecycle: A Practitioner's Guide to User Interface Design. Morgan Kaufmann, San Francisco (1999)
30. Miura, M., Sugihara, T., Kunifuji, S.: GKJ: Group KJ method support system utilizing digital pens. IEICE Trans. Inf. Syst. 94(3), 456–464 (2011)
31. Nunes, M., Greenberg, S., Neustaedter, C.: Sharing digital photographs in the home through physical mementos, souvenirs, and keepsakes. In: DIS 2008, pp. 250–260. ACM (2008)
32. Wallace, J., McCarthy, J., Wright, P.C., Olivier, P.: Making design probes work. In: CHI 2013, pp. 3441–3450. ACM (2013)