Buried in the Stream: Personal Information Management and Re-finding in High-Volume WhatsApp Groups

Akhil Cleatus

Department of Interaction Design
National Institute of Design, Bengaluru, India
akhil_cleatus@nid.edu

Course guide: Dr. Mamata N. Rao

Unpublished student research paper, prepared for the Design Research module, M.Des Interaction Design, National Institute of Design, Bengaluru, February 2026. Formatted in the ACM SIGCHI style for academic practice. Not submitted to or published by any conference or journal.

ABSTRACT

Whittaker and Sidner coined *email overload* in 1996 to describe inboxes that had quietly become task managers, schedulers and personal archives. Three decades later the same reappropriation has migrated to mobile group messaging, but without the folders, flags and filters that email eventually acquired. This paper reports a mixed-methods study of how Indian users store and re-find information inside WhatsApp's flat, strictly chronological group interface. We conducted eleven contextual enquiries with postgraduate design students ($N = 11$, aged 23–31), in which participants demonstrated live retrieval tasks, and a survey of WhatsApp group users ($N = 59$). We find that overload in group messaging is not primarily a volume problem: a majority of participants were untroubled by the size of their archives, yet only 32% were confident of retrieving a file shared weeks earlier. Instead, failure concentrates in seven distinct and separable breakdowns, of which *media non-indexability* and *keyword rigidity* are the most consequential. We further find that users respond by constructing *parallel archives*, most often a self-only group, that reproduce the very flatness they were built to escape, and by *outsourcing retrieval to other people* when the interface fails. Our contributions are (1) a seven-part taxonomy of retrieval breakdown in chronological messaging interfaces, (2) evidence that re-finding strategies in group messaging are resolved socially, through other people, and (3) design implications for messaging clients that must serve as habitats as well as channels.

Author Keywords

WhatsApp; information overload; personal information management; message retrieval; re-finding; group communication; contextual enquiry; digital wellbeing.

ACM Classification Keywords

H.5.3. Group and Organization Interfaces: Computer-supported cooperative work; H.3.3. Information Search and Retrieval.

INTRODUCTION

The seminal 1996 paper by Whittaker and Sidner observed that email inboxes contained far more than unattended incoming messages [20]. Inboxes held to-dos, to-reads, reference material and ongoing correspondence; the client had become, in their phrase, a Swiss-Army knife of information management. That observation generated two decades of research [1,2,5,8,9,19] and a generation of tools (threading, flags, filters, labels, priority inboxes) aimed squarely at the problem. Revisiting the question in 2006, Fisher et al. found inboxes still overloaded [9]; revisiting it again in 2014 with Gmail, Grevet et al. found overload persisting and in fact taking a new form, with personal accounts overloaded by message *type* rather than by status [10].

Meanwhile the centre of gravity of everyday communication moved. For a large share of the world's users, and particularly in India, the primary site of coordination is no longer email but WhatsApp. O'Hara et al. characterise WhatsApp less as a channel than as a place of "dwelling", a continuous presence rather than a set of discrete transactions [16]. Church and de Oliveira found users adopting it for its immediacy, cost and sense of social intimacy relative to SMS [6]. That a communication medium accumulates uses well beyond communication is itself long established: Mackay documented the same diversity in early electronic mail [15]. Cheng and Aflatoony recently documented the reappropriation that follows: users text themselves notes, dump messages into threads for later, and return to old conversations as a personal archive [4].

What has not followed the migration are the organisational tools. Email acquired folders in the 1980s, rules in the 1990s and labels in the 2000s. WhatsApp group chat, which now carries much of the same reference load, offers a strictly chronological stream, a single global keyword search, a flat list of starred messages, and a media gallery ordered by date. There are no folders, no tags, no date filters, no content-type filters within search, and no way to name or annotate what has been kept.

This paper asks what happens to personal information management under those constraints. Specifically:

- How do users of high-volume WhatsApp groups store information they expect to need again?
- When re-finding fails, how exactly does it fail?
- What strategies, inside and outside the application, do users construct to compensate?

We approached these questions through eleven contextual enquiries, conducted in participants' own study environment and including live retrieval tasks, followed by a survey of 59 group users. Our contributions are threefold. First, we offer a seven-part taxonomy of retrieval breakdown: keyword rigidity, media non-indexability, chronological burial, contextual decay, contextual fragmentation, non-linear navigation deficit, and the false-storage assumption. The taxonomy separates failures that the literature has tended to treat as one undifferentiated "overload". Second, we show that the dominant repair strategy is social rather than technical: when the system fails, users ask a person. Third, we draw design implications for messaging clients that are now habitats, and are still designed as channels.

RELATED WORK

Email overload and its definitions

Whittaker and Sidner's original formulation of overload concerned message *status*: an inbox mixing to-dos, to-reads and undetermined items [20]. A second and looser usage treats overload as synonymous with volume, meaning the sheer quantity of incoming mail that makes attending to it difficult [8], a dimension for which Hogan and Fisher proposed an explicit measurement scale [11]. Grevet et al. added a third dimension, *type* overload, observing that personal accounts accumulate structurally different kinds of message (bills, newsletters, correspondence) corresponding to the several facets of a life [10]. Their most consequential finding for our purposes is negative: a large archive did not predict a feeling of disorganisation. What predicted it was a high unread count. Size, in other words, is not the pain; unattended obligation is.

We find a parallel asymmetry in group messaging, but with a different pivot. Our participants were largely indifferent to the size of their archives and even to their unread counts; what distressed them was the gap between the confident belief that something had been kept and the inability to produce it on demand.

Filing, piling and re-finding

Malone's study of physical desks distinguished filers from pilers [13], a distinction Whittaker and Sidner carried into email as frequent filers, spring cleaners and no-filers [20], and Fisher et al. extended with folder count as a second dimension [9]. A persistent finding across this literature is that filing is expensive and frequently unrewarded: Whittaker et al. found that elaborate foldering did not improve retrieval speed or success relative to search, and argued that much organisational effort is simply wasted [19]. Barreau and Nardi observed that users prefer location-based browsing to search when the structure is theirs [3]; Jones surveys the field this literature became [12], Kaye et al. show that personal archives are kept as much to construct identity and meaning as to retrieve [14], and Teevan et al. described *orienteering*, in which a target is reached through a series of small, contextually-informed steps rather than one precise query [18].

WhatsApp group chat is an extreme case: it offers no filing mechanism at all, so every user is structurally a piler. Our interest is in what orienteering looks like when there is no structure to orienteer through, and the only landmarks available are social and temporal.

Messaging as habitat and as archive

Ducheneaut and Bellotti's notion of email as a *habitat*, an environment in which people dwell and make meaning rather than merely a tool for dispatching tasks [7], transfers directly. O'Hara et al. make the equivalent argument for WhatsApp, emphasising the sense of continuous co-presence it sustains [16]. Cheng and Aflatoony's study of instant-messenger reappropriation is the closest antecedent to the present work: from 81 survey responses and 14 interviews they identify self-texting for saving and remembering, message dumping, and emotionally-motivated revisiting of old conversations [4]. Their framing of users' "struggles and strategies to find information buried in conversations" is the point of departure for this study.

We extend that work in three directions. We focus on *group* messaging, where volume and the absence of a stable interlocutor compound the problem. We situate the study in India, where group membership counts are substantially higher than those reported in Western samples. Several of our participants belonged to more than a hundred groups, and one to roughly three hundred. And we decompose the single category "difficulty finding information" into separable failure modes with distinct design remedies.

THE WHATSAPP GROUP INTERFACE

Because retrieval behaviour is shaped by what the client affords, we briefly characterise the interface as it stood during fieldwork in February 2026.

Structure. A group is a single flat, strictly chronological stream. There is no threading, no folder, no tag, and no user-definable grouping of messages within a chat.

Search. A global keyword search spans chats; a per-chat search is also available. Search is lexical and matches message text and file names. It does not match the contents of images, audio, video or stickers, nor does it offer a date-range or content-type filter in the result list.

Starred messages. Any message may be starred. Starred items appear in a single undifferentiated list, in chronological order, with no labels, folders, notes or sort options.

Media, Links & Docs. Each chat exposes a gallery of its media, segmented by month and ordered by date, with small

square thumbnails. Links appear as URLs; shared map locations appear under links without a visible place name.

Self-chat and self-groups. A user may message themselves, or create a group in which they are the only member. Neither is presented as a storage feature, but both are widely appropriated as one.

Two absences matter most for what follows. There is no mechanism to annotate or name a stored item, and there is no filter on time or type. Every retrieval therefore reduces to one of three moves: remember an exact word, remember an approximate date and scroll, or remember a person and navigate through them.

METHOD

The study ran over three weeks in February 2026 as part of a graduate Design Research module at NID Bengaluru. It combined contextual enquiry, a survey, and a sequence of synthesis methods drawn from the IDEO Design Kit. Research question formulation was iterative, and the reframing is itself a finding we report below.

Contextual enquiry

We conducted eleven semi-structured contextual interviews with postgraduate design students (aged 23–31; 5 female, 6 male) in their own classroom spaces during ordinary academic hours, so that WhatsApp use could be observed in its authentic setting, not recalled in a laboratory. Sessions ran 20–25 minutes. Each had three parts: a discussion of the participant's groups and habits; a live retrieval task in which the participant attempted to find a specific item (a photograph, an identity document, a voice note) on their own device while thinking aloud; and a closing discussion of frustrations and wished-for features.

Topic headings covered usage and context, information storage behaviour, retrieval practices, emotional and motivational aspects, group dynamics and social behaviour, deletion and storage management, and attention and distraction. We recorded verbal responses, the live task demonstrations, observed interaction patterns, and emotional reactions during retrieval attempts. The live task proved the most productive instrument: several participants described search as adequate and then, minutes later, failed to complete a retrieval in front of the interviewer.

Survey

A questionnaire of 24 items was distributed online and returned 59 valid responses. It captured tenure and frequency of use, group counts, storage and retrieval behaviour, hardest-to-retrieve content types, time spent on retrieval, verification and recall cues, and six attitudinal statements rated on a five-point Likert scale. The sample skewed young and heavily habituated (Table 1), which we treat as a limitation in a later section.

Synthesis

The synthesis proceeded through six methods: contextual enquiry (observation, task analysis and semi-structured interviews), literature review, affinity mapping, survey research, data pattern analysis, and persona development.

Transcripts and task observations were first coded openly. Every discrete action a participant described or was observed performing was written as a gerund on a separate card, yielding 110 distinct behavioural codes. These were clustered into nine families (Table 5), from which the recurring themes were drawn. The code set deliberately preserved negative and paired forms, so that *labeling* sits beside *not-labeling*, *starring* beside *ignoring-starred* and *forgetting-starred*, and *finding* beside *failing-to-find* and *giving-up*. Retaining the absence of an action as a code in its own right is what later made the failure modes separable: several of the seven are visible in the data only as a negative pair.

The code set then fed back into the instrument. Codes that were frequent but thinly understood were turned into questions for a second round of semi-structured interviews, which is how items such as *assuming* and *expecting* were probed directly ("what did you assume about data you thought was there, and could not find when you needed it?") and how *backing-up* and *resaving* were separated from ordinary saving. Key terms were additionally clustered through card sorting to surface recurring themes. We then built an empathy map across all eleven participants (think, say, do, feel, plus retrieval strategy, key challenges and suggestions), a behavioural variables matrix along four axes (activities, attitudes, motivations and skills), and from the clustering of those variables, four behavioural clusters and three composite personas. A hierarchical information storage-and-retrieval framework was developed in parallel; its branches determined the columns of the pattern-finding sheet. The seven failure modes reported in Table 3 are the result of collapsing the nine code families against that framework, so the taxonomy grew out of the coded data instead of being imposed on it.

Reframing the research question

The study began with a narrow hypothesis: that the chronological layout was the cause of burial. Interview data did not support so simple an account. Participants who were insulated from chronology, such as those who had built self-groups or who searched by person instead of by time, failed to retrieve at comparable rates, for different reasons. The question was therefore reframed from a single-cause formulation to an open one: *how do Indian users employ diverse personal information management strategies to facilitate storage and retrieval within WhatsApp's non-hierarchical interface?* We report the reframing because the shift from

Survey sample ($N = 59$)	
Age 18–25	68%
Age 26–39	24%
Age 40+	7%
Female / Male	56% / 42%
Using WhatsApp more than 8 years	46%
Using WhatsApp 5–8 years	29%
Open it "almost constantly"	49%
More than 10 times per day	31%
Member of 11+ groups	47%
Member of 26+ groups	23%
4 or more highly active groups	53%
Keeps a personal self-only group	63%
Needs to find old information at least "sometimes"	73%
Has asked someone to resend	93%

Table 1: Survey demographics and baseline usage. Percentages are of all respondents; a small number of items were left blank.

Statement	**Mean**	**Agree**	**Disagree**
Continuous incoming messages push important information out of visibility	4.11	71%	5%
The chronological message flow makes it difficult to retrieve older messages	3.70	55%	12%
Important information gets buried in active groups	3.65	58%	16%
I feel confident I can easily retrieve files shared weeks ago	2.93	32%	33%
WhatsApp provides enough structured tools to manage information	2.77	25%	40%
Scrolling is my main method of retrieving information	2.51	16%	51%

Table 2: Attitudinal items, five-point scale (1 = strongly disagree, 5 = strongly agree). "Agree" is the proportion rating 4 or 5; "Disagree" the proportion rating 1 or 2. $n = 56$–57 per item.

a single blamed feature to a distribution of failure modes is the study's principal analytic move.

RESULTS

Volume is not the complaint

Our participants inhabit a far denser group environment than the email literature describes. Interview participants belonged to a median of roughly 100 groups; the largest reported approximately 300, of which around 50 were active. Reported storage footprints reached 20–30 GB. The survey confirms the pattern at scale: 47% belonged to more than ten groups and 53% described four or more as highly active.

Yet volume itself was not experienced as the problem. This mirrors Grevet et al.'s finding that large inboxes did not produce a sense of disorganisation [10]. Only 16% of our respondents disagreed that important information gets buried, but the complaint, when probed, was never about quantity. It was about the asymmetry in Table 2: 71% agreed that incoming messages push information out of visibility, while only 32% were confident of retrieving a file shared weeks earlier, and 33% actively doubted they could. The distress is located at the moment of retrieval, not at the moment of accumulation.

Two further results in Table 2 deserve note. Agreement that WhatsApp provides adequate structured tools was low (mean 2.77; 40% disagreeing), which is unsurprising. More surprising is that scrolling scored lowest of all six items (mean 2.51), with 51% rejecting it as their main method. Participants do not think of themselves as scrollers. The live tasks showed otherwise: scrolling was the observed fallback in almost every failed retrieval. We read the gap as evidence that manual scrolling is experienced as a failure state rather than a strategy. It is something done once the method has already broken down, and is therefore not counted as the method. The coding supports this reading independently: *scrolling*, *endless-scrolling* and *date-scrolling* were coded adjacent to *failing-to-find*, *retrying* and *giving-up*, not alongside the deliberate search actions.

Seven ways retrieval fails

Coding the eleven interviews against the storage-and-retrieval framework yielded seven failure modes. They are separable: each has a different trigger, a different observable signature in the live tasks, and a different design remedy. Table 3 reports how many participants exhibited each.

Keyword rigidity (6/11)

Search succeeds only on an exact lexical match, so retrieval is gated on recalling the words, not the thing. The signature failure is a near-miss. One participant searching for a PAN card document received results for "Pantone"; the document itself was eventually located in a chat with her mother, and turned out to be corrupted. Another found a birthday photograph only by searching for the phrase "Happy Birthday" from a previous year and reading forward. Participants described search as adequate and then demonstrated trial-and-error query entry, revising terms three or four times.

> *"Search is generally sufficient, but similar keywords slow retrieval."* (P5)

Media non-indexability (6/11)

Images, video, audio and stickers carry no searchable content, so the entire non-text archive is reachable only by browsing. This was the most emotionally charged failure. Voice notes were the worst case: with no transcript, preview or duration cue, participants reported playing messages one

at a time until they found the right one. Several had stopped trying.

> *"Audio and photo search is frustrating."* (P7)

Photographs were little better. Retrieval meant scrolling a month-segmented gallery of small thumbnails, and participants reported browsing month by month. Stickers, which are expressive artefacts deliberately kept, became effectively unrecoverable once buried, since they appear in neither search nor a dedicated collection. The survey corroborates the ordering: audio messages were named hardest to retrieve by 49% of respondents, ahead of links and documents (29% each), stickers and text (25%), and photographs (24%).

Chronological burial (6/11)

Information is displaced by subsequent traffic. It is not lost; it is simply no longer where anyone will look. Participants described this most vividly around "message dumping": bursts of low-information traffic, of which birthday wishes were the recurring example, that push a logistical detail out of reach within minutes of its posting. Notably, participants reported that casual conversation survives dumping better than serious conversation, because emoji and reply-quotes provide visual landmarks that administrative text lacks.

Contextual decay (6/11)

The memory required to drive retrieval degrades faster than the archive does. This is the characteristic failure of self-groups. Participants used them intensively for several days and then forgot what was inside.

> *"Most important files must be somewhere in my chats."* (P11)

One participant travelled to a print shop expecting to produce a document from WhatsApp and discovered it was not there. Another had starred messages for months and could not say what was in the list.

Contextual fragmentation (5/11)

The user remembers the item but not its container. Because an item lives in whichever chat it was shared in, retrieval requires first reconstructing the social context of the sharing. Participants searched for photographs by recalling who they had been with and then opening that person's chat. This is a retrieval strategy indexed on relationships, and it fails entirely for items shared by someone forgotten.

Non-linear navigation deficit (6/11)

There is no way to cut into the archive along any axis other than text. The absent filters named most often were date range and content type; the comparison participants reached for, unprompted and repeatedly, was Google Photos.

> *"Date-wise filtering like Google Photos would help."* (P4)

Failure mode	**Interview** ($n = 11$)	**Survey signal**
Keyword rigidity	6	64%
Media non-indexability	6	49%
Chronological burial	6	58%
Contextual decay	6	68%
Non-linear nav. deficit	6	56%
Contextual fragmentation	5	41%
False-storage assumption	3	63%

Table 3: Failure modes observed per interview participant, with the nearest corroborating survey item: reliance on keywords as the primary recall cue; audio named hardest to retrieve; agreement that information gets buried; reliance on date or sender as a recall cue; reliance on date-checking to verify a result; reliance on file name as a recall cue; and maintenance of a self-only group.

False-storage assumption (3/11)

Least common but most consequential. Participants treated WhatsApp as reliable storage while knowing, when asked directly, that it was not designed to be. The belief is not idle: it displaces real backup. Participants had lost material to accidentally enabled disappearing messages, to a corrupted file, and to deletions made during storage-pressure clean-ups without a record of what was removed.

> *"WhatsApp is not designed for storage"* (P6), who nonetheless maintained three self-groups for identity documents, study material and saved content.

Parallel archives

The dominant storage strategy is to build a private archive inside the application. Thirty-seven of 59 respondents (63%) maintained a self-only group; among interview participants the practice was near-universal, and several maintained more than one, topically divided. Reported uses, in order: saving notes (36), storing documents (32), reminders (18), temporary storage (14) and transferring between devices (10). Participants also used them to draft messages, test stickers before posting, and perform arithmetic.

The self-group is best understood as a response to contextual fragmentation. It collapses the many-container problem into one container. But it inherits the flat chronological structure it was created to escape, and the archive it produces is unlabelled: *not-labeling*, *not-sorting* and *not-captioning* were coded far more often than their positive counterparts, with *captioning* observed in only one participant's practice. Contextual decay therefore sets in quickly. One participant described the pattern precisely: intensive use for three or four days, then no memory of the contents. The self-group moves the problem somewhere tidier and leaves it intact.

Starring shows the same arc. It was the single most-reported management action in the survey (37 of 59) and among the

least trusted in interview. Participants starred compliments, appreciation messages and credentials, and then did not return. One had revisited the starred list exactly once. Because the list is undifferentiated and unlabelled, starring functions as reassurance at the moment of saving rather than as a retrieval path. The behaviour appears in the coding as a sequence, not a single act: *starring* is followed by *ignoring-starred* and *forgetting-starred*, with *unstarring* rare. Saving and retrieving had to be coded as separate events precisely because, for this feature, the second so seldom follows the first. Those who had moved to self-groups described starring as redundant.

Outside the application, participants screenshotted (20 of 59), saved to device storage (13) and pinned chats (15). Screenshotting is worth noting as a deliberate format downgrade: participants converted searchable text into unsearchable image in exchange for moving it into a gallery they trusted more than the chat.

Social retrieval

The most consistent interview finding is that the terminal fallback is a person, not a feature. When scrolling and search both failed, participants messaged someone in the group for a summary, called a close friend to reconstruct what had happened, or asked the original sender to resend. Ninety-three per cent of survey respondents had asked someone to resend information they could not find; 59% did so at least sometimes. The social repertoire was rich enough to need its own code family: *asking*, *calling*, *clarifying*, *delegating*, *requesting-summary*, *explaining-context* and, at the far end, *outsourcing-memory*.

> *"When too many messages accumulate, I stop scrolling and ask my friends to summarise."* (P9)

This has two implications. Retrieval success is distributed across a social network, not resident in the interface, which means that measuring the performance of search alone understates how often the system has in fact failed. And it means the people best served are those with strong ties inside the relevant group. One participant, who described herself as introverted and rarely posted, said she would ask a friend to post a question on her behalf instead of posting it herself. The retrieval cost falls unevenly, and hardest on the least socially central.

Behavioural clusters

Clustering the behavioural variables produced four groups, summarised in Table 4. They are not mutually exclusive: four of eleven participants fell into two.

The division that matters for design runs between the first two. Personal Archivists invest effort at the moment of saving and are defeated by contextual decay; Social-Manual Searchers invest nothing at saving and are defeated at retrieval. A single feature will not serve both: better filing tools help only those already willing to file, and better search helps the archivist not at all when the problem is that she has forgotten what she filed.

Cluster	Defining behaviour
Personal Archivists	Build a manual database outside the chat flow: self-only groups, forwarding to a trusted contact to "tag" an item, credentials in starred messages.
Social-Manual Searchers	Ask people for summaries or resends; scroll the media tab for ten to fifteen minutes; call someone when the system fails.
Selective Minimalists	Cut noise instead of organising it: mute and archive, treat @-mentions as the importance filter, abandon a retrieval after roughly 30 seconds.
Nostalgic Hoarders	Treat the archive as an emotional timeline: keep 100+ groups to avoid giving offence, re-read old chats, refuse to delete despite storage warnings.

Table 4: Four behavioural clusters derived from the behavioural variables matrix ($n = 11$; membership overlaps).

Three composite personas were developed from these clusters (the Systematic Archivist, the Nostalgic Searcher and the Selective Minimalist) and are reported in the supplementary material.

Emotional attachment and the cost of deletion

Deletion was consistently described as harder than it appears in the storage-management literature. Participants kept 100+ groups they did not read because leaving would give offence; they refused to delete media they were attached to even under storage pressure; and they were acutely anxious about disappearing messages, which several had enabled by accident. One participant re-read old chats after a breakup and then cleared them deliberately, which illustrates the point from the other side: deletion was an emotional act, performed for emotional reasons, not a maintenance task.

The consequence for retrieval is direct. An archive that cannot be pruned keeps growing, which degrades every browse-based retrieval path, which increases dependence on search, which fails on exactly the non-text content that carries the emotional weight.

Scepticism about automated summarisation

Participants were divided on AI assistance in a way that tracked what they used the archive for. Those retrieving functional material wanted summarisation of unread messages and transcription of voice notes. Those retrieving emotional material did not. One participant objected that an AI summary would normalise expressive typography, a capitalised "OMG" for instance, and strip the feeling that made the mes-

Code family	Representative codes
Attention	checking, rechecking, selective-checking, priority-checking, ignoring, muting, skipping, leaving-unread, night-checking, hourly-checking
Navigation	scrolling, fast-scrolling, slow-scrolling, endless-scrolling, date-scrolling, skimming, scanning, jumping, backtracking, thread-following, reply-tracing
Search	keyword-typing, guess-searching, name-searching, year-searching, retry-searching, common-group-search, media-browsing, audio-playing, audio-replaying
Outcome	finding, refinding, locating, failing-to-find, giving-up, retrying, cross-checking, confirming, verifying
Saving	saving, resaving, screenshotting, forwarding, self-forwarding, downloading, backing-up, restoring, starring, unstarring, ignoring-starred, forgetting-starred, pinning
Organising	sorting / not-sorting, labeling / not-labeling, captioning / not-captioning, archiving, hoarding, accumulating
Memory	forgetting, remembering, recalling, misremembering, assuming, expecting, depending, outsourcing-memory
Social	asking, calling, clarifying, delegating, requesting-summary, explaining-context, withdrawing, lurking, silencing-self, hesitating
Affective	venting, not-sending, drafting, redrafting, emoji-reacting, sticker-reacting, contextless-replying, conflict-reading, gossip-reading, nostalgia-reading

Table 5: The nine families into which the 110 open behavioural codes were clustered, with representative codes from each. Paired and negative forms were retained deliberately; several failure modes are visible in the data only as a negative pair.

sage worth keeping. What they objected to was narrower than automation. A summary that keeps the propositional content and throws away the affect fails the use they actually have for the archive. For an archive used as an emotional timeline, that is a lossy compression of exactly the wrong channel.

DISCUSSION

Overload has moved from status to retrievability

Whittaker and Sidner located overload in message status [20]; Grevet et al. relocated it to type, and found unread counts, not archive size, to predict felt disorganisation [10]. In group messaging we find a third locus. Our participants were largely indifferent both to archive size and to unread counts, several carrying thousands of unread messages with no distress, and acutely sensitive to one thing: the gap between believing something had been kept and being able to produce it. We propose *retrievability* as the operative dimension of overload in chronological messaging, and the subjective confidence gap in Table 2 as its most direct measure.

One structure doing two jobs

It would be a misreading of these findings to recommend abandoning the chronological stream, and our own initial framing made that error. Chronology is what makes a conversation legible as a conversation; it is the correct structure for the primary task. The problem is that it is the *only* structure, and it is therefore made to serve a secondary task, reference and retrieval, for which it is actively hostile. The design question is not how to replace the stream but how to lay a second, non-temporal index over it.

Design implications

Index the non-text archive. Media non-indexability was the most commonly cited failure and produced the strongest negative affect. On-device transcription of voice notes, object and face clustering in media galleries, and OCR over shared images would each convert an unsearchable region of the archive into a searchable one. Participants asked for these unprompted; the Google Photos comparison recurred across interviews.

Add axes other than text to search. A date-range filter and a content-type filter would support the orienteering behaviour [18] participants already attempt and currently execute by hand. Both were requested explicitly and repeatedly.

Make saving produce a retrieval path. Starring and self-groups both fail because saving produces no durable handle. Permitting a label, note or folder at the moment of starring, and surfacing starred items with their originating context, would convert reassurance into recall. This is the one place where the email literature's remedies transfer almost unchanged; Rodden and Leggett's work on giving Gmail labels the affordances of folders is a direct precedent [17].

Support the social fallback instead of ignoring it. Since asking a person is the dominant repair, there is a case for making it cheap and legible: a lightweight way to request a resend of a specific item, or to mark a message as the answer to a question asked earlier in the thread. The social channel should not be treated as a workaround to be designed away; it works, and it carries context the system does not have.

Design summarisation for two different archives. Functional and emotional retrieval have different tolerances for lossy compression. Summarisation should be available where the user is recovering logistics, and should not be the default where the user is revisiting a relationship.

Make the storage assumption honest. Participants rely on WhatsApp as a document store while the interface declines to acknowledge the role. Either the affordance should be sup-

ported as a real, named, verified store, with deletion confirmation for items marked important, or the reliance should be visibly disconfirmed. The present middle position is the one that loses people's documents.

LIMITATIONS

Our interview sample is small ($n = 11$), homogeneous and non-random: all were postgraduate design students at one institution, aged 23–31, and all were heavy users. Design students are plausibly more articulate about interface frustration and more inclined to propose remedies than a general population; the specificity of the feature requests in our data should be read with that in mind.

The survey sample ($N = 59$) is larger but skews young (68% aged 18–25) and habituated (75% using WhatsApp more than five years), and was recruited through the authors' own networks, which are concentrated in the same academic environment. The group counts we report are therefore unlikely to generalise to older or less connected Indian users, let alone to other markets, and our framing of self-groups as near-universal may be a feature of this population.

All retrieval times are self-reported except within the live tasks, where the sample of observed retrievals is small and the presence of an observer plausibly increased effort: participants persisted with searches they told us they would normally abandon. We did not instrument devices, so we cannot report actual message volumes, archive sizes or search success rates, and our storage figures are participants' own readings of their settings screens.

The study is a snapshot, following Whittaker and Sidner [20] where Fisher et al. [9] tracked behaviour over time. Retrieval behaviour plausibly changes across the semester, since submission deadlines, festivals and events generate both message dumping and retrieval demand, and we cannot speak to that variation. The open coding reported in Table 5 was carried out by a single researcher without a second coder, so no inter-rater reliability can be reported, and the collapse from 110 codes to nine families to seven failure modes involves judgements that another coder might have made differently. Finally, the taxonomy in Table 3 was derived from the same interviews it describes and has not been validated against an independent sample; the survey corroboration in the right-hand column is indicative, since the survey was not designed to test categories that did not yet exist when it was written.

CONCLUSION

Thirty years after email inboxes were found to be doing work they were never designed for, mobile group messaging is doing the same work with fewer tools. We studied how users of high-volume WhatsApp groups store and re-find information, through eleven contextual enquiries with live retrieval tasks and a survey of 59 users.

We find that overload in this setting is not a volume problem. Participants were comfortable with very large archives and with thousands of unread messages, and uncomfortable with one specific thing: not being able to produce what they were confident they had kept. Only 32% were confident of retrieving a file shared weeks earlier. Retrieval fails in seven separable ways, of which the inability to search non-text content and the brittleness of exact-match keyword search are the most consequential. Users respond by building parallel archives, most often a self-only group, which reproduce the flat, unlabelled structure that caused the problem, and by falling back on other people, which works but distributes the cost unevenly onto those least socially central in a group.

The implication for design is that a messaging client which has become a habitat needs a second index. Chronology should stay; it is what makes conversation legible. What is missing is any other way in.

ACKNOWLEDGMENTS

Thanks to the eleven participants who let their chat histories be examined in public, and to Dr. Mamata N. Rao, who guided the Design Research module at NID Bengaluru, for direction on hypothesis formulation and on the card-sorting synthesis. All participants are referred to by code; no participant is identified by name.

REFERENCES

1. Bellotti, V., Ducheneaut, N., Howard, M., and Smith, I. Taking email to task: The design and evaluation of a task management centered email tool. *Proc. CHI*, 2003, 345–352.
2. Bellotti, V., Ducheneaut, N., Howard, M., Smith, I., and Grinter, R. Quality vs. quantity: Email-centric task management and its relationship with overload. *Human-Computer Interaction* 20, 1–2 (2005), 89–138.
3. Barreau, D., and Nardi, B. Finding and reminding: File organization from the desktop. *SIGCHI Bulletin* 27, 3 (1995), 39–43.
4. Cheng, C.W., and Aflatoony, L. The reappropriation of instant messaging: Texting ourselves, message dumping, and revisiting conversations. *Proc. ACM Hum.-Comput. Interact.* 6, MHCI, Article 187 (2022), 23 pages.
5. Civan, A., Jones, W., Klasnja, P., and Bruce, H. Better to organize personal information by folders or by tags?: The devil is in the details. *Proc. ASIS&T*, 2008.
6. Church, K., and de Oliveira, R. What's up with WhatsApp? Comparing mobile instant messaging behaviors with traditional SMS. *Proc. MobileHCI*, 2013, 352–361.
7. Ducheneaut, N., and Bellotti, V. Email as habitat: An exploration of embedded personal information management. *Interactions* 8, 5 (2001), 30–38.
8. Dabbish, L.A., and Kraut, R.E. Email overload at work: An analysis of factors associated with email strain. *Proc. CSCW*, 2006, 431–440.
9. Fisher, D., Brush, A.J., Gleave, E., and Smith, M. Revisiting Whittaker & Sidner's "email overload" ten years later. *Proc. CSCW*, 2006, 309–312.
10. Grevet, C., Choi, D., Kumar, D., and Gilbert, E. Overload is overloaded: Email in the age of Gmail. *Proc. CHI*, 2014, 793–802.
11. Hogan, B., and Fisher, D. A scale for measuring email overload. *Microsoft Research Technical Report*, 2006.
12. Jones, W. *Keeping Found Things Found: The Study and Practice of Personal Information Management*. Morgan Kaufmann, 2007.
13. Malone, T.W. How do people organize their desks? Implications for the design of office information systems. *ACM Transactions on Office Information Systems* 1, 1 (1983), 99–112.

14. Kaye, J., Vertesi, J., Avery, S., Dafoe, A., David, S., Onaga, L., Rosero, I., and Pinch, T. To have and to hold: Exploring personal archives. *Proc. CHI*, 2006, 275–284.
15. Mackay, W. More than just a communication system: Diversity in the use of electronic mail. *Proc. CSCW*, 1988, 344–353.
16. O'Hara, K., Massimi, M., Harper, R., Rubens, S., and Morris, J. Everyday dwelling with WhatsApp. *Proc. CSCW*, 2014, 1131–1143.
17. Rodden, K., and Leggett, M. Best of both worlds: Improving Gmail labels with the affordances of folders. *Proc. CHI Extended Abstracts*, 2010.
18. Teevan, J., Alvarado, C., Ackerman, M.S., and Karger, D.R. The perfect search engine is not enough: A study of orienteering behavior in directed search. *Proc. CHI*, 2004, 415–422.
19. Whittaker, S., Matthews, T., Cerruti, J., Badenes, H., and Tang, J. Am I wasting my time organizing email? A study of email refinding. *Proc. CHI*, 2011, 3449–3458.
20. Whittaker, S., and Sidner, C. Email overload: Exploring personal information management of email. *Proc. CHI*, 1996, 276–283.