Information Structuring and Prioritization for Knowledge Collaboration in Online Communities

By

JINGXIAN LIAO

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

COMPUTER SCIENCE

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Hao-Chuan Wang, Chair

Kwan-Liu Ma

Wen-Chieh (Steve) Lin

Committee in Charge

2024

ABSTRACT

Information Structuring and Prioritization
for Knowledge Collaboration in Online Communities

In the last decade, the proliferation of online collaboration has vastly broadened the landscape of knowledge work. Individuals form online communities and accomplish complicated intelligent work together, such as developing software or planning healthy diets. These collaborations vary in levels of joint interaction and collaboration strategies, partially dependent on the technologies as well. While technological advancements have enabled large-scale flexible participation, they have also resulted in a proliferation of unorganized and overlapping knowledge contributions. Also, it may hinder misleading information from separating from credible knowledge. This overwhelming abundance and disorganization pose considerable challenges both for individuals and the community to leverage the knowledge to solve problems and make high-stakes decisions. These tasks demand not only cognitive resources but also advanced meta-cognitive skills. These obstacles highlight the need to concentrate on structures and distilling knowledge, from creation and integration to dissemination.

In my thesis, I explore the designs and understand the impacts of computational and visualization scaffolds that structure and highlight relevant and semantic knowledge, taking into account the socio-cognitive dynamics among collaborators. I aim to use technological support and social-cognitive mechanisms to mitigate cognitive and attentional limitations in online community-based cooperative work.

To help individuals actively explore and understand knowledge shared in online communities, such as videos, we first dive into the sharing from both video producers and audiences, and explore the idea of structuring semantic representations of video contents and audience comments, which aids in discerning high-quality videos and supports diverse video exploration compared to conventional video watching experiences. Empirical study results also illuminate the adoption and priority of structured overview for interpretation. Moving the understanding forward, accessing relevant knowledge doesn't guarantee analysis and knowledge integration. We investigate social nudging approaches to scaffold engagement in higher-order thinking for high-stakes topic analysis, and compare the influences between common documentation tools and a concept-mapping-based space which also plays as thinking scaffold, DeepThinkingMap. Two lab studies reveal the effectiveness of social nudging in fostering both reflective and critical thinking, and confirmed the synergetic effect of nudging with other thinking scaffolds. Finally, we shift focus to synchronous video-based interactions and non-verbal cues. Our secondary analysis of group brainstorming sessions demonstrates the significant impact of metaphoric hand gestures on both individual and partner creativity and found that the positive effect of metaphoric gestures is independent of media richness of communication medium.

In conclusion, this dissertation underscores the potential of computational and social support in reshaping how knowledge is explored, integrated, and co-created within online communities. Based on the empirical findings about the socio-technical-cognitive mechanisms and the design space, this dissertation paves the way for future research that promotes organized, reflective, and efficient knowledge collaboration in online communities.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This thesis would not be possible without the tremendous support of numerous brilliant people. Their invaluable contributions, support, and guidance have been the cornerstone of this journey. With profound gratitude and immense joy, I dedicate this work to each of them:

- My advisor, Hao-Chuan Wang, whose unending support over the past six years has been nothing short of transformative. His guidance and encouragement have been instrumental in my growth as a researcher and an individual.
- My dissertation committee and qualifying exam committee members, Kwan-Liu Ma, Steve Lin, Katherine Kim, and Michael Neff for their insightful advice and constructive feedback.
- Professor Vladimir Filkov and Prem Devanbu, who sparked my curiosity in research and computer science.
- My mentors, including David Kavaler, Wei Wang, Fu-Yin Cherng, and Nick Furlotte, among others, who have guided and supported me through various stages of my academic and professional development.
- My collaborators, colleagues, and mentees, notably Jason Xue, Xue Han, Farhad Hormozdiari, Mrinalini Singh, Yi-Ting Hung, Yi-Hung Lin, Guowei Yang, and many more, for their assistance and collaboration throughout my research and internships.
- My family — parents, grandparents, Uncle Wei, Aunt Dongyu, Shaoyang, and other cherished members — who have been my unwavering support system, regardless of the distance between us.
- My lab mates and friends—Jingchao, Xiaoyang, Yuanyuan, Shu, Kai-Hui, Xiaoyu, Mengyu, and Aoran—for their camaraderie and support, helping me navigate both the triumphs and challenges of research and life.
- My lifelong friend, Zijing, with whom I have shared the journey of PhD, offering each other support and companionship.
- Last but not least, to all the participants in my user studies and interviews, whose involvement has been invaluable to this work.

Chapter 1

Introduction

1.1 Motivation

Over the past decade, knowledge work collaboration has continued to embrace more diverse opportunities. The rise of peer production and other online collaboration strategies, fueled by technological advancements and the trend towards decentralized work environments, has reshaped how we approach collaborative tasks [25]. Unlike the traditional, organization-centric models of collaboration, the modern landscape of online, non-organizational collaborative work varies widely. It ranges from loosely organized cooperative efforts to well-structured and meticulously managed projects [110]. This spectrum is marked by a diverse array of collaboration models and tasks, each supported by technologies tailored to specific needs [110, 126]. These technology affordances vary in terms of project openness, power and organization, communication modality, and contribution pipelines. For instance, video conferencing tools like Zoom [1] and Microsoft Teams [2] are used in hybrid town hall meetings to foster social awareness among attendees. For design collaboration, online platforms such as Miro provide a comprehensive space for storing and communicating various project elements like diagrams, sketches, and documents. When we expand our view of knowledge collaboration to the asynchronous online sharing and reuse of knowledge, as seen in platforms like Stack Overflow, knowledge is also co-constructed and shared in back-and-forth Q&A communications [271].

However, while opening doors to more diverse collaborative opportunities, these technological advancements also introduce significant challenges to individuals and communities, particularly in community-based knowledge collaboration. In such collaborations, participants contribute flexibly based on their individual skills and availability, and share their knowledge in

[1] Videotelephony software developed by Zoom Video Communications, Inc., `https://zoom.us`

[2] Business communication platform developed by Microsoft Teams, `https://www.microsoft.com/en-us/microsoft-teams/group-chat-software`

free forms, such as collaboration in online forums (e.g., Stack Overflow) and designed collaborative spaces like GitHub [26, 77]. The lack of centralized control and the diverse range of inputs from numerous voluntary participants often result in an overwhelming amount of unorganized information available. People often struggle to filter crucial information and remove duplicated content, leading to potential delays in comprehensive information collection. Also, the increase in information modalities may not lead to informational gain. For instance, in streamlined video meetings, prolonged exposure to nonverbal cues can lead to "Zoom Fatigue", manifesting as overload and attention fatigue [17]. Such clutter of knowledge will further complicate the process of integrating and synthesizing due to the scale, which demands extra inference and association that may not be readily available for people. Moreover, a lack of organization and oversight in knowledge sharing can make more space for poor-quality and misleading content [149], as it is difficult for individuals to associate with accurate and relevant knowledge to verify and assess. These challenges sap cognitive effort and skills from individuals and groups, across multiple cognitive processes including understanding, analysis, evaluation, and creation [12]. They highlight *the critical need to organize and structure knowledge* to support knowledge retrieval, integration, and exchange processes in collaborations through the integration of automation and the enhancement of human capabilities.

Addressing these challenges, this dissertation proposes to *structure and prioritize crucial knowledge to mitigate the cognitive barriers to knowledge collaboration in online communities via computational technologies and social mechanisms*. As Drucker noted about knowledge mediation interfaces, an interface is not merely a gateway to information but an integral part of the information itself [67]. Technology affordance and designs for collaboration interfaces can not only facilitate access to knowledge but also shape individuals' thinking processes and collaboration paradigms [222]. I [3] develop computational and visualization scaffolds that structure and accentuate critical knowledge, and apply these scaffolds considering socio-cognitive mechanisms among collaborators. We incorporate computational technologies to automate, streamline, and optimize the processing and presentation of knowledge, thereby enhancing efficiency and alleviating human effort. Concurrently, we engage socio-cognitive mechanisms to more effectively harness collective wisdom and

[3]Although this dissertation is based on studies that I led, it consistently uses the pronoun "we" in acknowledgment and appreciation of the invaluable contributions made by my collaborators to the projects.

amplify individual capabilities. More specifically, my studies explore and understand which elements to prioritize and how to structure them in three key tasks that are cognitively interrelated and span different stages of knowledge work [12].

- **For knowledge retrieval and understanding**: We structure information into conceptual graphs across contribution sources through computational scaffolds to support active exploration and understanding
- **For knowledge integration and analysis**: Our approach combined structured interactions and social nudging, to support reflecting and critical thinking for high-stakes topic analysis and review.
- **For knowledge sharing and co-creation**: We prioritize semantic hand gestures in video-based conversations via real-time annotation, facilitating creative knowledge co-creation.

This work aims to explore and understand the socio-cognitive-technical mechanisms and their impacts on knowledge processing and social experiences in community-based collaborations. Investigating the interplay of social mechanisms, technology interventions, and knowledge production sheds light on the support for collaborative intelligent projects to empower individuals and communities, to unlock the full potential of community-based knowledge collaboration.

1.2 Scope and Research Method

This dissertation aims to explore potential knowledge structuring scaffolds, tailored to specific knowledge tasks and social interaction approaches. For a clear presentation, this section introduces the research scope of this dissertation and the main research method that guides the dissertation.

1.2.1 Knowledge Work and Online Knowledge Collaboration

Within the context of this dissertation, "knowledge" refers to a subset of information that is relevant, actional, and partially based on personal experiences, which may be subjective to individuals [154]. According to Polanyi, explicit knowledge is typically objective, articulate, and transferable through formal means, such as documents, videos, notes; while tacit knowledge is unconscious and may be transferrable through socialization [199, 218]. In this dissertation, we focus on explicit knowledge and help humans articulate their knowledge explicitly, which makes it accessible to others for collaborative work with appropriate prioritization.

To summarize the pharses in knowledge work, it's not easy to find a concise and clear one although the concept has attracted scholars across domains for decades and the publications have increased in recent years [220]. Multiple knowledge work models have been proposed from different dimensions, such as the SECI (socialization, externalization, combination, and internalization) model by Nonaka and Takeuchi which emphasizes the interactions between explicit and tacit knowledge and humans [198, 199], Bloom's taxonomy that classifies knowledge stages in educational and cognitive context [12], and Knowledge management life cycle model that focuses on organizational knowledge management [188]. Synthesizing these frameworks, we may distill knowledge work into the following stages: retrieve and filter information, organize and synthesize knowledge, analyze and evaluate knowledge, and share and create knowledge. In Section 2.1, we will elaborate on knowledge work by stages and the related research.

When knowledge work becomes interdependent and includes a group of distributed knowledge workers, it emphasizes on the processing of sharing, transfer, accumulation, transformation, and co-creation of knowledge [77]. Nowadays, knowledge collaboration often relies on information technologies. Thus, the way knowledge is organized, mediated, and shaped by the affordances of technology and interactions among users plays a pivotal role in the efficiency and effectiveness of knowledge collaboration [92, 93, 201]. My work leverages this opportunity to explore computational and visualization designs to support online community collaborations.

1.2.2 Community-based Collaboration

In parallel with organizations, online communities may form another important online collaboration group, community-based collaboration, in which open collectives of individuals contribute towards knowledge or information goods [25]. Individuals participated in the collaboration in an informal and often loosely structured way toward specific goals, depending on the collaborative forms. Besides clear collective goals, in some communities, collaboration occurs as a side product along with other primary focuses or individual goals (such as social bonding and individual knowledge sharing), often seen in in social media and discussion forums [77, 180].

The flexibility of such community-based collaboration leads to differences in multiple dimensions. Individuals in these communities often harness a diverse range of backgrounds, expertise, and participation motivations [26]. Through the lens of groups, community members share diverse levels of social network ties, decided by collaborative patterns and tasks. When the granularity of the unit contribution gets small, collaboration can operate by garnering discrete contributions from individuals who remain independent or have limited coordination with others [263]. By contrast, some community collaboration could also expect strong ties, which include regular personal interactions with other community members, self-disclosure, and commitments, such as core members of GitHub projects [26]. Haythornthwaite categorizes them into two primary collaborative patterns: heavyweight peer production, exemplified by Wikipedia and open-source software projects in which collaborators engage with others and have loose cooperative structures and social connections; and lightweight peer production, a crowdsourcing model where individuals are unconnected and participate in the micro-level [110]. There are also collaboration patterns between these two patterns, and dual-weight collaboration that blend these approaches to contribute to the same project goals. Individual performances and behaviors can be influenced by the differences in these dimensions and other factors in community-based collaboration such as trust and social transparency.

In this dissertation, we focus on community-based collaboration in commonsense knowledge work in which members are unacquainted and require no domain expertise beforehand. It covers a wide range of community-based collaboration scenarios and includes most online users who need support. Unlike crowdsourcing, which includes a set of authoritative individuals to enact task decomposition and implement technologies to support the deposit and integration of individual contributions, this dissertation is interested in natural collaboration without explicit hierarchy. In Chapter 4, we focus on the structured additions of personal thoughts and knowledge from a group of general users. Chapter 5 considers the real-time conversations between unacquainted collaborators for knowledge co-creation. Through knowledge structuring and prioritizing support design and experiments, we aim to help reconcile the nuances in communication-based collaboration, and provide design implications for building knowledge collaboration in online communities.

1.2.3 Research Method

As mentioned above, this work adopts a multidisciplinary approach, leveraging insights from social psychology and cognition, design, and computational technologies such as AI techniques. Inspired by the balanced integration of empirical observations and theoretical paradigms, this work adopts the framework proposed by Mackay and Fayard [177] to structure, reform, and evaluate the interaction of individuals with other individuals and technology artifacts moving between theory and empirical observation. We draw from theoretical paradigms from social cognition, computer-mediated communication, and learning science. Then we instantiate the knowledge structure and re-prioritize ideas into computational and interface experimental prototypes. We used lab experiments to understand the influences introduced by our new approaches through qualitative and quantitative analyses. Some of the work echoes back and contributes to understanding the nuanced dynamics of knowledge collaboration, and some focus on design implications for researchers and practitioners. We will elaborate on the contributions of each exploration in the following chapters.

1.3 Dissertation Overview

This dissertation is structured around a cohesive series of three projects, each contributing crucial insights to the overarching research opportunity across different knowledge tasks. For the rest of this dissertation, I first introduce the background, related work about tasks in knowledge work and collaboration, recent innovations in scaffolding the processes, as well as social and cognitive theories we used to understand community-based collaboration and guide our designs (Chapter 2). Then we demonstrate the knowledge structuring and prioritizing support for three interrelated key tasks in Chapter 3 - 5. [4]

Chapter 3 takes the YouTube community as an example to illustrate the opportunities of knowledge structuring for knowledge exploration among videos shared online, through computational information structured summary across two information sources. People spend significant time on passive linear viewing through recommendations, such as watching YouTube videos with

[4] Portions or some versions of these chapters have been previously published in peer-reviewed conferences and journals. The co-authors of these publications have granted their consent for the inclusion of projects in the present dissertation.

"playing next"; however, it risks hindering active knowledge assimilation, calling for a balance between active exploration and algorithmic recommendations. Interactions with peer audiences via comments provide both informational and social values that may support active exploration processes, but they are not as accessible to users before watching as video content and metadata. We designed the Kentaurus system, which blends semantic representations of both audience comments and video content into middle-level structured content overview to afford flexible exploration. Users are supported to explore videos in a top-down approach to foraging the video pools throughout and organizing their own viewing paths. In our study, Kentaurus aided participants in selecting higher-quality videos during the exploration, and helped participants sketch the conceptual overviews and understand the associations between knowledge objects. Meanwhile, we observed different knowledge filtering and selection strategies based on personal interests, backgrounds, and trust. A portion of the research featured in this chapter was previously published in Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing [157].

The first project helped individuals to understand the knowledge from the overwhelming information. However, access to appropriate knowledge doesn't always lead to further knowledge integration and analysis processes due to high demands on cognitive resources and thinking skills. Moreover, knowledge from user-generated content needs analytical review before integration. The following Chapter 4 explores social nudging for engagement in higher-order thinking with peers' thinking notes and triggers to structured thought sharing. We implemented a shared video-watching and documentation interface, DeepThinkingMap, to help collaborators document their interpretation of videos and personal thoughts and scaffold thinking in a conscious approach. We did two lab studies for asynchronous sequential and synchronous cooperative video reviews. The results showed that participants' engagement in both reflective and critical thinking was enhanced with social nudging. They also shared more analytical and reasoning thoughts during the interactions. These empirical observations also contribute to the understanding of social nudging techniques mixed with conscious thinking interventions as nudge plus. Portions of work introduced in this chapter were published in Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems [160] and Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing [158].

The preceding chapters primarily focus on recorded video resources and text-based collaboration around pre-recorded videos for tasks related to knowledge forage, interpretation, and analysis. With the advent of low-cost, widely accessible video-based communication, real-time video conversations have become increasingly prevalent. These synchronous video-based collaboration meetings bring forth new challenges, particularly in capturing and generating vital nonverbal cues that supplement verbal communication. As communication channels become more sophisticated, there is a growing need for additional support to help individuals prioritize and process specific cues essential for effective knowledge sharing and co-creation. In Chapter 5, we present a secondary analysis focused on the impact of hand gestures on fostering creative collaboration. This analysis is based on an experimental dataset that records performances and individual behaviors among previously unacquainted collaborators engaged in group brainstorming sessions. Our data analysis reveals that the use of metaphoric gestures, which convey abstract concepts with spatial cues, significantly boosts both personal and partner creativity performance. Furthermore, our findings indicate that the frequency of impactful gestures in video-based interactions is independent of communication media. These findings offer design implications for video communication support to prioritize semantic-rich non-verbal cues, particularly in enhancing creative knowledge co-creation. A rendition of the research in this chapter was formally published in Proceedings of the ACM on Human-Computer Interaction, Volume 3, Issue GROUP, 2020 [159].

In conclusion, Chapter 6 reflects the contributions and insights gleaned in this dissertation and discusses the potential directions for future research that may build upon these findings.

Chapter 2

Background and Related Work

In the complex landscape of online knowledge collaboration in online communities, understanding the intricate interplay of technical, social, and cognitive factors becomes imperative [234]. The nuanced influences can be found when we delve into the realms of cognitive processing of individuals. In this chapter, we first discuss different knowledge tasks, as well as the challenges and affordances presented by technology for knowledge work. Drawing upon rich theoretical underpinnings, we will briefly introduce the groundwork of relevant social factors that may affect online collaboration from socio-cognitive mechanisms.

2.1 Knowledge Work by Phases

As introduced in Section 1.2.1, knowledge work mainly involves capture, understanding, analysis, distribution, and creation of knowledge, where these phases are interconnected in the practices [12, 188, 199]. Recent work has also studied the learning and validation processes for knowledge on the web, such as misinformation detection [7] and informal learning [193, 258] when focusing on the consumption side. In this section, we present a succinct overview of these phases and related work into three sub-topics below.

2.1.1 Knowledge Forage

Knowledge forage here consists of searching and navigating the vast amounts of information and capturing knowledge that is most relevant to specific needs or goals. People make a series of decisions about whether and how to check the content, where to look for, when to switch to others, etc [215]. They tend to spend a significant time on it, especially as information explodes and forms of information enrich [182]. Prior studies reported that it turned cognitively taxing for individuals and groups [22, 210], let alone the efficiency and accuracy. There has been a wealth of research and commercial support designed to facilitate users or automate the task. General-purpose information retrieval techniques, for instance, aim to find the most relevant information based on the search query, which is refined into coarse-grained retrieval and ranking in commercial search algorithms [11]. Some approaches, tailored for specific purposes (e.g., learning) or professional domains (e.g., design), leverage domain knowledge and specialized processes to streamline the retrieval process. These methods often integrate with domain users' processes and strategies to facilitate individuals and enhance personalized recommendations, thereby reducing the workload on the non-expert users [279].

During the process, however, individuals and groups may still face different challenges, stemming from cognitive capacity, background knowledge, and technology affordances. In the early phase, people often need to view broadly with extra cognitive burdens to formulate a comprehensive overview of the information scope to start with [140], and then gradually develop the mental model which subsequently guides the refinement of search directions and pathways. Especially when people engage in open-ended or not clearly defined goals in exploratory knowledge retrieval (as opposed to lookup search, e.g., finding facts and visiting certain sites), they often need to iteratively search and browse to refine their understanding and define their needs, which can't be replaced by automatic algorithms [183]. The process is often influenced by algorithmic biases, leading users to more readily available knowledge while struggling to identify and integrate peripheral information. The lack of exposure to a diverse range of information can inadvertently foster limited perspectives and potentially biased decision-making [113]. Moreover, when people lack sufficient background knowledge, another paradox arises in that they don't know the efficient queries or related terms to search or gauge the relevance of the content encountered.

To address these challenges, previous research has investigated multiple approaches, including the development of tools designed to assist users in extracting essential information from webpages for subsequent interpretation, as well as the creation of algorithms capable of evaluating a vast array of search results to discern content of relevance based on various criteria. For example, research and commercial products such as Fuse, Unakite, and Notion, allow users to extract the selective web within clicks first leveraging HTML structures and Natural Language Processing (NLP) techniques, and then to annotate and organize them into structured knowledge later [144, 167, 270]. Advanced search and navigation algorithms use summarizing techniques and semantic comparison to extract the summary of topics and add faceted search filters based on metadata and semantic categories [3, 44, 135, 193]. For scholarly articles, for instance, Semantic Scholar provides "TLDR" and estimates "the fields of the study" and "publication types" of papers, as well as supports filtering by relevance and publication period [3]. Research systems like Connected Papers and Apolo, utilize citation relationships to help researchers find and visualize related work [44]. Besides textual materials, to support search among videos, multiple systems have also been explored to capture the sub-topics in the videos and sequential relationships between topics using the narrations, images, and video metadata, and provide easy-to-connect visualizations to guide users, especially novices [135, 193].

2.1.2 Knowledge Integration and Verification

Following the browsing, collecting, and extracting information, individuals and groups move to integrate pieces of knowledge into structured views that are accessible for diverse tasks. They often encompass in-depth interpretation, presentation, or the transfer of knowledge, aimed for particular objectives, such as learning, problem-solving, and decision-making. A related field in Library and Information Science is knowledge organization (KO), which focuses on the processes of activities such as document description, indexing and classification and related knowledge organizing systems used [116]. Many knowledge integration innovations could be traced back to the pioneering visions of Vannevar Bush before the boost of computer science in this century [222].

Depending on the context and goals of the integration, many approaches are used to synthesize the raw knowledge, including classification and grouping, graphs, wikis, panels, and others [283]. Here, we provide a concise overview of two methodologies pertinent to this dissertation: classification and graph.

Classification is one simple systematical approach for knowledge organization and integration [145], and the task has notably evolved from being predominantly human or crowd-driven to increasingly relying on machine intelligence. Prior studies have incorporated classification options into the browsing and organization stages for individuals and communities to leave tags and other customized labels. For instance, online websites such as LibraryThing and MovieLens allow users to catalog their reading and viewing collections based on unique systems and categories [13, 237]. Meanwhile, researchers found that tags in these communities may also function as facilitators of social navigation, enabling individuals to share and discover knowledge contributed by others, which is intimately connected to knowledge sharing discussed in the subsequent section, and map the identities of the online community itself [13, 194]. Regarding automatic support, research systems leverage state-of-the-art models in artificial intelligence and machine learning to streamline and enhance the classification process for people. For instance, ForSense used natural language understanding models to suggest themes and groups for webpage clips [221]; Paper-Poles clustered and visualized papers according to the relevance and citation network among papers [111]. classification task has also become a cornerstone in the ongoing journey of other knowledge-related complex tasks, boosting the collaboration between human ingenuity and machine intelligence.

As classification and groups help integrate knowledge into distinct clusters, another structure – graph, including networks, trees, concept maps, helps identify the relationships and connections between different pieces of knowledge. Overall, the visual representation of a graph can scaffold users to organize content consumption and interpretation. Ssystems provides free-form or structured features to support individuals and teams to create and share online concept maps or similar graphs-like documents. CmapTools, and the following designs including R-CMap enable users to establish links to different types of resources (e.g., documents, videos, audios) and add interaction designs for clear visualization [19, 39]. Apolo also visualized literature sets based on citation networks and progressively expanded the paper scope to reduce the information overload [44]. Some work provides additional help, such as adaptive feedback and expert templates, to improve the quality of maps and learning performance [115, 278]. For instance, MindDot monitored the concept mapping process and provided adaptive feedback to help learners develop comparative strategies and improve reading outcomes [278]. Kit-Build concept map assessed users' learning status based on partial or full decomposition of instructors' concept map, to support both learning and teaching [115]. However, as concept mapping can be time-consuming and challenging for non-experts, there are also many attempts to generate concept maps or graphs automatically in another vein, so that people can follow the learning paths or gain a comprehensive outline. For instance, ConceptGuide provided concept-map-based graph visualization that generated recommendations of learning paths for videos on generic video platforms for informal learning [258]. For text-based materials, many studies used data mining, named-entity recognition (NER), relation extraction, and others to discover the associations between concepts [66].

In conjunction with these studies and systems, scholars also work on the accuracy and trustworthiness of knowledge synthesized. With the prevalence of user-generated content, which is an important requisite of knowledge collaboration online, credibility concerns in society are also introduced, adding extra work for individuals and communities to evaluate the truthiness of knowledge and capture substantiated knowledge. Similar to the strategies to structure knowledge into an organized framework, prior work has explored various human-centered or algorithm-dominated innovations. Algorithms mainly focus on automatic misinformation detection models utilizing rich information including the content itself, metadata, and external resources; and human-centered solutions design working flows and interfaces, to leverage experts and general crowds to identify and debunk misinformation [7]. More specifically, algorithm-driven automatic correction offers promising results while alleviating cognitive strain on humans. These algorithms typically utilize linguistic, semantic, and image features, along with other metadata, to identify misinformation [78]. Efforts by researchers are focused on enhancing both the data aspect (including human labeling and the integration of contextual features) and the modeling and engineering aspect (encompassing the development of novel transformers and architectures), aiming to boost the accuracy and reliability of these models. Alternatively, to support fact-checking with crowdsourcing, for example, Hoaxy and CrowdTangle created platforms to scale up and speed the crowd efforts and track misinformation on social media [74, 238]. Some other work developed efficient visualizations to highlight experts or crowd-based disputes, and help users evaluate information from different perspectives [29, 137, 209].

Reviewing the related work on knowledge integration and evaluation, it is evident that scholars have made substantial strides in both human-centered and algorithmic approaches. However, while these efforts have yielded fruitful results, the limitations of automatic approaches, particularly concerning their generalizability and accuracy, cannot be overlooked in the practice. Similarly, human-centered solutions face scalability challenges and inherent human limitations, such as bias and fatigue, which hinder their adaptability to the ever-expanding knowledge scope. Consequently, research is increasingly pivoting towards exploring synergies in human-AI collaboration to harness the complementary capabilities of both, thereby forging a more nuanced and robust framework to integrate and structure the knowledge.

2.1.3 Knowledge Sharing and Co-creation

As an integral and prevalent step, knowledge sharing and exchange are crucial in benefiting the wider community and fostering the co-creation of knowledge [26]. Early research in this domain focused on organizational support, emphasizing knowledge transfer within corporate structures [196]. With the rise of knowledge communities and peer production, a surge in studies has examined knowledge sharing in online communities. Sharing useful knowledge can help community members learn from each other's experiences or expertise, and solve problems by reusing previous insights from similar scenarios [25, 248]. Example projects include but are not limited to collaborative projects and platforms such as thread-based discussions, Q&A forums (e.g., Stack Overflow), Wikis of projects (notably Wikipedia), and open-source software development (e.g., GitHub), investigating various dynamics influencing sharing behavior [148, 181, 293]. One line of work concentrates on understanding individual-related factors such as motivations and willingness, contextual elements, and the types of knowledge involved, and design strategies tailored to specific tasks. (Subsequent sections will discuss social-related mechanisms, expanding on the interpersonal aspects of knowledge sharing.)

Departing from traditional economic models centered on rational self-interest individuals [205], the motivations driving knowledge sharing and co-creation that range from obscure one-time cooperation (e.g., fact-check certain news) to sustained, large-scale projects (such as Firefox) present a unique paradigm. From the angle of self-determination theory, intrinsic motivations rest on factors, such as internal curiosity, personal enjoyment. For instance, open-source software and hardware development share similar enjoyment-based motivations, related to joy and fun [109, 248]. More extrinsically, people may be motivated to share pursuing some personal needs, such as learning, building reputation, or filling reciprocity needs [291]. For instance, learning can encourage developers to share their code to not only practice their skills but also invest in future job opportunities as an approach to increase human capital [286]. Besides learning and personal needs, prior studies also point out the power of reciprocity: consumer communities share their experiences and purchase suggestions for subsequent consumers as they receive help from previous members [42]. As the motivations of individuals vary, they also evolve over time and as communities develop, driving them to continue sharing or leave after their needs are addressed.

With the innate human desire to share and exchange knowledge, collaboration and communication technologies emerge as indispensable tools to scaffold users in sharing. These technologies not only offer platforms for the presentation of shared content but also enhance coordination among individuals for collaborative creation. Platforms like GitHub and Wiki design specific collaborative spaces for code development and knowledge documentation, respectively, with integrated version control and management systems to coordinate edits in structure and maintain cohesion [1, 2]. Besides these famous peer production projects, the landscape of knowledge exchange extends to discussion forums and multimedia spaces. To give an example in thread-based discussions, tools like Wikum and Wikum+ are developed to enable user-led curation and organization concurrently with the process of knowledge sharing, thus fostering structured knowledge discussion [264, 293]. Recent AI technologies are further blurring the lines, offering fine-grained levels of summaries of chat threads and real-time discussions, effectively distilling the essence of knowledge exchange while reducing manual efforts [56, 212]. Human-centered collaboration designs also incorporate attention-based designs to prioritize and enhance important information, promote social ties among members, and address other pertinent considerations for video-based knowledge sharing [156, 288]. Moreover, as we discussed, the full spectrum of knowledge also encompasses tacit knowledge, which's is inherently non-verbal and unwritten [276]. In response, researchers are exploring conversational tools that can elicit appropriate levels of detail within artificial sharing environments, bridging the gap for tacit knowledge transfer [132].

2.2 Social and Cognitive Influences in Online Collaborations

For knowledge collaboration among online communities, the dynamic and flexible member participation and interaction patterns primarily impact knowledge co-construction. In this section, we first introduce related cognitive theories that connect with the challenges discussed in the knowledge collaboration at individual level. And we list social interaction types to discern distinct knowledge collaboration models and clarify the interaction engagement in the projects discussed in this dissertation. Then, we delve into the social capital to elucidate how group-level elements influence individual contributions to knowledge sharing and the extent to which these contributions are adopted and used. As this work only focuses on applying these interpersonal factors, this section will not cover the related work that shapes social dynamics.

2.2.1 Intrapersonal and Interpersonal Cognitive Factors

To support knowledge collaboration, few would deny the importance of understanding social interactions, and how individuals process the stimuli or knowledge relevant to understanding other social agents and their interactions (i.e., social cognition). Many of the following chapters start from findings of socio-cognitive processes. In this section, we briefly summarize related work that underscores the intrapersonal and interpersonal factors that connect to knowledge work in our following chapters, and we use decision-making as a common example to simplify the illustration.

To explain many aspects of human cognitive processes, dual process theory, a well-known model popularized by Kahneman, attributes thinking as a result of two different processes: System 1 and System 2 thinking [129]. System 1 thinking refers to the automatic unconscious process by which people employ their heuristics to make intuitive decisions;. In contrast, System 2 thinking is analytical and conscious but slow, which is needed for most of the knowledge work we mentioned in this dissertation. Modern psychology theories believe that they may operate one system or both when people make decisions, and proposed parallel-competitive and default-interventionist two main conflict resolutions to explain the collaboration between two systems [18, 73, 244]. When people apply cognitive heuristics with System 1 thinking, they don't need to invest significant cognitive load, i.e., the amount of mental effort being used in the working memory [256]. Therefore, there are at least three primary heuristics that may introduce bias and challenges for users that need the technical support introduced in Section 2.1: availability heuristic, representativeness heuristic, and anchoring heuristic [269]. For instance, people tend to remember information that they saw recently or frequently online, thereby influencing their decision-making accordingly, as in the famous example about the risk of traveling by different transportation introduced by Tversky and Kahneman. When engaging in System 2 thinking, people experience a considerable increase in cognitive load while thinking rationally. Even though, many of our decisions and thoughts are still influenced by beliefs and biases originating from System 1 [129]. Take overconfidence bias as an example, people can still believe they are less biased than average people in a group after analytical reflection [119]. HCI researchers have spent an increasing amount of attention investigating and mitigating the negative impact of cognitive biases on topics such as human-AI collaboration and information retrieval [16, 273].

In terms of interpersonal aspects, there are different social and situational factors that impact whether people engage actively [80]. Usually, people as social animals tend to get along with others, looking for belonging [106]. Such needs may trigger automatic behaviors to follow others (e.g., reciprocity or mimicry) [75] or conform to social norms they perceive [61, 114], where automaticity is often neccessary to drive people. As an example in Chapter 4, when participants perceive others' reflection process, they still are more likely to think consciously as well, following the social norms. The motive to control may lead to controlled, deliberate decisions for the feeling of control, especially in outcome-dependent scenarios such as collaboration with AI [81]. Social influences are more significant when individuals see the social presence of others or the authority in the group since they value social evaluation or image, no matter it's in person or online. And when people don't identify themselves as being in the group, the influences are relatively weak [30]. Other aspects, such as time, affection, trust, and cognitive load, can also impact the engagement level of System 2 thinking.

2.2.2 Social Interaction and Collaboration Types

Following the outlines in the introduction, members communicate to understand each other, and co-create new knowledge from different channels [173]. Due to flexible collaboration patterns and the large scale of community, a large number of online collaborations aren't limited to direct interactions within a small group of people. Jeong et al. introduce A3C framework to understand individual and collective processes of online collaboration: Attendance, Coordination, Cooperation, and Collaboration [126]. In attendance activities, people "look around" the community knowledge, such as watching videos or reading community-shared knowledge, reacting to others by "likes" or comments. They normally pursue their own goals and do not engage in joint work. Many people fall into attendance on YouTube and social media groups. Chapter 3 investigates how to restructure existing knowledge pieces, both videos and comments, to help people explore diversely and comprehensively. For Coordination, people still work on individual goals but they need to coordinate the processes and the outcome is interdependent. Edited book [1] is one exam-

[1] Edited books are books with chapters written by multiple authors respectively.

ple. In the next stage, cooperation, members share the same goals. They work independently and then combine their individual contributions toward the final artifacts. In the video review task in Chapter 4, participants often chose to cooperate to take notes and share individual thoughts. Collaboration, the full collaborative process, refers to the scenarios where community members pursue a shared goal and work collectively, similar to heavyweight peer production, as we see in open-source software development projects. In Chapter 5, we simulate a collaboration scenario in which participants brainstorm together in collective efforts through real-time video-based communications.

2.2.3 Intergroup Social Capital For Knowledge Sharing and Assimilation

As introduced in Section 2.1.3, there is a range of group-level factors underlying the knowledge-sharing processes. They also impact the adoption and integration of knowledge after they are shared. Among the different facets of social capital [48], we talk about social ties, trust, social norms, and grounding which may be considered in the following chapters.

Based on the time, frequency, and intimacy in social interactions, individuals form certain levels of social ties. Relatively stronger ties can help secure the breadth and quality of knowledge shared, as they are comfortable in the community to share and have some common groundings about the tasks [266]. More sharing activities may also positively associate back to social ties among community members. Meanwhile, online communities also need newcomers and individuals with weak ties to bring new information and diverse perspectives. Consequently, online communities need to find a balance to integrate the shared knowledge and ideas from both established members and newer or less tied individuals, where diverse contributions are valued and effectively assimilated into the collective knowledge base.

Trust in online communities mainly focuses on integrity in individuals in terms of their behaviors and knowledge contribution. Another category of trust, system trust, is also relevant to online communities, as the perceived reliance on the group or even the online collaboration platform.

People often need to build trust during iterative interactions, but may be hindered by limited visual cues and face-to-face contact and potential mask of demographics [228]. Also, due to the fluidity in these communities, the trust development process fluctuates and usually varies depending on the members and tasks [46]. High trust may fuel open communications, build community reputations, and the feeling of community belonging to encourage more contribution. It will also reduce the need for monitoring and content verification. However, communities need to establish efficient and sustainable mechanisms to build and maintain quality interpersonal trust between not only core collaborative contributors but also system trust that will impact attendance people. Without sufficient trust in the community, it may hinder the opportunities for attendance people and general users to value and benefit from the knowledge artifacts.

In the knowledge facet, common ground, the shared knowledge and assumptions about the mutual goals, is paramount for effective communication and collaboration, even all collective actions [55]. For online communities, it's crucial to keep exchanging common ground between individuals ensuring everyone stays on the same page, as they are fluid with changing members and topics. Grounding activities include coordinating the division of tasks, resolving conflicts, and asking and answering questions to clarify ambiguity and confusion. Since most people are geographically separate, the communication channels are reduced compared to face-to-face interactions, which causes more constraints on the grounding process [55]. To enhance communication and grounding in online settings, a wide range of applications have been designed to facilitate richer communication mediation with video and enhanced video tools (e.g., VR and AR).

In practice, these social factors, among others, become a vibrant ecosystem, intricately interwoven and influencing the processes of knowledge sharing and adoption. For instance, strong social ties are related to high trust between individuals, and they tend to share sufficient levels of common ground. For collaboration in online communities, different social interaction and collaboration patterns co-exist and they influence different facets of the social contexts of collaboration. Conversely, collaboration and interpersonal interactions are also molded by these social contexts, as observed in previous work and this dissertation.

2.3 Summary

Overall, there have been adequate research and design systems that contribute to assisting individuals and groups in finishing different phases of knowledge work, and many prior works have acknowledged the benefits of restructuring and prioritizing knowledge for these online intelligent tasks. In parallel, social and information science studies have delved into how people perceive, understand, and interact with others from individual-level and group-level lenses.

This dissertation draws upon these socio-cognitive frameworks, technical innovations and their studies, and state-of-the-art algorithms. Grounded by these solid prior studies, the following chapters of this dissertation unveil my explorations and observations that further empower individuals to address the challenges in knowledge forage, integration, analysis, and co-creation.

Chapter 3

Structure Middle-level Representations across Sources for Understanding

As videos become an increasingly popular medium to understand and learn new knowledge, passive viewing through recommendations risks hindering active learning, calling for a balance between active video exploration and algorithmic recommendations. Existing video navigation support focuses mainly on video content, which underestimates audience comments' informational and social values for active learning. In this chapter, we propose to structure semantic representations of both audience comments and video content to promote active learning and afford manual exploration. We instantiate our hypothesis in a prototype called Kentaurus. It analyzes video transcripts and identifies thematic focus from comments, offering users structured overviews of both content and audience insights to help users start their exploration. In addition, it provides limited video recommendations for users to select videos, customize their own learning paths, and preview the content of individual videos. Through a user study with 60 participants, Kentaurus aided participants in selecting higher-quality videos than a video playlist determined by YouTube's algorithms and crafting personalized learning paths more diversely compared to a content-centric video learning prototype, without imposing additional cognitive demands.

This chapter is expanded from the following short paper: Liao, J., Singh, M., Hung, Y. T., Lin, W. C., & Wang, H. C. (2023, October). ConceptCombo: Assisting Online Video Access with Concept Mapping and Social Commenting Visualizations. In Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing (pp. 362-364). [157]

3.1 Introduction

As the barriers to video creation and sharing of videos are reduced, videos have emerged as a dominant medium for acquiring factual knowledge. The sheer volume of video resources, coupled with the state-of-the-art recommendation algorithms, has enabled users to access the videos based on their search queries on generic video platforms, such as YouTube [128]. However, such passive video watching by simply viewing a list of ranked recommendations, risks stifling users' active cognitive engagement in video exploration and selection, an important aspect of video-based learning processes. Active learning theory suggests that users are more likely to have satisfactory learning experiences and results when they engage *deliberately* in tasks that include identifying appropriate video materials, interconnecting key concepts across resources, and interacting with peers [47, 65]. Present end-to-end personalized relevance-based automatic algorithms provide video recommendations, which may mismatch the pedagogical processes needed. On the other hand, users, particularly those with low technology literacy or domain expertise, still need facilitation to help them navigate a small fraction of quality videos and related resources as their cognitive bandwidth is very limited for video-based resource filtering, especially when they focus mainly on content assimilation by watching videos [141]. Therefore, learning-oriented video exploration support that balances active user exploration and automatic AI support emerges as a promising avenue for amplifying active learning experiences while tempering cognitive demands [112].

Video exploration and navigation designs in the HCI community have utilized semantic representations from the content of videos to facilitate video exploration and selection, especially video sequences, via learning object collection and relevant recommendations. For instance, transcript-based textual cues, video metadata, and visual cues in images (e.g., images and snippets) have been used to find inter-video connections and recommend learning sequences suitable to incremental learning pathways [135, 258, 298]. Previous work also explored the interaction and visualization of these video connections. For example, several approaches advocate for the visualization of video interlinkages through keyword hierarchies or map structures [236, 298]. Others advocate the use of image snippets, allowing users to glean video content at a glance [20]. An example of such innovations is ConceptGuide [258], which generates a concept map-based content summary from video transcripts, offering links between concepts and videos to provide diverse learning pathways and a conceptual overview of video content (see Figure 3.6).

However, when comparing the existing video learning support with diverse active cognitive engagement activities, the social activities related to other audiences (e.g., connect with peer learners or check others' reactions) remain rarely engaged in the video exploration process. Social components, such as time-anchored comments (i.e., Danmaku), and YouTube comments, are mainly used to augment social presence of video-watching processes and highlight informational video snippets, especially when watching single videos [153]. Moving beyond the applications of comments for viewing support, a deeper investigation of the usage of comments reveals a new dimension. Users are observed to seek related knowledge from the comments, and sometimes use them to redirect to other videos [50]. Thus, comments may serve to connect videos in terms of video content. They also reflect the learning states of the learners, which could be valuable logs for instructors and other learners to gain learning feedback from existing audiences [253]. An examination of YouTube comments on scientific videos also revealed that comments and discussions on generic video platforms have evolved from mere knowledge sharing to the encouragement of argumentative thinking, which led to a higher level of collaborative understanding [68]. To balance the benefits of comments and their overwhelming scale, the same dilemma in video content, there's latent potential to integrate social comments into the process of video exploration and selection phases, which may both help users utilize the audience comments for video exploration and enhance their active learning experiences by increasing access to social reactions from different phases of learning processes.

In our work, we explore blending semantic representations of videos and comment resources to enrich video exploration experiences for video-based active learning. To scaffold users among dozens of videos and thousands of comments, we propose to preprocess the videos and audience comments with automatic computational pipelines to offer structural overviews of video content and audience comments with middle-level representations (e.g., keywords and learning objects), along with their connections with associated videos. And users can explore videos based on structured previews from different information channels, followed by fine-grained video recommendations for each representation, a glance of content, and comment highlights of videos. We instantiate the idea in our video exploration prototype on YouTube videos, *Kentaurus* (**K**nowledge

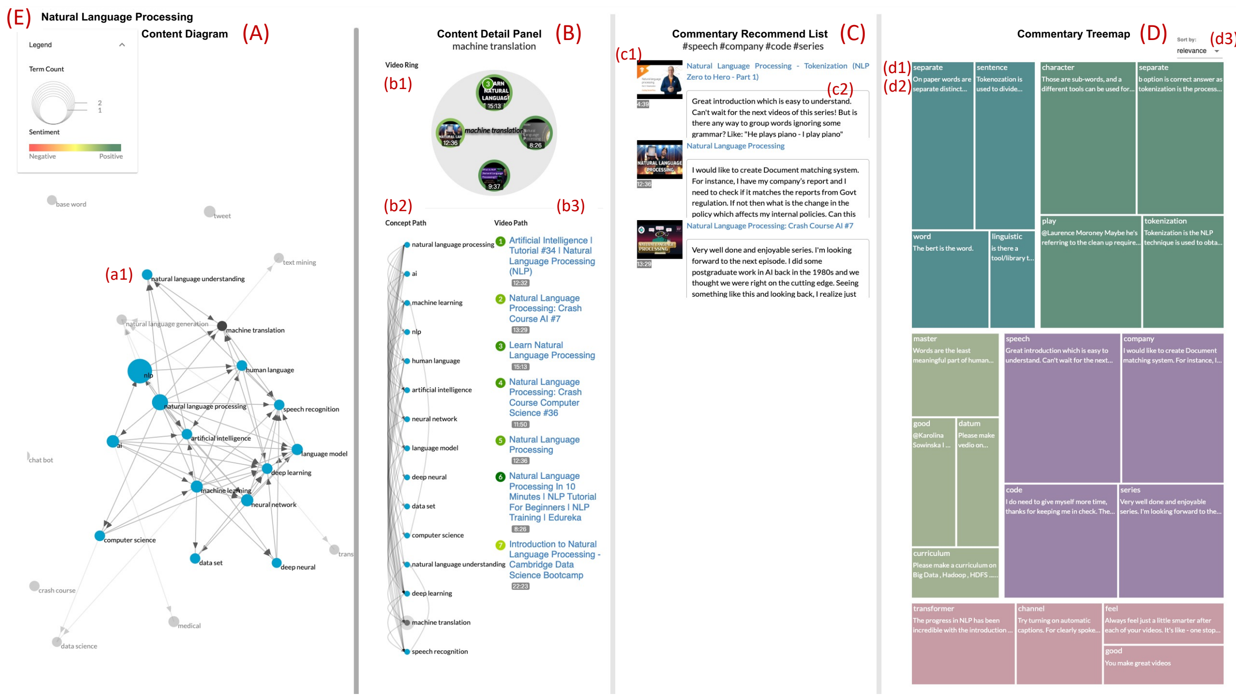


Figure 3.1: Screenshot of Kentaurus Overview. As a user chooses a video topic ("Natural Language Processing" (E)), Kentaurus first presents the structure overview of video content – Content Diagram (A) and comments – Commentary Treemap (D), on the two sides. Content Diagram presents top concepts (a1) and their directional relationship identified from video content. When a user interacts with the Concept Map by clicking on a particular concept node (e.g., "machine translation"), a Content Detail Panel (B) emerges. This panel offers multiple video recommendations associated with the chosen concept. Specifically, it displays a circular arrangement of recommended videos for the chosen concept (b1), a sequential pathway of related concepts and associated video recommendations (b2 and b3), enabling users to engage in a structured learning journey from foundational concepts to their concept of interest. Simultaneously, the Commentary Treemap visualizes the social focus from the comments around videos with multiple keywords for each cluster (d) and a snippet (d2) of a representative comment. By choosing a theme node (e.g., "speech"), users can activate the Commentary Recommend List (C). This panel surfaces a short curated set of videos (c1) accompanied by comments (c2) that align with the chosen thematic focus, offering social video recommendations. Users can also adjust the ranking criteria for keywords in each theme (d3).

Exploration and **N**avigation for **T**opics from video content and **A**udience insights with **U**sable middle representations and **R**ecommendations for general **US**ers) [1]More specifically, as users explore a topic, Kentaurus (1) analyzes video transcripts to identify the important learning objects as concepts and the prerequisite relationship between concepts as a map-based overview, and (2) identifies topic-relevant comments and categorizes the comments into overarching themes that indicate audience focus and connect with other videos. With these data, users can first access the conceptual knowledge of concepts and prerequisite relationships via a directed graph (Content Diagram) and a hierarchical treemap that visualizes thematic summary of keyword breakdowns from comments (Commentary Treemap). They can explore the video landscape through the representations of these overviews which are associated with videos, and also triage the content and authors of resources actively and flexibly. Figure 3.1 presents a screenshot of Kentaurus, showcasing the topic "Natural Language Processing," collected in 2022 [125].

We investigate how the mixed affordance of structural overviews of video content and audience comments influence learning processes and video selection through a user study with 66 participants. As such video exploration design works to find a balance between automatic support and active exploration, a crucial element to understand is users' acceptability given diverse sources and synthesis levels of information for active learning, as well as their usage strategies. We compared this idea with two prototypes with different video information channel overviews: Content-centric system, ConceptGuide [258], which has a structured overview from video content alone, and Default List, a ranked video playlist generated by the deterministic algorithm from YouTube with the same video set which only has individual video metadata available. Through the analyses of quantitative surveys and interview results, we found that blended representations of both video content and comment channels can help users find videos of better quality and actively select their own watching paths compared to Default List, when Content-centric System didn't. Learners using Kentaurus did not find it significantly more cognitively demanding than the other two conditions. Moreover, learners' attitude toward comments in general and from the visualization of Kentaurus largely impacted their usage and usage timing of comment-generated learning support. These results provide empirical evidence on learning experiences with learning support affordance from different middle-level representations of resources, and the design space

[1]Kentaurus is named after centaurus, the mythical creature that combines two creatures, human and horse.

for active learning facilitation on generic video platforms.

3.2 Background and Related Work

This work builds upon the insights from the video-based learning literature, especially about social interactions, and designs of video navigation interfaces.

3.2.1 Active Learning, Social Engagement, and Comments

Videos have rapidly grown as a primary source for people to acquire knowledge in both the personal and professional sectors [257]. Within the recent HCI research on learning, two kinds of video topics are mainly studied and supported: procedure videos and factual knowledge videos [135, 136, 166]. Procedure videos cover topics related to specific skills or procedures through demonstration or guided practice, such as recipes, lab experiments, how-to, foreign languages, and surgery [135, 136]. In contrast, factual knowledge videos primarily present information, facts, and elucidations on specific subjects, including academic presentations and documentaries. For factual knowledge videos, they present information mainly through speaker narration when visuals and graphs work as aids [35]. The presentations will primarily follow the logic order of knowledge, which is not sequential as in procedure videos. Our work primarily focuses on topics in these domains of factual knowledge. This focus drives our video and comment summary techniques to rely on textual video transcripts and comments to identify central learning objects on given subjects.

The literature on active learning emphasizes that learners can engage cognitively and meaningfully in learning activities to improve their learning outcomes [86]. In terms of overt behaviors that learners can engage in active learning, the Interactive, Constructive, Active, and Passive (ICAP) framework suggests actions such as browsing and selecting videos (active mode), comparing concepts with other materials (constructive mode), and discussing with peer learners (interactive mode), for video viewing scenarios [47]. Dodson et al. also introduced a detailed framework for active video viewing that treats videos as a non-linear learning medium, including video search, triaging with textbooks and other resources, and tagging video content [65]. As this work extends the video-based active learning process within video exploration, we aim to provide more opportunities for users to participate actively from the aspects of video selection, constructing conceptual and structured understanding, and increasing social connections, to improve their active learning experiences and video selection results.

Inspired by the interactive mode designs for learning, one avenue to facilitate active participation is to create a collaborative and social learning environment [121]. These interactions could occur in asynchronous and synchronous media, which leads to different positive influences on users [120]. For live streaming online lectures, real-time discussions between learners or with instructors, such as questions and shared notes, improve social presence for both roles and encourage more knowledge sharing and feedback in a timely manner [76, 172]. More often, learners access recorded videos on generic video platforms where people leave comments under the videos or time-anchored comments that display as the video plays [120, 121]. When users interact through these discussion channels, researchers found that the discussions have produced high levels of collaborative understanding and new knowledge, akin to more formal learning settings [68]. Even when users don't actively participate in the interactions, Tanprasert et al. presented that witnessing vicarious dialogues between peer learners and instructors alone can augment learning concentration and overall satisfaction [259].

Besides comments, social annotations and other interactions also provide diverse supplementary learning materials, social connections, and guidance for users about their online learning [76, 142]. For instance, social annotation of video transcripts may reduce learning distraction with collaboration annotation designs [76]. Also, reactions to personal comments with tags that reveal detailed opinions for comments could improve social feedback and support constructive learning processes [142]. Based on these findings, our research explores the potential to increase the utility of comments in video exploration by integrating comment summaries in the content overview and in the video selection phases. We also aim to understand how such an overview of social comments as additional video metadata may impact users' video selection and navigation.

3.2.2 Enhanced Video Learning Interfaces

To facilitate video navigation and viewing on genetic platforms for constructive and active learning-related actions, existing HCI research has pursued two primary avenues. The first focuses on aiding navigation within one single video with different cues, employing techniques ranging from automatic algorithms to crowdsourcing [166, 185]. For example, user-generated comments and upvotes are used to build the summary of live-streaming and pre-recorded videos to help learners grasp the highlight of video content [153, 172]. And some studies focus on time anchoring between comments and videos, and have designed the visualization of time-anchored comments embedded with video timestamps, in order to help both learners and instructors understand learners' learning status with content and sentiment analysis and timing of important video clips [153, 253]. For long videos, researchers notice the difficulty of navigating around different sections of the video. Multiple systems have used machine learning algorithms to extract chapters using narration, video snapshots, and other sources, enabling users to preview and navigate through the content more effectively, which's now available by multiple commercial video platforms [185]. Other work also designs crowdsourcing approaches and manual labeling tools to help decompose the long video content into smaller sections [166, 230]. As such research revealed the benefits of audience and user comments on content navigation, many studies collected comments from their participants exclusively for their studies, which has different quality concerns than natural comments left on YouTube. This work explores to utilize the existing comments from the real world, where users have pre-existing perceptions about their quality and trust tendency, to understand the potential and challenges users may face in video navigation.

The second trajectory delves into navigation support across multiple videos, which is the focus of our work. They mainly propose to find the connections between videos into different structures and provide interactive visualizations to help users navigation in the video pool. For instance, MOOCex recommends videos via sequential relationships between videos and video topics, and further visualizes the semantic relationship of related videos on a 2D space with related videos in Voronoi-alike diagram [298]. Surch [135] displays a semantic graph of operation procedures and enables step-level video comparison. Content Flow Bar provides "snippets" of videos in one display with menus of keywords and definitions [214]. As most of these systems leverage content and metadata of videos, few studies have explored integrating social comments, especially semantic comment content, across different videos for video navigation on generic video platforms. The lack of such opportunities can limit the opportunity to interact with peer learners in the video selection process. Our work synthesizes ideas from these systems to provide structured overviews of video resources, and is situated with the process of video exploration to enable exploration with previews from different connected information channels.

3.2.3 Learning on Generic Video Platforms

Video-based online learning does not always lead to satisfying and efficient learning experiences, especially on generic video platforms [242]. It depends on factors such as learning resource quality and user trust and motivation. From the perspective of resource diversity, the unrestricted nature of these platforms allows both experts and general users to disseminate educational content across a broad spectrum of topics, employing different presentation styles, and tailoring content for different learner groups [242]. Potential access across the world also offers opportunities for all users to share their knowledge, experiences, and opinions around the videos and to access others' thoughts. It not only enriches the information sphere of topics, but also creates a community for collaborative and self-directed learning of the booming video collections [68, 151, 176]. Recognizing its benefits, challenges about video navigation, credibility, and quality have also gained significant attention from general users, practitioners, and researchers [41, 123, 191]. Videos uploaded by different authors do not consider consistency and overlap largely with each other, which is not optimal for learning that requires watching multiple videos [41]. Videos with high views and significant comments do not guarantee quality and credibility [151]. And the work auditing misinformation on YouTube recommendation algorithms found that watching the history of misinformative videos leads to more misleading videos on search results, up-next, and top recommendations [123]. For comments, there are noises around informative comments, including spam, memes, self-promotion, and even toxic comments [179]. A portion of comments also includes misleading content, such as refusing scientific health suggestions based on personal experiences and advertising conspiracy theories [8]. Moreover, as single videos may receive hundreds of comments, it's also difficult for users to refer to all the comments and look for informative information, which may not rank consistently with liking scores or default display order [50].

With these, users now face extra activities that turn into learning barriers rather than active learning activities suggested by theories that originate in formal learning settings. The efforts required to retrieve relevant and accurate videos and comments may put extra cognitive burdens on users, especially novice learners who don't have sufficient background knowledge [22]. Users must rely primarily on sources, presentation quality, language styles, and social popularity. Redundancy from materials can waste users' time on duplicated content and limit access to various subtopics, which can cause misconception for users about the topic [9]. Inconsistency on terminologies may also confuse some users when comparing concepts across videos. Regarding comments, prior work about Facebook comments reveals that conflicting comments in economic and political news (hard topics) could reduce their attention to the comment field which may reduce their engagement with other audiences, similar as their impact on their corresponding news articles [69]. The presentation of comments is substandard for interactions, such as most comments that are pushed down into the scroll area within a few minutes of their publication may be barely noticed by users [235].

In our work, we propose to address these challenges with automatic algorithms by creating middle-level structured learning representations that aggregate individual information. Our approach aggregates individual videos and comments and extracts learning concepts or keywords, to fold a portion of irrelevant information and use knowledge to select within videos. It allows users to practice active learning behaviors around video selection and knowledge construction.

3.3 Methodology

We instantiate our idea in the design of Kentaurus, which allows users to actively explore videos by providing middle-level structured representations from video and comment channels, along with video recommendations and interaction designs.

3.3.1 Design Goals of Kentaurus

Following the ICAP framework, which categorizes user activities that link to advanced learning engagement modes (i.e., active manipulating, constructive generating, and interactive dialoguing) [47], we break down the overall video exploration and learning processes into three pairs of user actions and automatic support designs.

R1 *Embed comment summary in the overview:* To enable users to access social comments during the video exploration processes, the system should present a commentary summary, which will facilitate users to enjoy the benefits of social commenting and leverage them for video selection.

R2 *Integrate content into structured overviews:* Videos and their accompanying comments in generic video platforms tend to be unorganized for the overall topic, as they are generated by different creators [47]. The system should provide direct structures of content or guide users in constructing their own inference of the topic and the framework of knowledge.

R3 *Provide connections between comment and video content channels:* To support users in finding a video and the next videos to watch, the system should allow quick previews from both video and comment content channels as users explore actively. It may design visualization connections to help users curate their own video-watching list based on their interests and learning focus on the spot.

3.3.2 Interface and Interaction Design

Following Shneiderman's mantra "overview first, zoom and filter, then details on demand" [240], the interface of Kentaurus is designed in two layers – overviews and details with four main panels, and provides pop-ups for details. Users start the experience by searching for a topic in the search bar or picking a learning topic from the navigation menu, and Kentaurus will return the main overview page, as shown in Figure 3.1. Our system initially visualizes the structured summary from video creator's transcripts, *Content Diagram (A)*, and the semantic themes of learners' comments in *Commentary Treemap (D)*, respectively. *Content Detail Panel (B)* introduces recommendations for concept-specific videos and sequences upon selecting a concept in the Content Diagram (A); similarly, *Commentary Recommend List (C)* presents a curated list of videos whose comments are relevant to the comment theme chosen. In this section, we describe the details and interactions of the Kentaurus interface, as well as the underlying rationale.

Content Diagram (A) visualizes the concept-based outlines of video content and associated video recommendations. We adopted the visualization and interaction paradigms initially outlined in a prior study, ConceptGuide [258]. In the Content Diagram, we present a network of salient knowledge concepts extracted from video transcripts. Basic and core concepts are positioned at the center of the canvas, encircled by more advanced or less relevant ones. (For instance, "deep learning" is close to "language model" in the center where "data science" is in the bottom corner, in Figure 3.1.) The size of nodes represents their relative frequency in the video corpus.

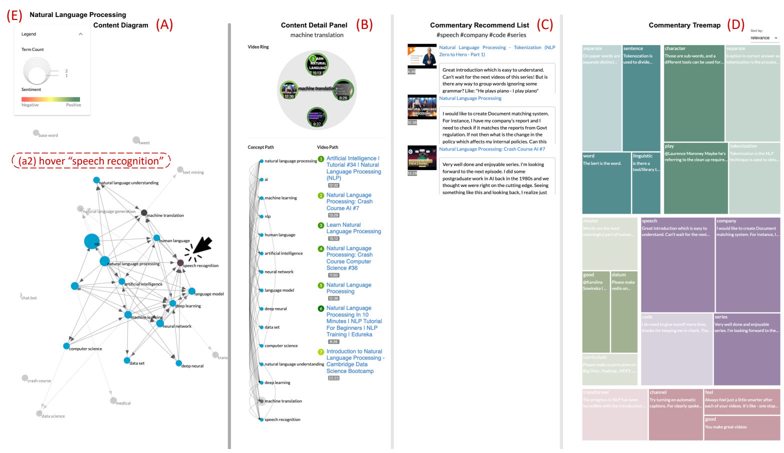


Figure 3.2: Highlights when hovering over a node in Content Diagram (similar interaction when hovering in Commentary Treemap)

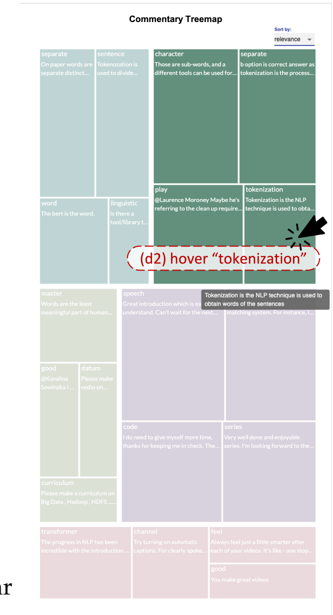


Figure 3.3: Hover a comment and its comment theme on Commentary Treemap (D)

To better convey the structure of concepts across videos (**R2**), we incorporate the directed links to indicate the direction from early to advanced knowledge and visualize the network in a force-directed layout. The size of each concept is proportional to its frequency across videos. When users hover over a concept node, the links of interconnected concept nodes in the canvas are highlighted in black and nodes in blue (as in (A) in Figure 3.2). Meanwhile, the system affords interactive connections between Content Diagram and Commentary Treemap, to support users in relating the summary of comment channels with videos and activate users to read comment-based summary (**R1**). More specifically, whenever users select or hover over a concept, the interface will identify the first video in the video ring (i.e., the video with positive comments and the highest frequency of the concept), and highlight the keywords available in its comments in the Commentary Treemap dynamically by fading others.

Content Detail Panel (B) displays three different details to visualize the conceptual contexts and video recommendations as users select a concept node: video ring (b1), concept path (b2), and video path (b3) in Figure 3.1, when any concept is selected (a1). In the video ring, a small group of videos are recommended for the selected concept with video thumbnails, ordered clockwise, available for users to pick videos. The color of its ring indicates the sentiment of its comments:

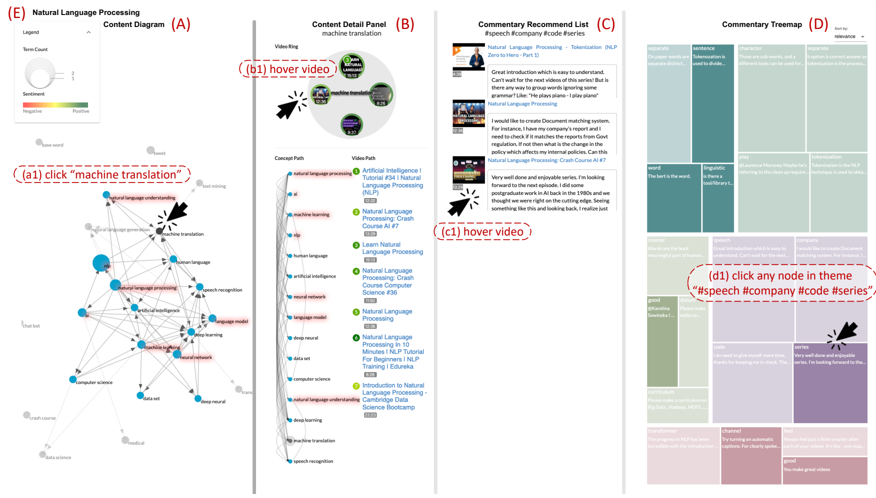


Figure 3.4: Highlights when hovering over a video on Content Detail Panel or Commentary Recommend List

the darker the green, the more positive their comments. The concept path shows a curated path of the selected concept and its prerequisite from the basic concepts. As intricate video recommendations, video path specifies a sequence of videos with titles from the concept path that users may follow from the most basic concept to the concept selected. Content Detail Panel also provides interactive previews of video transcripts and comment summaries to help users decide whether to watch them (**R3**). Specifically, when they hover over any video title or thumbnail that represents a video, Kentaurus will check its transcript and comments, and highlight the ones that appeared on the corresponding Content Diagram using red background and Commentary Treemap by fading other keywords. The visualization is shown Figure 3.4, sharing the same effect when a video is hovered in Commentary Recommend List (c1). In this way, we created a content-oriented video meta-data panel that augments users to pre-select videos with previews of connected content concepts and comment highlights.

Commentary Treemap (D) To facilitate users attain a preliminary insight into the comments that illustrate social reactions as an approach to interact with peer learners (**R1**), we designed the Commentary Treemap, a topical summary. Users may quickly capture the themes that synthesize comments based on what users keep mentioning in groups (**R2**), such as key video content, Q&A (for example, "crash course" as video type and "transformer" as key learning objects). Overall, the treemap will display multiple comment clusters along with their representative keywords in squarified layout as in Figure 3.1. The number of keywords per theme and themes was practically calibrated for the treemap visualization, ensuring an optimal balance between representational clarity and information density. The details of layered treemap generation are in the section 3.3.3. In the treemap, each rectangle corresponds to one keyword, attaching a snippet of a comment that best exemplifies that keyword. To cater to personal user interests in comment topics and afford personal selection (**R3**), we allow users to choose the topic terms ranking criteria by which keywords are more ”representative”: frequency or semantic closeness to the topic (Figure 3.1 (d3)). Users can switch the thematic summary to either reflect the prominence of terms or the essence of the theme through semantically resonant terms. The colors in the treemap serve to differentiate between divergent themes and the size and positioning of these rectangles convey the importance and ranking of the respective keywords. When frequency criterion is selected, the themes and keywords are ranked based on the occurrence in the comment corpus, and arranged with the largest cluster at the top left and the smallest one at the bottom right. The size of rectangle areas also indicates their relative frequency. In comparison, when users choose to visualize keywords that are semantically close to the topic, the themes and keywords per theme are ordered according to their cosine similarity to the topic and videos from text embeddings.

To better convey the connection between the comment keywords and other content, users may hover over these rectangles to reveal the complete text of the associated comment snippet, as shown in Figure 3.3. The remainder of the treemap also triggers a fading effect to emphasize the chosen theme, to help users focus on the keywords in the theme and the corresponding comments to help understanding. Simultaneously, following the same interaction designs across Content Diagram and Commentary Treemap, when one keyword rectangle is hovered in the treemap, Content Diagram responds and highlights knowledge concepts mentioned in the video source of the comment, offering users a seamless awareness of interconnections between two content channels. This Commentary Treemap provides an intuitive structured overview of comments over the videos, and detail-on-demand connections with video content.

Commentary Recommend List (C) As prior work suggests the connection of videos through comments, we provide comment-oriented video recommendations to provide video recommendations and previews. It populates with a small curated list of videos linked to the comment theme chosen in the Commentary treemap, as social recommendation shortcuts for video selection. These videos are mainly corresponding videos of the comments in the theme or contain semantically similar keywords to the keywords selected to represent the theme. To help users grasp the content in these videos, this panel presents multiple vital video metadata, incorporating the title, thumbnail, and a few comments resonant with the selected theme. Similar to Content Detail Panel, with the purpose of offering a more enriched contextual preview for video selection (**R3**), hovering over any video in the list accentuates the relevant concepts from the video content on the Content Diagram and its comment on Commentary Treemap (Figure 3.4 (c1)). With this, users may explore social video recommendations driven by comment topics (**R1**), and examine the video content through related video content concepts on the left panels.

Once users pick any videos to watch by clicking the middle panels, Kentaurus will redirect users to a video view, which includes the video player, comment panel, and Content Diagram. The video player supports standard video controls, such as speeding, turning captions on/off, jumping between chapters, etc., to fulfill personalized video watching habits. Comment List next to the player displays content-relevant comments selected by algorithms introduced in section 3.3.3. Content Diagram are still available to further guide users on the structure of concepts to help construct the knowledge framework and guide the learning process (**R2**).

3.3.3 Computational Framework

Kentaurus relies on NLP, recommendation algorithms, and external knowledge resources to create middle-level learning representations and video recommendations from video resources. Shown in Figure 3.5, as a search query is requested, we collect the top 50-100 non-short videos searched from YouTube Data API [261], and get their captions and user comments if available, and other video metadata such as titles, thumbnails, like counts, duration, video descriptions, channel names, channel descriptions, etc. After that, we select video candidates based on two practical rules: 1) videos that have English captions, automatically created or manually uploaded. Our system works only for English for now. 2) Videos whose duration is between four and thirty minutes. The duration threshold is chosen from our empirical observations about videos. Although short videos are hugely popular among young generations, they tend to be small nano-content and edutainment [134], so we decided not to include them. Usually, these videos account for at least half of the original video retrieved. Meanwhile, we filtered out comments that are very short or irrelevant content, which we will introduce in more detail in Section 3.3.3. With the selected videos

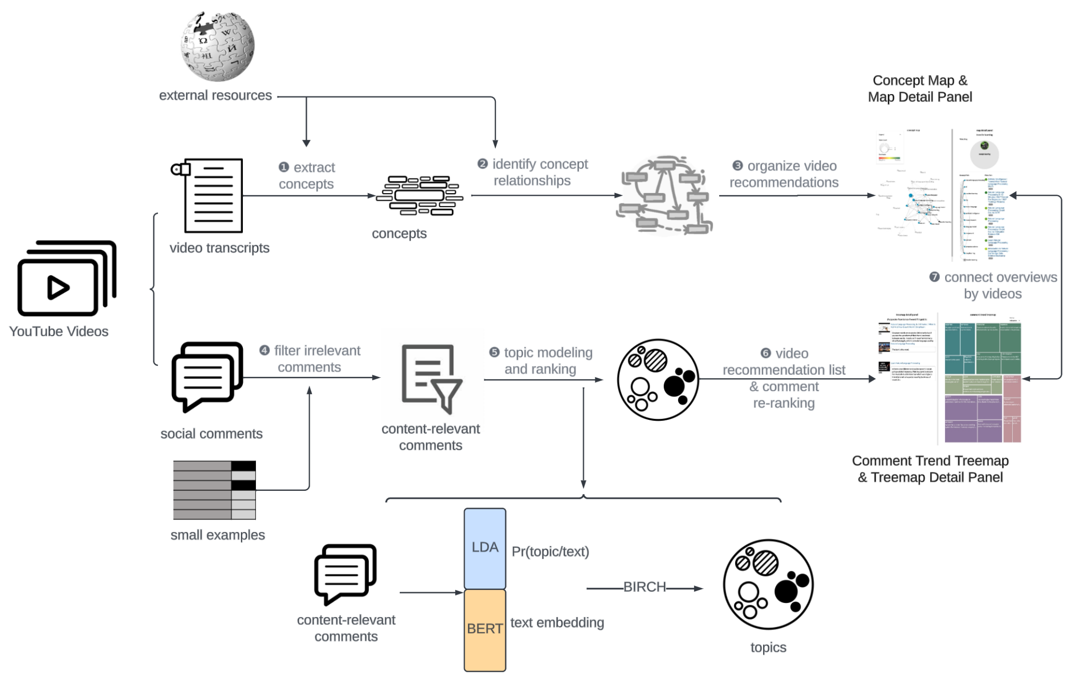


Figure 3.5: Kentaurus Computational Pipeline

and corresponding comments, we generate conceptual map and topical treemap of semantic content from videos and comments, in Content Diagram and Commentary Treemap. Furthermore, we present relevant videos according to the semantic similarities that support dedicated refinement for video selection in Content Detail Panel and Commentary Recommend List. In this way, we offer middle-level overviews and guidance of video resources available to users, in order to activate their active learning activities, as well as investigate how users may interact with these representations to navigate the video learning processes.

Generating Content Diagram from Videos

To create a directed map that captures the key concepts from the collections of video transcripts and their association, we extracted concepts first and identifying the relation and prerequisite of concepts. We adjusted the pipeline used in Tang et al.'s ConceptGuide [258]. With the rapid innovations in Large Language Models (LLMs), it now turns relatively easier to generate the map directly via transformer-based large language models with prompt engineering. In this work, we opt for NLP solutions that further fine-tune or start with pre-trained models, such as BERT or ERNIE, and stack with statistical methods and rule-based approaches. It helps to balance satisfactory performance with computational efficiency and further video recommendations [60, 252].

Concept Extraction: We stacked several keyword extraction techniques to identify the key learning objects from videos of different video producers, including statistical methods, transformer-based model, and external knowledge resources. We used KeyBERT, a self-supervised BERT-embedding keyword extraction algorithm, to extract keywords in two rounds: the first round from individual transcripts separately and the second round from the pool of all transcripts together to identify potential keywords across videos [99]. Due to local self-attention and memory limitations, we decided to split them into multiple groups practically and merge the transcripts into one full document based on the size of the transcripts. In addition to the BERT-based method, we incorporated manual help from video uploaders and external knowledge resources. We also collected the tags of videos as concept candidates, as these tags frequently encapsulate the core content of the videos from the producers' angle. Furthermore, to effectively extract keywords that are domain-important, external knowledge graphs and dictionaries, such as DBpedia [14] and Wikipedia [281], may be used to look up domain-specific words. We used Google's NLP API [124] to ascertain the domain of the video topic query and then match the terms in this domain glossary in the transcripts as concept candidates as well. Finally, we ranked the importance of these concept candidates based on a ring of weights, such as TF-IDF scores, lexical length, frequency across multiple videos, and presence in video titles, similar to in [258]. In the end, we picked the top 30 concepts as final concept list. The number 30 is determined through iterative experimentation focused on balancing the visualization and information density.

Relation Mapping: To create the structures of concepts into a map, we established the connections and pre-requisites based on their semantic similarity and their prerequisite anchored both in the videos and external references (e.g., Wikipedia). We measured four aspects to discern prerequisites between pairs of concepts (a, b) using rule-based approach. It's built upon methodologies introduced by Pan et al. [208] and modified in [258]. The summary of these measures is listed in Table 3.1.

- *Semantic similarity* This measure quantifies the semantic closeness of concepts by evaluating three different dimensions: cosine similarity based on BERT-embeddings, co-occurrence frequency in videos, and references in Wikipedia. Utilizing the BERT embeddings generated by KeyBERT, we first compute cosine similarity as a foundational semantic metric. Additionally, we analyze the co-occurrence ratio of concepts a and b across the video corpus as local closeness. And if concept b is referenced in concept a's Wikipedia entry, a Wikipedia-based similarity score is assigned, which represents the global connections from external resources. The composite Semantic Similarity score is the mean of these individual metrics.

- *Video reference prerequisite and Wikipedia reference prerequisite* Both measures assess the prerequisite relationships between concepts a and b through their relative positions in videos and Wikipedia, respectively. The main idea is that foundational concepts are expected to be highlighted in introductory or early videos, while more advanced concepts are typically introduced later. The same logic applies to their appearance on Wikipedia pages. If either measure yields a non-zero value, a prerequisite score is inferred between a and b. Detailed calculations can be found in Appendix A.1.

- *Complexity level distance (Cld)* This metric captures another facet of prerequisite relationships by examining the frequency and temporal distribution of concepts across videos. This metric comes from the idea that basic concepts are more frequently addressed and last for an extended duration over time. Formally, $Cld(a, b) = avc(a) - avc(b)$, where $avc(a)$ represents the proportion of videos in V that refer the concept a.

We synthesize these measures into a comprehensive prerequisite score $P(a, b)$ through weighted summation, post-normalization into the [0,1] range. A positive value signifies that a is a prerequisite for b, and vice versa. When $P(a, b)$ exceeds a predetermined threshold, it will be visually represented as a directed link in the concept map.

Metrics	**Detail**
Semantic similarity (Ss)	Mean of cosine similarity (BERT) of a and b, co-occurrence in videos, and Wikipedia pages.
Video reference prerequisite (Vrp)	Relative positions of a and b in videos. See Appendix A.1 for calculations.
Wikipedia reference prerequisite (Wrp)	Relative positions of a and b in Wikipedia. See Appendix A.1 for calculations.
Complexity level distance (Cld)	$avc(a) - avc(b)$, where $avc(a)$ is the proportion of videos referencing a.

Table 3.1: Measures for the prerequisite relation between concepts a and b

Video Recommendation To allow users to quickly navigate the videos based on concepts from the Content Diagram, Kentaurus recommends a small ring of video recommendations for any selected concept in the concept ring in Content Detail Panel and corresponding video sequence recommendation from the basic concept to the selected concept, visualized in video path. To achieve this capability of concept ring, we employ a multi-criteria approach that recommends the videos relevant to the concept by occurrences and is rated positively from comments. Videos are ranked by the frequency of concepts mentioned in the transcripts. We also visualize the sentiment of their comments using the color of the outer ring of video thumbnails. We use TextBlob to measure the sentiment of each comment, and assign the sentiment for each video by averaging the sentiment scores of all comments [171]. To generate the video sequence recommendations, Kentaurus initially uses topological sorting to identify the shortest path from the basic concept to the selected concept (i.e., concept path), and then selects one representative video for each concept. Subsequently, an optimal viewing sequence is determined based on the frequency at which each video appears in the concept path.

Creating commentary treemap from comments

As shown in Figure 3.5, Kentaurus aggregates highlights from comments across videos and presents social-oriented video recommendations. Due to the quality issue of YouTube comments, we started by curating comments that are relevant to the video content or initial query. Then, to summarize the comments by topics, we built a topic modeling algorithm by hybrid BERT and LDA embedding and clustered the relevant comments into topical groups. For each group, we present top common or semantically relevant keywords to help users review audience reactions and thoughts for active learning.

Relevant Comment Curation: As we discussed above, comments under videos could be valuable for active learning [68], but they may also include non-relevant messages, such as spam, general greetings, or memes [179]. Thus, for preliminary analysis, we fine-tuned a pre-trained knowledge-enhanced language model, ERNIE, to classify relevant and irrelevant comments. Similar as BERT, ERNIE is a large knowledge-enhanced transformer-based language model, capable of natural language understanding tasks with exemplary fine-tuning and zero-shot experiment performances [252]. For the text classification, we froze all the layers of the ENRIE model (i.e., keep all the model parameters and layers) and attached a linear layer to the output vector of the [CLS] token of ENRIE, serving as our comment classifier. Two coders were invited to manually collect and code a small collection of YouTube comments with labels, such as “greetings”, “questions”, and “opinions”, as in [179]. Then we merged the original labels into binary labels – relevant and irrelevant, considering their video topics. For instance, ”greetings” and ”spam” are merged as ”irrelevant”, and ”summary” and ”opinion” are merged as ”relevant”. The comment classifier was then fine-tuned with this labeled YouTube comment dataset.

In refining the topic-specific comments for our following analysis, we initially narrowed down our focus to a maximum of 100 or 40% most-liked comments per video, forming a raw comprehensive comment pool. This was employed to guarantee not only relevance to the video content but also a uniform representation across the videos, balancing the comment spectrum from highly engaged videos to those with lesser audience interaction. As our prototype currently supports English only, we also filtered out non-English comments. Furthermore, we excluded comments comprising fewer than four tokens, as their tend to lack substantial content-rich keywords and often embody brief expressions or spontaneous reactions. Subsequently, the dataset underwent further filtering using the fine-tuned classifier to sift out comments deemed irrelevant to their topic. This meticulous approach resulted in a distilled subset of comments, characterized by their elevated quality and topic pertinence, thereby providing users with a focused and insightful comment list for users to process.

Comment Topic Modeling and Representation: Kentaurus provides treemap-based visualization of commentary topics to scaffold users exploring the comments systematically with low cognitive efforts. To achieve this, we customized a clustering model that integrates results from Latent Dirichlet Allocation (LDA) and embedding of sciBERT [24, 32] as mix embedding. A large portal of topic modeling algorithms, such as LDA, aims to cluster documents and long texts that self-contain sufficient information for modeling. However, video comments are largely contextual to their videos and have a character limit on YouTube. Word co-occurrence based methods like LDA may miss the topics when texts are incoherent in terms of word choice or lack clear context explanations. To address this issue, we developed the clustering model for comment clustering shown in Figure 3.5, inspired by [251]. The model begins with the extraction of latent topics from the comments using LDA as well, and outputs a topic distribution representing the comment’s relevance to each topic. The model then leverages sciBERT for contextual sentence embedding. SciBERT is a BERT-based language model trained on full-text scientific publications on multiple domains from Semantic Scholar [24]. SciBERT performed as SOTA for tasks such as text classirfication and Named Entity Recognition (NER), which may help better renders the embedding for scientific corpus as in our video topics. Following this, the model concatenates the topic probabilities from LDA with weights and the sentence embeddings from sciBERT, as the joint features with both contextual and topical information. This concatenated data is then fed into a BIRCH clustering method [294], to learn an efficient latent representation of the comments and cluster them into several groups. We pre-set the group size as 4 to 6, and the model will select the topical group size based on the coherent score. With this model, Kentaurus is able to group the relevant comments into multiple topical themes.

To represent these comment themes with a small set of keywords, Kentaurus visualized two sets of keywords to support users in capturing social focus from occurrence and semantic relevance perspectives. For the common keyword set, we identified a core set of four to five keywords that recurred across various themes, creating a frequency-oriented keyword ensemble, utilizing Term Frequency-Inverse Document Frequency (TF-IDF). In contrast, the second set of keywords per theme was curated with an emphasis on thematic congruence with the video content. Starting with the top pproximately 20 TF-IDF scored candidates, we refined this pool by assessing the sentence-level semantic textual similarity of these terms with their corresponding video transcripts using cosine similarity in sentenceTransformers [225], and retaining four to five keywords with a high degree of semantic similarity to visualize. The size of themes and keywords is pragmatically determined to ensure effective visual representation for the treemap layout. To help users understand the connection of keywords to the topic, we also presented an example comment when users hover a keyword as in section 3.3. For example, for the keyword ”renewable energy,” mentioned in Energy Crisis, the representative comment chosen says ”The infrastructure to produce and run green renewables is also powered by conventional resources like coal.” To achieve this, we also measured the semantic similarity between the keyword and the comments that include the keyword. Then the comment with the highest semantic similarity is marked as the ”representative comment” and is prominently displayed below the keyword within the treemap visualization. With these explanations, the two lists of keyword representations and comments provide users with an intuitive and immediate understanding of major themes in the comments.

Video recommendation and Comment Re-ranking: Similar to video recommendations for

Content Diagram, the interface presents a fine-grained video recommendation list for each comment theme on Commentary Recommend List. For each comment theme, the system measured the semantic matching between video transcripts and keywords in the keyword set, using sentence-level semantic similarity with sciBERT embeddings. We then curated the top relevant videos based on the average similarity to all the keywords as final recommendation list.

Furthermore, so as to facilitate users reading relevant and diverse comments when they watch the videos, we followed a comment re-ranking mechanism to re-order content-relevant comments selected from previous stages. Given the absence of a ground truth for optimizing the ranking such as using normalized discounted cumulative gain as loss function, we adopt a pragmatic approach. Given the semantic similarity with video transcript and like count of comments, we weighted the rankings of two orders into a combined score and visualized these comments according to the weighted score.

3.4 User Study

We designed a user study to investigate the potential of dual-channel middle-level overviews of video resources for video exploration, and understand how users customize their learning processes given different scope and granularity of video overview and metadata, when learning videos actively online. We conducted a lab experiment with three different video learning prototypes, two prototypes with middle-level overviews with different data scopes, and a video list determined by YouTube as a baseline. Participants explored and watched videos using one of the prototypes in an hour. We examined the videos participants watched, the responses to questionnaires, and the interviews to address the following questions:

- How do participants behave actively when exploring videos with the dual overviews of video resources?
- What will influence the acceptance and usage of comment-generated support for online learning on generic video platforms?

3.4.1 Conditions

In this study, we designed three conditions, each differing in the availability and scope of middle-level representational overviews provided to participants for their video-based learning experiences.

Default List. In the *Default List* condition, participants were presented with a predetermined video list only where there's no video overview support, similar to the conventional video learning experiences (shown in Figure 3.7). Participants can only watch the videos in this list. To ensure uniformity in video resources across conditions, this list comprised identical videos featured in other experimental conditions. The sequencing of these videos adhered to the default algorithms employed by YouTube. To generate this list for each specified topic, we utilized the YouTube Data API [261]. The videos were then arranged in the order they were returned by the API. This approach was deliberately chosen to minimize the impact of individual user profiles on the recommendation outcomes. In addition, it's worth noting that *Default List* provides basic guidance to users compared to learning on YouTube, as they have limited videos available and presented in specific order. In the real-world scenario, users often search the keywords iteratively as they form their learning list [183]. To simulate a more real-world learning scenario on YouTube, participants in the *Default List* condition watched the videos on the YouTube website directly.

Content-centric System. For *Content-centric System*, we provide a video learning prototype that presents middle-level representations of video content alone. It follows the designs of ConceptGuide by Tang et al [258], which focuses on providing a structured overview of video content and recommending video sequence through computational pipelines. More specifically, the content-centric prototype visualizes a concept map derived from the video transcripts and detailed video recommendations. In other words, this condition presents users Content Atlas and Content Detail Panel only, as shown in Figure 3.6. Compared to Kentaurus, we use this system as an example system that primarily utilizes video content (e.g., texts or images) and metadata, which captures a common approach to video navigation.

Kentaurus. As we discussed above, participants were facilitated with both video content and social comment representations and detailed exploration panels (i.e., Content Diagram, Content Detail Panel, Commentary Treemap, and Commentary Recommend List).

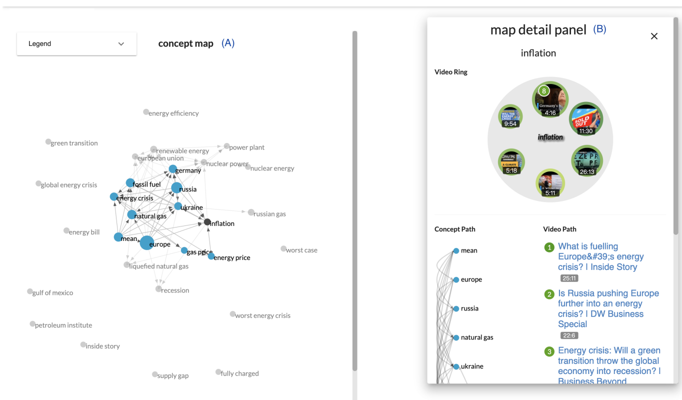


Figure 3.6: Content-centric System [258]

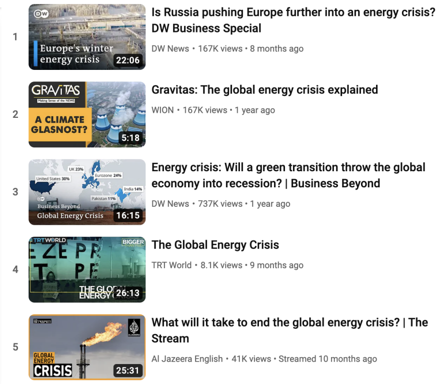


Figure 3.7: BaseList condition (on topic "Energy Crisis")

In summary, these three conditions consist of different scopes of resource overviews and video recommendation approaches for online video learning. As we propose the potential of comments for interactive mode, users still need support for video content, as they are the main information sources related to more fundamental learning modes (e.g., active and constructive). Thus, we selected *Content-centric System* and Kentaurus, one content-centric and another with both channels, but did not include a comment-only system in the experiment. To mitigate the influence of video resources on learning experiences, the videos available in three conditions are all the same. In terms of user interfaces, however, we acknowledge that the video list and video player in the *Default List* are more familiar and visually appealing to users as they access YouTube directly, while Kentaurus and *Content-centric System* are prototypes we built as shown in the Figure 3.1 and 3.6.

3.4.2 Participants

We recruited 66 participants from two university campuses in different countries, through emails and school social media. The participants were required to be at least 18 years old, fluent in English, and watched informative videos online. We selected for a diversity of backgrounds, including college or majors, video watching frequency, and online learning frequency. Among them, three were pilot participants and three others were removed due to missing data or attention check failure during the study. It led to 20 participants in each condition, aged 18-37 (mean=22.98, sd=3.90). Thirty-five of them self-identified as female, 23 as male, and the other 2 preferred not to disclose or disclosed as non-binary. In terms of their video learning frequency, 58.3% of participants watch online videos for learning one to multiple times a week, 16 participants learn through videos everyday, and others learned via videos less often. As participants' disciplinary backgrounds differed, we screened them before the study, and no one self-reported as an expert in the topic they were assigned to watch. The study lasted for about 90 minutes, and participants were compensated with about $12 for their time. The study was approved by both campuses' institutional review board.

3.4.3 Study Protocol

We designed this study as a between-subject experiment. Participants were randomly selected to learn one of two distinct topics: Natural Language Processing (NLP) or the energy crisis. Each participant was given 60 minutes to learn the videos in one of three conditions, with a 27-inch monitor and a desktop. Participants first signed the consent form and completed a pre-study survey for demographics and prior learning experience. Participants in Kentaurus and *Content-centric System* conditions then spent 5 minutes watching a brief tutorial explaining the main functions in the interface in their respective conditions. Participants also completed a hands-on practice session with another example topic, to increase their familiarity with their system and mitigate the learning effects of the video learning task. Participants were instructed to learn the topic assigned broadly and deeply, and learn at their own pace. They could take notes, change video player speed, etc., as they learned before. Upon completion, participants were asked to complete a survey that evaluated their experience and reflected on their learning processes, particularly focusing on perceived social interactions, learning activities. We also interviewed participants to garner more nuanced feedback about their video-watching strategy and experiences. Depending on participant availability, these interviews were either full-length or condensed to a brief short session. A total of 34 participants engaged in full-length interviews — 12 from the *Default List*, 9 from *Content-centric System*, and 13 from Kentaurus.

3.4.4 Videos Materials

As mentioned earlier, we selected two emergent video topics for participants to learn in this experiment: Natural language processing (NLP) and Energy Crisis in Europe in 2022. NLP, a field that has seen remarkable growth over the last decade, intersects with a multitude of real-world applications, such as chatbots and machine translation. It is noteworthy that the videos set for

NLP were retrieved before December 2022 and therefore none of them referred to ChatGPT. On the other hand, Energy Crisis in Europe emerged as a focal point of discussion in 2022, significantly impacting socio-economic structures across the continent and receiving widespread media attention [125].

For the NLP topic, there were 17 videos selected with upload dates ranging from 2016 to 2022. This selection provides a longitudinal perspective on NLP's evolution. The videos mainly adopt lecture formats, featuring slides, whiteboards, and occasional animations. They originate from industry leaders like Google and IBM, as well as reputable YouTube channels such as Crash Course. Regarding the Energy Crisis, we included 24 videos uploaded between late 2021 and December 2022. These videos predominantly comprise news reports, interviews, and documentaries, produced by mainstream media outlets. Videos in each topic vary in presentation style, content, and social popularity regarding view count and comment count, representing the diversity of videos available on general video platforms. The inclusion of these two distinct topics enables a robust assessment for Kentaurus, while also allowing us to explore potential variations in user behavior based on the nature of the topic and the sources of the content.

3.4.5 Data and Analysis

Video Selection and Logs

To investigate the results of active video exploration across conditions, we qualified and measured the video dropout rates and the quality of videos participants watched. As different conditions have separate interface designs, we focused on the final user decision: videos participants checked in the video player and videos actually watched. We compared the average quality of video participants fully viewed based on the expert evaluation, referred to as **full-view video quality**. Specifically, We logged participants' browser activity from the website. We extracted the start and end timestamps of each video, enabling us to calculate the watch duration. Additionally, we tabulated the sequence and count of videos each participant fully viewed. Multiple participants mentioned that they dropped out of videos when the content didn't fit their expectations after watching the introduction or skimming through the video. Thus, we also counted videos that were watched for less than 60 seconds or with a watch rate (time watched / video duration) lower than 10% as videos that participants dropped out, as **video dropout count**.

As an approach to understanding the differences between videos participants watched, particularly for the video quality, we invited one graduate student specializing in NLP to rate the NLP videos, and another Ph.D. candidate with expertise in economics and political science to assess the Energy Crisis videos. Videos were rated on a scale of 1 to 5 across six criteria: presentation clarity and coherence, presentation structure, content credibility, accessibility to a general audience, inclusion of contextual background (i.e., Does the content incorporate contextual background of the topics discussed?), and depth of content (i.e., Does the content present different perspectives or complex concepts?). The expert score for each video was calculated as the mean of these six metrics. To ensure scoring consistency between the two experts, we standardized the scores within each topic category to yield the final video quality score for each video. (When the score is positive, the video is above the average quality of the videos.) See the equations below for more clarification about the measures for video selection.

$$\text{video\_watch\_rate}_{ij} = \frac{\text{duration of video watching by participant } i}{\text{length of video } j} \tag{3.1}$$

$$\text{video\_dropout\_count}_i = \text{\# of video } j \text{ whose video\_watch\_rate}_{ij} < 0.1 \text{ or watched } < 60s \tag{3.2}$$

$$\text{full-view\_video\_sequence}_i = [\text{video}_{i1}, \text{video}_{i2}, \text{video}_{i3}, \ldots, \text{video}_{im_i}] \tag{3.3}$$

$$\text{full-view\_video\_quality}_i = \frac{1}{m_i} \sum_{n=1}^{m_i} \text{expert score of video}_{in} \tag{3.4}$$

where i =1,2,...60 participants, j refers to video index, and m_i is the size of video finished for participant i.

Perceived User Experiences

After participants finished the learning session, they shared their general learning experiences on a 7-point Likert scale, from 1 (strongly disagree) to 7 (strongly agree). Specifically, we gauged participants' perception of social connection during video watching through three items (e.g., "In the experiment, I felt connected to other users who commented under the videos") derived from questionnaires in [118, 133]. Participants also self-reported their perceived guidance when interacting with the system or list navigation via another three items (e.g., "In this experiment, the interface helps me to know what's next to learn") from [258]. Participants also completed the NASA TLX to report their cognitive load [108].

Analysis

With the quantitative data available, one-way mixed-effect ANOVA was used to examine the influence of Kentaurus on video-related metrics. The video topic and recruitment site were used as random variables. Participants' familiarity with the video topic, frequency of online video learning, and other demographic information, such as gender and education level, were used as covariate variables. We performed Tukey comparisons for post hoc comparisons. For self-reported scales, we performed Kruskal-Wallis H tests to compare three conditions because the data were not normally distributed.

We conducted semi-structured interviews and utilized thematic analysis to understand more about learning processes and the factors that influence participants' video exploration and selection. We transcribed the interviews into transcripts and segmented them into two main topics: video exploration and comments related to learning results and efficiency. We applied thematic analysis combining both axial and open coding. We first identified recurring codes related to these two topics. Three researchers then refined the codes and merged the initial codes into broader themes and revised and reassigned the codes until consensus among the authors.

3.5 Results

In this section, we present the findings from overt active learning behaviors and users' learning experiences from their sharing. We mixed the objective measures such as video logs and survey and interview results. Similar to the discussion in ICAP framework [47], we note the boundaries between learning modes and activities are not rigid and exclusive; and there may be potential discrepancies between observed behaviors and the underlying learning processes participants had, attributable to individual and contextual variances. Thus, we used diverse proxies to reflect the influences of Kentaurus on users' behaviors and perceptions. These results were introduced from relatively passive to more active actions, while these actions were not rigidly assigned to specific active learning modes.

3.5.1 Active Video Selection

Participants watched 7.45 videos on average in the learning session (sd=2.02) and these videos mentioned similar amounts of video concepts in the Content Diagram (mean = 24.07, sd= 3.11), and keywords in Commentary Treemap (mean = 25.89, sd = 4.97). These content coverages were not statistically significant using T-tests across conditions. With the same information visited in the end, however, we found significant differences in the quality of videos that participants watched **full-view video quality** ($F(2, 50) = 5.11$, $p < 0.01$, $\omega^2 = 0.13$). Post-hoc Tukey test indicates that participants using Kentaurus (mean=0.28, sd=0.27) watched statistically significant higher quality videos compared to *Default List* (mean=0.03, sd=0.30): Kentaurus - *Default List*: $z = 3.09$, $p < 0.01$; *Content-centric System* (mean=0.21, sd=0.31) also helped participants to watch videos of marginally significantly better quality ($z = 2.16$, $p = 0.07$). There is no difference between Kentaurus and *Content-centric System* ($z = 1.02$, $p = 0.56$).

In addition to the expert quality of videos, the frequency with which participants stopped viewing videos (**Video Dropout Count**) offers ancillary insights about video exploration efficiency with varied aids. ANOVA result shows that there is a statistically significant effect of prototypes on **video dropout count** ($F(2, 50) = 12.91$, $p < 0.01$, $\omega^2 = 0.31$). Tukey tests further reveal that participants using *Default List* (mean=5.05, sd=4.12) dropped far more videos than participants in both *Content-centric System* (mean=1.10, sd=1.37) and Kentaurus (mean=1.30, sd=1.75): *Default List* - *Content-centric System*: $z = 4.61$, $p < 0.01$; *Default List* - Kentaurus: $z = 4.03$, $p < 0.01$.

From the video selection and dropout results, content-centric learning representations contributed significantly to improving the video selection quality compared to the default video list where participants selected average-quality videos (mean of normalized video quality is near zero), and reducing redundant video previews in the video player. Moreover, comment-oriented helped to further enhance the quality of the videos participants selected, potentially due to the additional video preview and resources available to choose the videos.

3.5.2 Video Exploration Experiences

Besides the influences on video quality selected and video dropout, we analyzed participants' video learning processes via a blended analysis of their video exploration process and guidance. By doing so, we explore the nuanced user experiences different conditions provided.

As proposed above, for active learning on generic video platforms, it is both important to encourage participants to explore actively and provide structured facilitation via automatic algorithms and interface designs. To investigate how participants tell the guidance from middle-level representations, we analyzed the survey reponses post the learning session. Self-reported guidance remained similar across all conditions (*Default List*: mean=5.31, sd=1.04; *Content-centric System*: mean=5.01, sd=1.54; Kentaurus: mean=5.43, sd=1.25), as supported by a non-significant Kruskal-Wallis test result ($\chi^2 = 0.69, p = 0.71$). Across the system support, users didn't perceive one method as inherently superior in aiding their video exploration. The fact that we did not observe the statistical difference is consistent with the quantitative results about overall information visited in the section 3.5.1. This speaks to the strength of commercial recommendation solutions and participants' trust and reliance on automatic algorithms.

Nevertheless, this inconclusive result masks substantial qualitative distinctions when given different support diverse in format and channels. For instance, six participants attributed their sense of guidance to the simple list as they didn't need manual search and filtering. For instance, P108 further explained *"This Default List is helpful for me because I believe it's not random. It's a list similar to search result recommendations from YouTube, but sometimes [on YouTube] you have to pick."* In Kentaurus, participants utilized some common video exploration strategies by checking video uploaders, titles, views, and duration. More importantly, they developed personalized video exploration processes, reflecting their attitude toward comments, evolving understanding of the topic, and their individual impression of the recommendation and summary capability of our computation framework. We will delineate these findings in the following subsection. Some insights were gleaned from participants using *Content-centric System* as they shared the designs of Content Diagram and Content Detail Panel.

Diverse structured summary and personal interests shaped the video exploration.

The availability of structured overviews and diverse recommendation channels helped participants start the video learning, combined with their personal preferences. Many participants checked the Content Diagram and then followed the visual cues by selecting important concepts to choose their first videos. E.g., *"I click on the biggest node in the map, watched the first video, and then moved to the second largest one to watch videos"* (P308). And their video selection improvised based on their interests once they have a general understanding of the topic. E.g., They may *"explore the outskirts of the graph"* (P310) or *"look for videos that caught their eyes"*(P314). Participants also noted they utilized the connections between the concepts to find relevant videos to watch, and to explore broader understanding of the topic with subtopics that they hadn't known. P303 noted that *"It [Content Detail Panel] automatically shows other points with the connection, which gave me more directions and options"*.

The presentation of Commenetary Treemap, on the other hand, facilitates users to prioritize and streamline their learning process. Participants noted that they saw the Treemap to *"prioritize the concepts they needed to know to grasp the whole topic in a short limited time, especially key and fundamental concepts"* (P305), and decided *" whether to explore deeper or change another direction"* (P303). Also, participants considered the unfamiliar comment themes and keywords to discover minor subtopics they were interested in, but not available from the content summary. For instance, both P312 and P313 remembered that they *"saw the comment 'green transition' and selected a video there [Commentary Recommend List]"*. Some representative comment details also triggered participants to revisit some concepts in the Content Diagram. E.g., *"Some comments were not the same as the idea of the video, I would go back to the knowledge points mentioned in the comments to watch some video."* (P3010)

As for selecting one specific video among some candidates provided in the video recommendations, coverage of concepts as preview of the video content worked as a crucial indicator. For instance, P3007 mentioned *"videos that have more concepts and subtopics covered"*. Meanwhile, three participants used comments-related features (i.e., Commentary Treemap and color of Video Ring) to verify video selection, such as *"choose well received, commented videos"* (P315). In the *Default List*, on the other side, participants also shared their selection strategy to narrow down the search. They mainly relied on video thumbnails, titles, and duration to select videos, to estimate the video content and presentation. These considerations match the efforts of SEO strategies to optimize video visibility from the producer side [170]. These selection criteria were consistent across conditions, revealing the common requirements for video exploration.

Attitude toward comments impacted their usage and timing of usage of comment-related information.

In the Kentaurus condition, not all the participants actively consumed Commentary Treemap and Commentary Recommend List. From the browser logs, 15 of 20 participants using Kentaurus selected and fully watched at least one video from Commentary Recommend List, but some participants checked the summary from comment-related panels but didn't finish watching a video. For participants who hesitated to read comments or use these two comment-related panels, they were concerned about the quality and relevance of comments, and the limited organized recommendations from social comments. For comments under the video player, two participants (P215, P219) in *Content-centric System* condition shared that they found comments distracting as some of them are *"off-topic"* or subjective. P3007 also shared comments not focused on learning as *"I felt the comments only related to a part of the video which may not represent the whole content"*. Another two participants using Kentaurus also mentioned they preferred Content Atlas as Commentary Treemap was less organized in terms of video sequence recommendation, visualized in *"bullet point styles only"* (P3009).

Compared to participants who are resistant to comment-related knowledge and overviews, participants who were interested in comments and other audience connected them with video content, explored new subtopics, as well as inferred social feedback in the videos, as described in section 3.5.2. We further investigated the timing of selecting videos by checking the Comment Trend Treemap. By checking the time series logs of browser history, we found that 80% of participants started watching videos via Commentary Recommend List 20 minutes after the start of learning session. Participants also reported similar patterns as exploring Treemap after browsing video content summary from their interviews. For example, P315 mentioned that *"after watching videos through that (Content Diagram) to get basic ideas, I saw people asking harder questions in the comments (Commentary Treemap) and clicked on those videos."*. P313 further explained their reason: *"To* *decide my watching sequences, I started from the left (Content Diagram) since it's clear, and when I saw the comment when watching videos, I found Commentary Treemap was also helpful and clicked on those video recommendations. But I also felt we shouldn't overtrust them to form my opinion [about energy crisis]."* Related to the similarity in perceived video exploration guidance across conditions, our granular qualitative analysis reveals the dichotomy in participants' usage and attitude of comment-related overview, video recommendation, and comment details. Overall, multiple factors contributed to the acceptance and usage timing of comment-related information, including the attitude and trust toward social comments, and the roles of comments and video content, among others.

3.5.3 Learning with Kentaurus

Building upon the analysis of active video exploration and middle-level comment representation usage compared to content representation, this section delves into the active learning processes with Kentaurus, from the perspective of comprehension of the topic, cognitive burden, and mixed system satisfaction.

Exploration Process

From the interview, participants using Kentaurus mentioned the benefits on overall understanding and details, content coverage and diversity. They remarked on their holistic and systematic understanding of the video topic with the grounding of the structured overviews from middle-level representations. Nine participants pointed out that Content Diagram and Commentary Treemap supported them in capturing the big picture and the connections between subtopics. For instance, *"I can use this system to quickly understand several important keywords and which videos are related to these keywords."* (P3006). Similarly, P305 recalled, *"I compared the tree maps from the videos and logically located specific concepts with the system and understand where I am"*. Two participants also mentioned the overviews helped *"filling the knowledge gap"*. In terms of details, participants also mentioned they learned about the development and context of events and different perspectives for the same event (Energy Crisis) with reasonable video sequences. For example, *"the first two [videos] explained the different context and the third video summarized the reasons and connected the context together"* (P313), and *"the first long video helped me get the overview, and I watched some news reports to understand the connections of events."* (P3009). P221 mentioned that *"I feel I can't trust ev-*

erything I read, but some people did make good points from the treemap. The comments made me skeptical and kept thinking".

System Usability

As for system satisfaction, participants expressed mixed usability feedback besides the availability of content-based and comment-based guidance, partially due to the learning curve and familiarity with the system operation and capability. Participants discussed the trade-off of benefits and drawbacks of Kentaurus design. P311 mainly remarked on the benefits of the time spent on video search that *"I spent a lot less time trying to figure out what video to watch next, which gave me more time actually to watch the video"*, with 10 others. Some, however, pointed out the learning efforts and time needed to start learning efficiently, for instance, *"As the experiment goes, I gradually felt the help this system (Kentaurus) provided, especially at the end of the learning session. But I was reluctant to learn with it when I watched the tutorials, as I didn't think it would be different from searching on my own."*. *Content-centric System* and Kentaurus introduced a learning curve to participants on interface operation, which may add cognitive burden to users and mitigate the learning efficiency [207]. Therefore, we analyzed the cognitive load scale with the Kruskal-Wallet test. Results indicated no significant differences among the conditions ($\chi^2 = 5.27, p = 0.08$; *Default List*: $mean = 3.77$,$sd = 0.65$; *Content-centric System*: $mean = 3.36$, $sd = 0.89$; Kentaurus: $mean = 3.83$, $sd = 0.81$). This lack of difference suggests that Kentaurus's richer, two-channel representations did not necessarily impose a higher cognitive load than the simple *Default List*. The structured overview from the content-centric overview slightly reduced the cognitive load, consistent with prior study [138], but there's no statistical difference.

Social Connection

Following the design principles (**R1** and **R3**), participants in Kentaurus accessed the thematic summary of comments across videos as they explored videos. Section 3.5.2 talked about how participants utilized the content of comments based on their personal interests and preferences, and learning process. According to prior work on social support for video learning [172], another important facilitation mechanism relies on the ambient social environment and social connections. Following this line, we compared the perceived social connection from the survey items. The

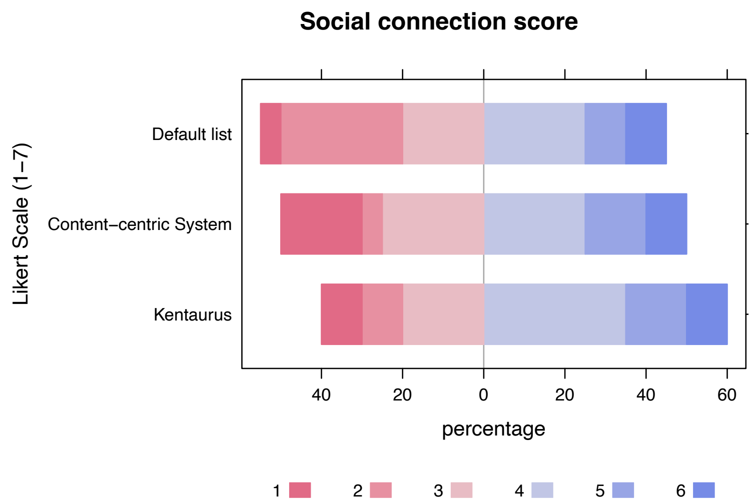


Figure 3.8: The distribution of social connection scores in three conditions

differences were not statistically significant from Kruskal test result ($\chi^2 = 0.38, p = 0.82$). Still, the distribution of the social connection score showed that participants in Kentaurus felt more connected with audiences who shared the comments, and trusted them, as shown in Figure 3.8. (*Default List*: $mean = 3.42$, $sd = 1.40$; *Content-centric System*: $mean = 3.40$, $sd = 1.53$; Kentaurus: $mean = 3.65$, $sd = 1.37$). In summary, Kentaurus didn't introduce direct improvement in social connection and social guidance; however, it helped participants apply social comments whenever they need.

3.6 Discussion

Learning actively on generic video platforms like YouTube requires diverse user video exploration behaviors and appropriate levels of knowledge guidance and automatic support, ideally across information sources. In this paper, we instantiate a system, Kentaurus, which presents middle-level structured overviews of learning materials and social comments for users to explore a learning topic, alongside pertinent video recommendations. Through a user study, we have gleaned empirical insights into how users participated in active learning given such affordances and selectively utilized comment-based support. In this section, we discuss these findings and elucidate their implications for the future design of online active video learning support.

3.6.1 Audience-generated Comments for Active Learning and Video Exploration

Kentaurus harnessed and integrated thematic summary from audience comments to enrich the active learning experiences. We observed that access to such comment-based overview and recommendations changed the video explorations from different angles, while users didn't perceive a higher social connection with other audiences. the additional opportunity to access audience sharing didn't significantly improve the social transparency of other learners. One potential reason is that the structured overview highlighted the common themes of comments. It might reduce the social presence of individual commenters, which may help create a more conducive social environment for learning [151]. From the content side, participants referred to the comment keywords for video popularity and followed the video recommendation ranked by comment sentiment positivity. Notably, many of them draw upon the structured comment overview (i.e., the Commentary Treemap and corresponding video lists) as social shortcuts to adjust their learning processes and exploration directions. They helped participants refine their learning processes and exploration pathways, leading them to significantly higher-quality video content as in Section 3.5.1.

As the study unfolded, participants embraced comment-based support in different ways. To decide whether to use them, some participants expressed their concerns about the credibility and relevance of the content through the system managed to reduce a portion of irrelevant comments, and they chose to neglect them when they couldn't verify the correctness. This dichotomy mirrors tensions highlighted in previous research about users' trust and authenticity consideration on audiences' notes and summaries [172]. When it comes to the timing of using them, individuals appeared to develop their own strategies, based on their perceived comprehension of the topic at the moment, time budget, and others from interviews and surveys. Partially due to these different usage considerations about comment-based support, participants perceived better facilitation for social connection and less guidance compared to the Content-centric System, while none of them is statistically significant. Drawing together the aforementioned findings, it becomes clear that incorporating flexibility into the design for audience-generated support and structural visualization is not just beneficial but essential. Future systems may consider the approaches to cater to users' learning strategies and interests about when and how to present other learners' comments for video exploration, leaving the options to users or adapting as users learn.

3.6.2 Active Learning Under Recommendation and Structure Support

Despite featuring three distinct experimental conditions with diverse recommendations and video selections within a pool of about 20 YouTube videos, participants across conditions visited a similar amount of videos and engaged with a comparable range of concepts and keywords within the 60-minute session. This suggests a high degree of content overlap and redundancy within the videos on generic video platforms, highlighting the saturation of the conceptual landscape.

To support learning on option-dense yet knowledge-sparse resources, Kentaurus introduces middle-level representations with structures and interactions between two information sources. These representations foster active exploration by structuring and interlinking two primary information sources. Beyond merely enhancing video quality through active selection, the majority of our participants recognized the value of both structured overviews. Not only did they guide learners through a blend of video and social data, thereby avoiding a linear watching strategy [242], but they also facilitated the construction of a broad topic outline and anticipated information encounters (as designed in **R2**).

Considering the manual efforts, our results highlighted the balance needed between active learning activities and reliance on automated support. The comparison between the Content-centric System and Kentaurus indicates that integrating both approaches did not necessarily increase cognitive strain—a promising finding for future work pursuing a hybrid approach to encourage active learning experiences by increasing the knowledge sources. However, the lack of a distinct advantage from participants suggests that more exploration options do not always equate to better user satisfaction: for example, an overabundance of video choices might lead to selection overload which drops the perceived system guidance. This observation raises a critical question about the extent to which the amount and type of active learning activities should be supported, ensuring appropriate integration without exceeding cognitive load or creating a steeper learning curve. While our results don't provide definitive answers, they underscore the importance of design decisions to manage this trade-off. Future research might delve deeper into these design challenges, taking into account factors such as users' digital literacy, background expertise, and attitudes towards data sources and algorithms.

3.7 Limitation and Future Work

Kentaurus currently relies on computational pipelines based on pre-trained transformers and can take a considerable amount of time to generate data, depending on the video corpus and computational capacities. Thus, we do not claim that our pipeline is the most accurate in this line of work. For instance, there was not a one-to-one correspondence between abbreviations and their full phrases, as evidenced by "AI" being displayed exclusively with "Artificial Intelligence" in Content Diagram. With the rapid development of Large Language Models (LLMs), future studies can work on generating better summaries from videos and comments, as well as texts and data from other modalities. On the other hand, we chose to use network and treemap to deliver the structure of content for videos and comments and provide "overview first, then details" exploration interface. Though they are common visualization options for video-based learning or community information visualization [214, 298], future work may explore alternative visualization designs for better user experiences and learning performance.

Our lab experiment has several limitations. The lab study cannot infer the long-term impact on participants. Some mentioned they did not have a full mental model of the computational pipeline of Kentaurus until the end of the experiment (P312 and P3007). As users resolve the learning curve, we may observe different feedback when users are familiar with the interface and algorithms. We selected a predefined video playlist displayed on YouTube as an alternative to Kentaurus and Content-centric System, while there are other natural options, such as learning on YouTube directly. We chose it to control the video pool participants may access, and it turned out as a competitive baseline: multiple participants mentioned it as a simple and guided learning condition compared to YouTube itself. However, this also happens when experts or YouTube provide a watching list for learners, and future work may explore more real-world alternatives. The two learning topics might not be what participants were interested in learning. We mitigate these risks in the study through practice sessions and choosing the tasks based on YouTube trends and public search on Google. In addition, a few participants reported that they planned to learn the topic for their professions or out of interest after the learning session (P108). Future work could transfer similar designs into plugins and conduct longitudinal field studies to understand participants' long-term usage and the impact on learning strategies.

To provide diverse recommendations based on past learners, future work may also consider more resources to personalize the learning navigation and provide adaptive recommendations in real-time. For instance, adaptive overview guidance with varying levels of detail could cater to learners' learning progress and improve their perceived guidance. Other user feedback and learning processes could be utilized for video navigation support, such as peer learners' learning paths [193]. We could also leverage crowdsourcing or LLMs with domain knowledge to generate high-quality comments and personal learning notes to improve the quality of comment-based overviews for user adoption [166].

In this work, we focus on helping people learn on generic video platforms, like YouTube, as avenues for educational content. As evidenced by our user study and prior studies, people can significantly benefit from these resources when supported by structured learning scaffolds. However, the entertainment and open nature of YouTube also presents challenges, notably in assessing the credibility of information and maintaining focus amidst distractions revealed in our results. This is particularly true for user-generated content such as comments and videos from non-experts [242]. Contrasting with YouTube are platforms like MOOCs, Vimeo, and TikTok, each with unique video scopes, community attributes, and interaction dynamics. As such, people may adopt different strategies for learning when supported by scaffolds like Kentaurus. Future research may delve into these diverse learning environments to understand how they influence learning strategies and outcomes, particularly in the context of active learning.

Chapter 4

Structured Thought Documentation and Sharing to Nudge Reflective Mind for Analysis

---

For high-stakes domains (e.g., healthcare), it's crucial to analyze knowledge beyond surface-level understanding, such as assessing credibility and integrating different views. However, people may not have sufficient cognitive resources to trigger these processes. To address this, we introduce to leverage structure-documented personal thoughts as social nudging, offering users to see peers' notes and share their own analysis and evaluation thoughts related to the same content. Social nudging may engage individuals in active thinking and foster in-depth thoughts through unconscious social mechanisms. We conducted two studies to investigate the potential of social nudging in coordination and cooperation scenarios and compared the influence of social nduging across normal and structured documentation tools. Results reveal that social nudging enhances the engagement of both reflective and critical thinking especially in cooperation settings, and elevates users' attitude confidence. These findings also contribute to expanding design space of nudge technology to nudge plus, which combines the nudging mechanisms and thinking scaffold such as structuring sharing via concept mapping.

This chapter is expanded from the following short paper and demo: (1) Liao, J., & Wang, H. C. (2022, April). Nudge for Reflective Mind: Understanding How Accessing Peer Concept Mapping and Commenting Affects Reflection of High-stakes Information. In CHI Conference on Human Factors in Computing Systems Extended Abstracts (pp. 1-7) [160]. and (2) Liao, J., Singh, M., & Wang, H. C. (2023, October) and DeepThinkingMap: Collaborative Video Reflection System with Graph-based Summarizing and Commenting. In Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing (pp. 369-371). [158]

4.1 Introduction

With the remarkable popularity of generative artificial intelligence and other technological innovations, the authenticity, credibility, and quality of content available online have become focal points for both researchers and the general public. This trend fuels the rising necessity to analytically approach information in high-stake domains, including healthcare and financial investments [83, 178]. When interpreting information, individuals have to not only assess the validity of content and detect misinformation, but also apply this knowledge to specific applications (e.g., take appropriate diet suggestions based on current health status). Factors like confirmation bias, anchoring bias, and sociopolitical influences can obstruct a comprehensive interpretation of online content [4]. Therefore, empowering people to transcend inertial thinking and consciously process and reflect upon online information is essential for the well-being and safety of both individuals and communities [89].

To process information beyond mere literal comprehension, individuals may make substantial cognitive efforts to integrate, analyze, and reflect upon such information. This involves extensive engagement in higher-order thinking (colloquially, practices that go beyond memorizing information or retelling stories), which is operated consciously by individuals [73]. Higher-order thinking encompasses two major categories: reflective and critical thinking, both essential to decide how the content is applied to their lives (e.g., taking vaccination or not) [90]. For example, reflective thinking can potentially help people deconstruct the underlying structure of information across various sources, identify what's relevant and meaningful, and bridge the gap between online content and existing mental models they hold [90]. In parallel, critical thinking may help keep people open-minded while grappling with contradictory information, especially when it challenges previously held beliefs [146].

Despite the recognized significance of these cognitive endeavors, higher-order thinking is often so cognitively taxing that they are not uniformly accessible across diverse circumstances and among different individuals [79]. This leads to a tangible need to facilitate people to trigger their reflective and critical thinking. Inspired by the boosting effect of nudging technology on behavior change (e.g., showing others' behavioral records to motivate people start similar behaviors, such as exercising for health) and reflection (e.g., reflecting on their political views by seeing other people's views) [97, 98], we note that social nudging may have potential to support individuals in this complex and cognitively demanding process. Social nudging, as a choice-preserving intervention, leverages social influence to guide individuals, such as through invoking reciprocity and increasing the visibility of peers' actions [40]. It aims to alter individuals in a predictable manner without constraining any options or changing incentives [262]. Cognitively, nudging is often less demanding than formal thinking activities like reflective journaling, and can minimize the likelihood of burnout and negative reactions [262, 272]. Also, social nudging doesn't limit the types of social heuristics to reach the desired goal, which is flexible for different social interaction or collaboration scenarios [40]. The implicit and flexible nature of social nudging makes it a promising tool to foster critical and reflective thinking as it allows open expansion of thoughts and leaves final decisions to individual thinkers.

Many studies have utilized normal individuals' and groups' efforts and wisdom to support general interpretation, sensemaking, and other specific tasks, such as misinformation detection [40, 63, 98, 152]. These approaches range from crowdsourcing dispute warnings within or across platforms [186], detecting misinformation from social comments with NLP [243], and redirecting users to papers with different views and comments [209]. While these studies have designed different ways to curate social contributions to assist people, only a few have considered using peers' sharing as thinking prompts rather than information sources (e.g., display of comments and alternative opinions) and examined the thinking engagement of such support when social interaction level varies. Moreover, in terms of technology designed to mediate personal sharing as nudge cues during the nudging processes, how deeply individuals could engage has not been thoroughly examined. Investigating these nuanced socio-cognitive processes over technological mediation will help understand the underlying processes needed for thinking processes and following behavior change.

In this work, we aim to explore the impact of peers' personal thoughts as social nudging on engaging individuals in higher-order thinking—namely critical and reflective thinking—and assess the effects of thought documentation and sharing tools on this engagement. We hypothesize that observing how others engage with and react to the same information may provide valuable behavioral cues (e.g., "Do others reflect on this content?") and cognitive stimulation (e.g., "How are others responding to this content?"), thereby fostering engagement in higher-order thinking.

To further understand whether social nudging will perform differently on thought documentation methods, we implemented an interface, DeepThinkingMap, for simultaneous concept mapping during information consumption. It collects peers' thinking activities and guides users to organize their thoughts and connect with others' notes. As an alternative mediation tool to document and share personal thoughts, it may also serve as a cognitive scaffold as sharing via concept mapping helped individuals better convey their understanding and engage in robust inference making and critical thinking [45, 104, 277]. As depicted in Figure 4.1, DeepThinkingMap consists of two components – an interactive concept-mapping-akin workspace and a content viewer displayed adjacently. Users may not only share their takeaways and discuss them with others on the workspace, but also connect notes in multidirectional and nonlinear manners, which may help preserve the complexity of individuals' own thoughts and encourage the quick links of thoughts among each other [282]. By comparing DeepThinkingMap with others common documentation approaches such as Google Docs, we may uncover nuanced insights into how documentation tools might modulate the influence of social nudging on higher-order thinking processes.

We conducted two studies to delve into the complex interplay of social nudging and documentation tools: Study 1 focuses on the impact of social nudging; Study 2 is a comprehensive exploration of social nudging and personal sharing tools. In Study 1 ($N = 44$), we used notes of median quality from prior users as social nudging cues to explore the effects of non-interactive social nudging in a relatively well-controlled setting when participants all used DeepThinkingMap. The findings from Study 1 revealed that merely viewing average-quality notes from prior users could stimulate reflective thinking, even through such asynchronous interaction scenarios. Building on the insights garnered from Study 1, Study 2 ($N = 52$) incorporated a more complex synchronous setting to examine the impacts of interactive social nudging and the difference between DeepThinkingMap and Google Docs (as an example of common documentation approach) [96] in a scenario mirroring real-world complexities. Study 2's findings not only corroborated those from Study 1, but also underscored the pronounced influences of social nudging and DeepThinkingMap on critical thinking and attitudinal changes. Results from both studies provide consistent results showing that access to peers' thoughts and social interactions with others can (1) enhance the engagement of reflective and critical thinking, (2) promote the analytical level of thoughts during interactions, especially when using the concept-mapping-akin documentation space, and (3) strengthen users' confidence in their attitude toward the topic. Moreover, the influence of social nudging was further strengthened with thinking scaffolds via documentation and sharing tools.

In summary, the primary contributions described in this paper are two-fold:

- We investigate social nudging as a non-forcing thinking intervention for high-stakes information processing. By examining social nudging in two studies, we demonstrated its potential to enhance individuals' reflective and critical thinking and solidify their confidence in the topic.
- Our empirical findings reveal a significant synergistic effect when combining intuitive social nudges with deliberate thinking interventions (i.e., Nudge Plus), such as concept mapping, shed new light on the design of nudging blended with other thinking scaffolds.

4.2 Background and Related Work

4.2.1 Higher-order Thinking and Tool Design in HCI

Higher-order thinking is crucial in today's era of information proliferation, as it paves the way for innovative ideation and solutions [103]. According to Lewis and Smith, higher-order thinking occurs when a person takes new information and information stored in memory, then integrates and extends them to achieve a purpose. [155]. Online information consumption often requires higher-order thinking, including decision-making for content selection, integrating diverse sources, creating a coherent mental model of the topic, and evaluating opposing viewpoints [90]. We adopt Geertsen's framework to explore how social nudging may facilitate two main types of higher-order thinking: critical and reflective thinking [90]. Reflective thinking involves activities that tend to connect different thoughts and enlarge the space for thinking, such as analyzing the logical relations between ideas for sensemaking, and finding possible similarities between the known and the new learned. Critical thinking consists of establishing the accuracy and validity of content through self-examination, such as critically evaluating a news article's sources to determine its credibility. Despite the distinctions in their definitions, they are both vital in processing high-stakes online information that may include unverified content [89]. Taking diet suggestions as an example: people need to not only assess one source of so-called facts that support the benefits of certain diet strategies (i.e., critical thinking) but also compare them with health knowledge they have known and collect other resources (i.e., reflective thinking).

We note that some HCI frameworks categorize reflection and introduce both reflective and critical thinking as different levels of reflection, including Mezirow [190] and Fleck and Fitzpatrick [82]. Opting for Geertsen's framework, we use the term "higher-order thinking" to distinctly discuss these processes and highlight their importance. By avoiding readers' pre-existing notions about "reflection," we hope to sidestep potential ambiguities.

To support higher-order thinking, designers have explored various strategies to design tools to collect and display relevant extra information for users as scaffold [21, 27, 49]. Bentvelzen et al. outline four design approaches for developing design interventions for thinking activities: temporal overviews, conversations with other human or computer actors, comparison of alternatives, and discovery of new perspectives [27]. For example, NewsCube [209] clusters different event viewpoints and offers an unbiased clustered overview to nudge people to support reflective thinking of different perspectives in the consumption of news. Similarly, Trackly [15] employs customized pictorial trackers of personal data to prompt reflective moments and preserve individual agency.

When it comes to social components fostering reflection and critical thinking, studies have made initial attempts to utilize community senses, social norms, emotional support, and content sharing. Halbert and Nathan has explored the role of social discomfort, such as discussing problematic scenarios, as a design opportunity for elevating critical thinking [102]. TalkReflection supports collaborative reflection in work settings through sharing and commenting experience reports about interactions with clients [219]. In summary, previous work on tool design mainly concentrates on augmenting individuals' cognition by lowering the barriers to accessing overviews or information comparison. While some research explores the role of social contexts and generic collaboration in fostering higher-order thinking, our work aims for a nuanced understanding of social nudging, focusing on the behavioral and cognitive utilities of enabling social transparency of others' actions as the basis of higher-order thinking support.

4.2.2 Social Nudging

With the proliferation of nudging techniques, the HCI community has integrated nudging into various intervention designs in the domains of health informatics, recommendation systems, and news consumption [40]. As different definitions apply to different sectors, *nudging* is generally viewed as a lightweight intervention directing people towards particular choices while preserving all options [254]. Nudging was initially designed to happen automatically without much conscious awareness and trigger low cognitive capacity in behavior public policy [262]. And as the nudge theory and applications developed over the past decade, nudging strategies have advocated the inclusion of reflective elements to improve the effectiveness and efficacy period of nudging [18]. Several literature reviews have discussed different aspects of nudging within HCI literature, such as working mechanisms and design patterns [28, 40]. As our work specifically focuses on social nudging to promote higher-order thinking activities, we focus on the work related to nudging via social influence in this section, aligning with Caraban's categorization [40].

Social nudging utilizes socio-cognitive mechanisms like social comparison and visibility of user actions to guide individuals toward optimal outcomes [40]. As people tend to copy others' actions to emulate behavior in certain situations based on social proof theory [54], both the peers' actions and content are important for social nudging as they provide references and evidence to how others behave, and how others react. For instance, Pinteresce, a tool to motivate senior citizens to connect to online communities, used family members' past stories to prompt online engagement [36]. For active video watching, tagging and rating comments are designed as extra nudging actions to encourage critical reflection and compassion [63, 152]. For health behavior changes, such as fitness tracking and calorie control, scholars have researched various subtle social nudging designs. With the design of glanceable behavioral feedback for self-monitoring, Normly was evaluated with various visualization designs [98]. Another line of work by DiCosola and Neff studied the timing of nudges on e-grocery shopping and introduced comparisons from both in-groups and out-groups [174]. Aligned with related research on technological support for higher-order thinking, these studies offer a deeper dive into the social-related nudging mechanisms and empirical insights.

As researchers have tailored nudging designs for specific tasks, the effectiveness of nudge designs is context and user-group dependent. For example, disputes to nudge people from overeating may be futile due to fatigue over time [254, 255]. Additionally, there have been mixed empirical results when users receive nudging content from groups far different from their own [64, 175]. When a reference group (i.e., the group of people that matters to individuals) is not relevant to specific users, disseminating information about that group's behavior has little nudging effect [30]. Moreover, the desire to conform to social norms can inhibit individual thinking, and fear of judgment may lead to aligning with group views which leads to groupthink and worse group performance [61, 114]. Thus, it is crucial to account for these nuances in nudging design. Learning from these observations, our study employs peer thoughts as social nudging content, using similar individuals as a reference network to encourage thoughtful engagement with nudging cues.

4.3 Social Nudging and DeepThinkingMap Design

4.3.1 Social Nudging Mechanisms

To facilitate general users to engage in higher-order thinking with voluntary, cognitively manageable opportunities, we posit that social nudging may have positive influences and is applicable to diverse social scenarios. Social nudging guides users' behaviors and attitudes by leveraging social influences and cognitive bias in a non-coercive way [40]. It often starts with identifying the prevailing peer behaviors or social norms in the community. Then users adhered to the common behaviors they perceived or observed, and reinforced their behaviors based on feedback if available. This process may take advantage of different social facilitate channels, depending on the social interaction scenarios. For instance, observing peers engage in thinking tasks may serve as a social proof heuristic, guiding individuals to use similar strategies [53]. This tendency is irrelevant to the quality of peers' thinking results [43]. Especially when the majority of the community (large or small) engages in higher-order thinking, people are more likely to spend more effort due to bandwagon effect [233]. To leverage this line of social heuristics, we installed behavioral transparency among peer users by making visible how others process the online information, for instance, by reflecting or criticizing the information source. As a result, users may not only observe the final thinking results (e.g., comments), but also observe a part of the behavioral process. Both could act as nudging cues for users. Building upon the behavior cues, we take the nudging content into consideration. As peers share more personal thoughts, there will be a larger chance that people will receive information inconsistent with their own mental model. Then individuals tend to lower their rates of quality of content, and become more careful (regret aversion bias) [267]. It can nudge them to stay open to different perspectives and evaluate every possibility, such as re-processing the content to find verdicts or adopting a new perspective to reduce the cognitive conflicts experienced [162].

There are a few related research questions that our social nudging does not address. First, we don't aim to narrow down the scope of higher-order thinking to specific tasks, such as misinformation detection and decision making. In our experiments introduced later, we implemented a health-related information consumption task which may lead to decision making (e.g., "choosing turmeric or not") or misinformation debunking (e.g., "the safety of vaccination"), depending on the videos participants watched. Still, our work focuses on the higher-order thinking processes triggered, as one prerequisite for these final intelligent decisions. We believed that social nudging helps to mitigate these issues when higher-order thinking is at a high level, and we also investigated behaviors as proxies that surface as the result of higher-order thinking. We discussed these applications later in the discussion section. Nor are we attempting to address the lack of motivation to process online information on casual occasions. Fogg's Behavior Model suggests three necessary components for a behavior: sufficient motivation, sufficient ability, and an effective trigger [84]. We agree that limited motivation is one important barrier to starting higher-order thinking online, and the entertainment-focused lens is exacerbating this challenge. Although our cognitive mechanisms connect with motivation activation, this work concentrates on processing high-stakes information from the angle of cognitive trigger. StoryWell by Herman Saksono et al. [231] and Birk et al's study on intrinsic motivation in games [31] examined motivation articulation and instantiated solutions in particular contexts.

4.3.2 Design and Implementation of DeepThinkingMap

Design Principles in DeepThinkingMap

DeepThinkingMap is designed as an alternative documentation tool that allows people to share their notes and briefly interact with others. Meanwhile, it is also an example of conscious

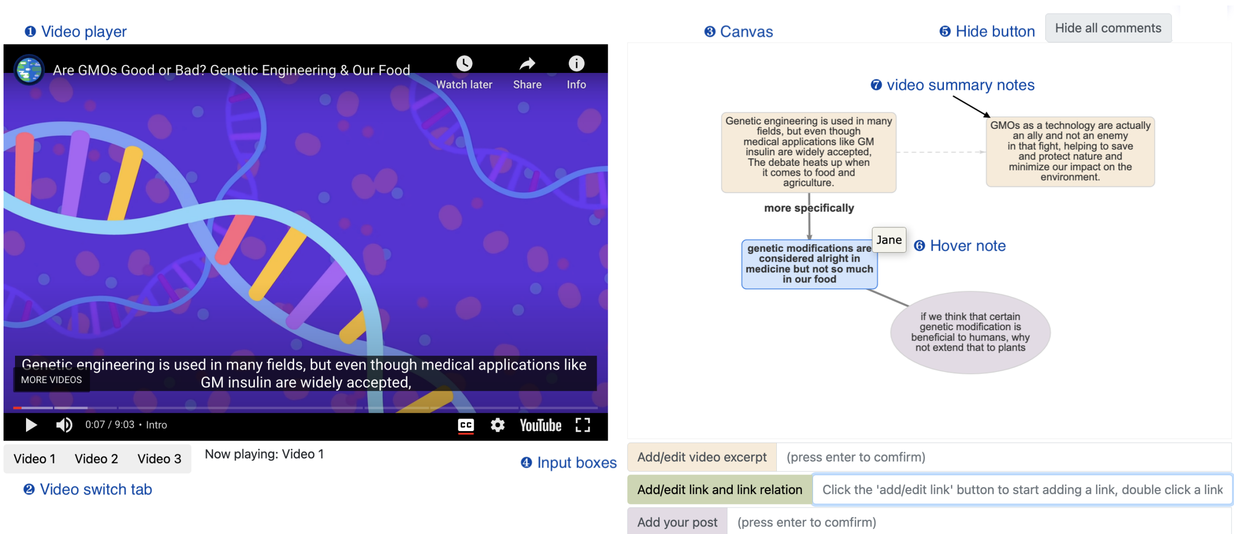


Figure 4.1: DeepThinkingMap Interface. It includes a standard video player with a video switch tab on the left, allowing for in-tab video switching. The note canvas for real-time note-taking and discussion is on the right. Below the canvas are input boxes for video summaries and individual comments. Additional features like the ”hide comments” button and hover-over identity indicators also facilitate collaboration.

thinking facilitators, to encourage users to exert greater cognitive effort and elevate their higher-order thinking engagement. As discussed, concept mapping, which is the core of DeepThinkingMap, offers substantial benefits in both visual representation and activity levels. By externalizing their thoughts on a canvas, concept mapping facilitates individuals to enhance the clarity and organization of internal ideas. Also, the graphical structure of documentation enables users to map out and customize the relationships between pieces of information, to foster deeper thinking processes, such as synthesizing evidence, and evaluating the assumptions of conclusions [150, 268]. From the aspect of visualization, the graphical display of nudging content may also help. The positioning of information is another common approach to tap into or mitigate the influence of status-quo bias, such as default option [28, 40]. Compared to linear documentation such as Word and comment threads, a concept-map-like structure of documentation may help mitigate the anchoring bias as users can read the nudging content freely in the canvas [150]. Thus, we designed and implemented DeepThinkingMap as a sharing space, to serve as a documentation tool and scaffold for higher-order thinking processes.

In terms of high-stakes resources people need to think, we use health-related videos as illustration in this work. Videos serve as a key medium for knowledge dissemination [223]. Prior studies found that conventional video players available on online platforms like YouTube find it difficult to afford active watching activities, such as comparing video content with other materials or social comments for content validation. To perform motivated higher-order thinking activities, users may need to interrupt or pause the video, memorize the content, switch their attention to another source of information, and then switch back for comparison and analysis, resulting in usability hurdles that can interfere with users' performance and motivation in higher-order thinking tasks.

Informed by the socio-cognitive mechanisms of social nudging we identified as well as basic considerations of usability, we identify three design principles (DP) that guide our designs:

DP1: Fostering social transparency. For utilizing peers' thinking activities to nudge higher-order thinking, it is crucial to ensure that both the presence of peers and their thinking activities (e.g., identifying an important idea from the video or adding a personal reflection) are salient and accessible to the users [27]. Our interface should explicitly highlight social presence either in real-time or asynchronously with saved notes contributed by peers. And it needs to help users compare their own thoughts with others.

DP2: Presenting temporal connections between personal notes and video content. Since users often anchor their comments to specific video segments, it's vital to allow seamless integration between these notes and the video content for two-way references (from video segments to notes, and from notes to segments) and inter-operatability [285]. Thus, the interface should anchor user-generated comments at specific time points of the video where these comments are generated. This time-anchored design will allow users to easily view and make sense of available contributions from self and peers, as well as add new contributions to the workspace without disrupting the visibility and accessibility of existing comments.

DP3: Differentiating video content and personal thoughts. Higher-order thinking builds upon the original content from the video. To ensure that higher-order thinking acts are visible, and not mixed with mere extraction or rephrasing of the video content, the interface should be designed to distinguish the two types of contributions.

Interface Design

DeepThinkingMap interface consists of the video player and a concept mapping canvas for social discussions (Figure 4.1). The video player panel features essential video controls, allowing individuals to control their own video screen. And video tabs enable users to switch between pre-selected videos.

The mapping canvas enables real-time collaborative note-taking and sharing of personal thoughts. Users may use two input boxes to contribute different types of notes: the ”Add/edit video excerpt” box for adding short summaries of specific video segments and the ”Add your post” box for adding individual comments for noting and sharing personal thoughts with peers and for possible group discussions. The ”Hide comments” button allows users to toggle the visibility of personal thoughts, enabling users to choose between seeing only the summaries of video content or seeing all contributions made by self and peers. The specific design aims to allow users developing a personal understanding about the video content without social inputs first, and then engage in exchanging personal thoughts with others afterward.

For clarity, contributions added to the mapping space will appear in specific shapes and colors to differentiate different types of content: box-shaped and light-yellow for video excerpts and violet ellipsoids for comments, with the hues resonating with their respective input boxes (**DP3**). To further help users see and contemplate the relations between different nodes (e.g., how one video segment relate to another, or what's the video context of a personal thought) and compare one contribution to another (**DP1**), directed links can be added between notes, users may specify the directions of these links and add text to explicate their beliefs about the relations as the labels to the links. Users can position the notes at will to avoid visual cluttering and to help them compare notes contributed by self and other people. When users hover over any note, the note is highlighted and the name of its last author will pop up to help collaborators know the author and increase their social presence. According to **DP2**, our system adeptly captures the video timestamp when notes are created. Double-clicking any note will navigate video to the associated timestamp, to help others understand the context of the notes from videos and compare them with their own thoughts.

The DeepThinkingMap prototype is implemented as a web application for synchronous and asynchronous social discussions around the videos. It utilizes the YouTube Data API for client-side video sourcing and D3.js for interactive functionalities. Real-time note synchronization is achieved through Firebase database.

4.4 Study 1: Social Nudging in Asynchronous Cooperation

4.4.1 Study Design and Hypotheses

In Study 1, we explore how visibility of others' thinking activities influences subsequent users' engagement with higher-order thinking activities of videos with DeepThinkingMap in a relatively controlled setting. We design an asynchronous cooperation scenario where we can isolate the visibility of other users' thoughts and understand the impact of static nudging cues. Therefore, we invited two batches of participants to finish a video consumption task in sequence. The first batch of participants watched a video with a blank DeepThinkingMap canvas and wrote down their notes (control condition); the second batch of participants finished the same task with the notes of one participant from the first batch as nudging (nudge condition). This became a between-subject study comparing the influences of the existence of others' thoughts (e.g., comments or notes) on engagement of thinking activities and note-taking results.

As the initial study, we focused on verifying the core assumption of social nudging – whether seeing other people's thinking activities can nudge oneself to contemplate. We hypothesized that accessing peer thoughts would positively increase the extent of understanding of the content, and engagement in higher-order thinking activities (i.e., critical and reflective thinking). By seeing peers' thoughts, the follow-up comments and notes on the videos in the nudge condition would be more thoughtful and constructive than the control condition, showing greater reflection toward video content.

4.4.2 Participants

We recruited 44 students (28 female, 14 male, 2 non-binary), aged 19-29, from one university campus, after screening to ensure that they can understand health information, make health decisions, and are fluent in English. Participants were enrolled in two phases: 24 initially for the control group and 20 in the nudge condition later. They reported a median level of self-assessed healthcare knowledge ($mean = 4.93$, $sd = 0.99$, on a 7-point scale where 1 = "novice level" and 7 = "expert level") based on questions from [122] (e.g., "I am familiar with preventing minor and chronic problems such as allergies and dry skin."). As background knowledge is fundamental to shaping the directions of reflective and critical thinking [72], we compared healthcare knowledge scores between conditions and found no significant difference ($t = -0.15, p = 0.87$), which confirmed no sampling differences between conditions. The duration of the study is 30 minutes, and they received about $7 USD for their time. It was approved by our institution's ethical review board.

4.4.3 Material Preparation

Video Materials

We selected two healthcare videos on different topics from YouTube. One introduces human immune systems ("Immune System") , explaining multiple lines of immune defense against bacteria. The other talks about the scientific evidence about the influence of turmeric in golden lattes ("Golden Latte"). Appendix B.1 provides further details on these videos. Both videos are geared toward daily health practices to foster participant engagement. We shortened the videos to approximately five minutes, to ensure enough time for thinking activities and notes. About 95% of participants were unfamiliar with their assigned video, while some already had background knowledge. Participants were randomly assigned to one of the two videos.

Nudge Content Construction

To prepare nudging content for the second batch of participants, we collected notes from participants in the first batch. These 24 participants were invited to perform the task alone (i.e., the control condition). Given a blank canvas on DeepThinkingMap, they were instructed to produce two types of video-related notes when they watched the video assigned: (1) a concept map to represent key takeaways from the video (denoted as concept map of video content later), and (2) any comments they have relevant to the topic or video content (e.g., opinions, related experience, and questions). Participants were instructed to link their video takeaways and comments with links and labels, demonstrating the connections they saw.

Since the 24 participants were randomly assigned to one of two videos, there were 12 individual notes for each video. To control the nudge quality in the nudging condition, we chose two individual notes of average quality from each video to use as nudge content. Quality was assessed by two researchers who independently identified unique concepts mentioned in each video — finding 28 for "Immune System" and 36 for "Golden Latte." We then calculated the conceptual coverage (i.e., *#concepts_mentioned* / *#concepts_ever_mentioned_by_anyone*) for each participant's concept map and chose those whose coverage ranked at the median. Although we selected the nudge content by the concept maps of videos, we displayed the complete notes these four participants wrote, including their comments or other notes. In summary, four participants whose concept maps and comments from the first batch were selected for the nudging content. These four participants and their data were excluded from subsequent analysis to prevent confounding effects.

4.4.4 Procedure

We conducted Study 1 via DeepThinkingMap website. Upon recruitment, participants received an invitation to complete tasks and questionnaires online on our website. They were first introduced to the study's procedure and given a tutorial on using DeepThinkingMap and concept mapping. Depending on the study condition, the DeepThinkingMap either displayed nudging content or remained empty. Participants were asked to produce two types of video-related notes on the canvas of DeepThinkingMap. The task instruction for the control condition is described in Section 4.4.3. For the nudge condition, the instruction was slightly different to match the availability of nudging content. Participants were told that they were extending a sequential task and were given a pre-existing concept map and comments to improve and add their own comments. We imposed a 20-minute limit for the task in both conditions. Upon completion, participants filled out a post-study survey on their engagement levels of critical and reflective thinking and understanding thinking from [130].

4.4.5 Measures and Data Analysis

Engagement in Higher-order and Lower-order Thinking

In the post-task survey, participants rated their engagement in reflective and critical thinking using a 5-point Likert scale (1-strongly disengaged, 5-strongly engaged). The survey items, adapted from Kember et al. [130], are provided in the appendix. We also asked about participants' engagement level of understanding as a comparison. Understanding is a relatively lower-level thinking process from Bloom's taxonomy [143]. Cronbach's alpha is 91.6%.

Behavioral Logs

To understand how participants actively watch the video and write notes during the process, we logged video player interactions and edits on concept mapping canvas. We tallied video pauses and rewinds as indicators of active video watching and quantified individual canvas operations—adding, editing, or deleting nodes and edges—to measure note-writing participation. These logs serve as side proxies for participants' activeness on higher-order thinking.

Comment Category Analysis

Like reflective journals [217], participants' comments reveal the depth of high-order thinking using code schemes, as another metric to understand the effect of social nudging. Two coders first identified the comments from participants' video-related notes by marking those notes not explicitly mentioned in the video. Then, the coders classified these comments into three cognitive levels of thinking based on Anderson and Krathwohl's revision of Bloom's taxonomy [12]: understanding, analysis, and evaluation. The categories of analysis and evaluation represent higher-order thinking. "Because the generation process of the comments of these two categories may need both critical and reflective thinking, we do not intend to separate the two or match them with thinking processes exactly." The coders reached a high level of agreement, with a Cohen's Kappa higher than 0.9.

4.4.6 Results

To evaluate the influence of social nudging on engagement in higher and lower-order thinking (i.e., critical and reflective thinking) and understanding, a lower-order thinking activity, we conducted a mixed-effect one-way analysis of variance (ANOVA). The main independent variable is the study condition (control vs. nudge). The video topic and the specific video-related notes used for nudge content were considered categorical random variables. Participants' self-reported score of healthcare knowledge was included as a confounding variable. For post-hoc analysis, we reported omega square (ω^2) for significant or near-significant effect sizes [290]. As the sample size of

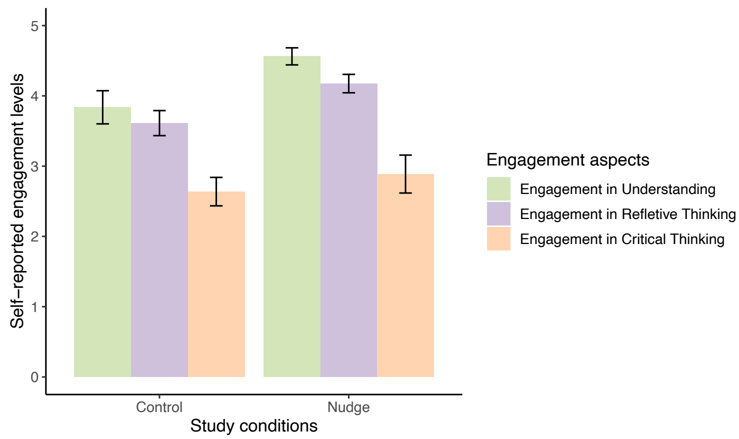


Figure 4.2: The self-reported engagement of understanding and higher-order thinking for both control and nudge conditions in Study 2. Error bars indicate the standard error.

each condition is relatively small, omega square is a more unbiased effect size statistic compared to eta squared (η^2) [6, 290].

Engagement in Higher-order Thinking

For engagement of reflective thinking, ANOVA results show a significant difference between the control condition and nudge condition in reflective thinking ($F(1, 37) = 6.59$, $p = 0.01$, $\omega^2 = 0.13$) shown in Figure 4.2. Moreover, ANOVA shows a trend toward significance in healthcare knowledge on reflective thinking engagement, but it is not statistically significant ($F(1, 37) = 3.94$, $p = 0.05$).

For engagement of critical thinking, no significant differences were observed between the two conditions ($F(1, 37) = 0.26$, $p = 0.70$). Similarly, healthcare knowledge shows no main effect on critical thinking engagement ($F(1, 37) = 2.27$, $p = 0.14$).

To further investigate whether the difference between reflective and critical thinking is evident at individual level, we performed a paired t-test to compare the two engagement scores for each participant. It shows that participants were significantly less engaged in critical thinking than in reflective thinking overall across both conditions ($t(39) = 3.76, p < 0.01$). This suggests that participants found it more difficult to engage in critical thinking, which is consistent with prior work [130].

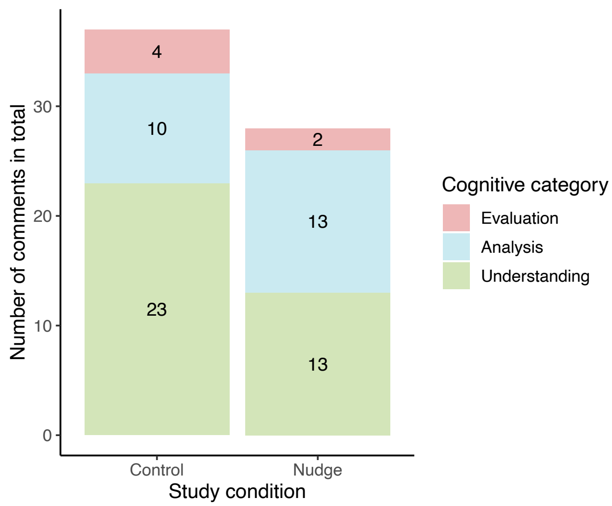


Figure 4.3: Cognitive levels of comments across two conditions

Engagement in Understanding

As reported in Figure 4.2, social nudging significantly increases engagement in understanding (nudge condition: $F(1, 37) = 7.75$, $p < 0.01$, $\omega^2 = 0.15$). Additionally, participants with a high score of healthcare knowledge were also statistically more engaged in understanding the video ($F(1, 37) = 3.94$, $p = 0.05$, $\omega^2 = 0.07$, $\beta = 0.26$).

Comment Category and Behavioral Logs

We extracted 37 comments in the control condition and 28 comments in the nudge condition. Figure 4.3 shows the cognitive types of comments in both conditions. Over 50% of comments in the nudge condition represented higher-order thinking (i.e., analysis and evaluation), compared to 37% in control condition; however, the absolute number of such comments was similar in both conditions: 14 in control and 15 in nudge.

From Table 4.1, the reduced frequency of video watching and canvas editing was found in the nudge condition. t-tests did not reveal any significant differences between the two conditions in video rewinds and pauses, suggesting that participants exhibited similar levels of video-watching activity with or without nudging cues. One potential reason might be the saturation of content summarization, given the first participant worked for 20 minutes to write a concept map of a 5-minute video.

Table 4.1: Mean and standard deviation for video and map operation counts in DeepThinkingMap across two conditions

Condition	Map editing	Video rewind	Video pause
Control	42.05 (18.39)	1.45 (1.47)	11.35 (10.83)
Nudge	18.75 (9.15)	1.85 (1.23)	17.85 (10.77)

In summary, in Study 1, our results partially supported our hypotheses. From the survey results, receiving social nudging improved engagement in understanding and reflective thinking processes of the subsequent collaborators, but had no significant impact on critical thinking. For the comments participants left, both conditions produced nearly the same number of higher-order thinking comments, which did not support the efficacy of social nudging.

4.5 Study 2: Social Nudging in Synchronous Cooperation

Study 1 explored the impact of social nudging cues during video consumption using the DeepThinkingMap tool. This provided insights into the role of social nudging on various thinking processes within a controlled asynchronous setting. Progressing from there, Study 2 aimed to deepen our exploration by situating social nudging within a more dynamic and realistic context of interactions. In this context, participants witnessed peers creating video-related notes in real time and have the chance to engage in synchronous discussions. Moreover, we compared two documentation approaches to afford discussions among participants: DeepThinkingMap and another traditional approach, which doesn't provide extra support. Through Study 2, our aim was to provide a richer, more nuanced understanding of social nudging and to evaluate the influence of social nudging with different technical mediation tools in a less controlled, real-world-like scenario for better ecological validity of the findings.

4.5.1 Study Design

In Study 2, we used a between-subject experiment design to understand how people are engaged in different thinking activities in synchronous interactions with two peers compared to thinking alone (i.e., with or without social nudging), as well as using different types of workspace to document and share thoughts about the videos (i.e., using DeepThinkingMap or not). To inves-

Table 4.2: Study conditions in Study 2

	Individual	Social
Google Docs	Individual video review using Google Docs (Individual-Docs)	Collaborative video review using Google Docs (Social-Docs)
DeepThinkingMap	Individual video review using DeepThinkingMap (Individual-DeepThinkingMap)	Collaborative video review using DeepThinkingMap (Social-DeepThinkingMap)

tigate the effects of the two factors on higher-order thinking activities, the experiment is designed as a 2×2 factorial design: individual (review the videos individually) vs. social (in a 3-person group collaboratively), and Google Docs (conventional documentation) vs. DeepThinkingMap (graphical documentation as cognitive facilitator). Table 4.2 outlines the conditions and their abbreviations. Participants either completed the experiment individually or in triads using designated documentation tools. Participants in the Social-DeepThinkingMap and Social-Docs conditions did not know each other beforehand. Moreover, to foster a more collaborative environment and streamline communication, a Discord channel is integrated with these two conditions.

4.5.2 Procedure

We used a similar task as in Study 1, i.e., health video consumption. We instructed this task as a video review to help participants utilize their video processing heuristics. Since we noticed the sign of contribution saturation in the original video mapping task in Study 1, which may prevent multiple collaborators from making contributions to the workspace, we increased the complexity of the task by inviting participants to compare information from different sources, which also demands higher-order thinking [90]. We therefore increased the number of videos available for participants to watch in the task to create opportunities for critical thinking. Before the study session started, we introduced the task and played a pre-recorded tutorial to help participants familiarize themselves with the task and tools. Then, participants had 60 minutes to review three videos about a topic randomly assigned to them, which is either about the characteristics of Genetically Modified Organism foods [GMO] or how vaccination works [vaccination] (detailed descriptions of videos are in section 4.5.3). Specifically, participants were asked to review these three videos on the same topic and discuss a thought-provoking question corresponding to the specific video topic, ”Do you think it’s safe to vaccinate eligible people?” for the vaccination topic or ”Do you think GMO food is safe?” for the GMO topic. They didn’t need to write a final report together; instead, they were instructed to write down their stand and reasons individually afterward.

Similar to Study 1, participants needed to (1) write down important points of the videos, and (2) read and share their personal thoughts and review opinions. For the participants in social conditions, they were encouraged to form group communications. We asked participants twice about their attitudes and confidence about the topic assigned, once before watching the video, and once after the video review. They also completed the same self-report survey on thinking activities as in Study 1. The study ended with a semi-structured interview with researchers for about 5–15 minutes. Study 2 took about 90 minutes and participants were compensated with a $15 gift card for their participation.

4.5.3 Material Preparation

To situate social nudging in a realistic context where misinformation and controversies tend to emerge, we used two health-related topics that frequently appear in public discourse: GMO foods and general vaccinations. Despite broad scientific consensus on these two topics, online debates and discussions about their safety and impact remain and public misunderstandings still pose threats to people’s wellbeing. Therefore, they serve as important subjects and relevant matters to participants, suitable for exploring higher-order thinking in the landscape of public opinions. We curated 5–8 popular videos from YouTube for each topic and invited two professionals holding PhD degrees in pharmacology or public health communication to assess the veracity of the videos. Ultimately, we selected three videos per topic that align with the current scientific consensus. By adopting the videos consistent with scientific agreement, we aimed to avoid the ethical dilemma of exposing participants to misinformation. However, it’s important to note that this task doesn’t intend to correct or reinforce users’ attitudes. The total duration of the three videos is about 30 minutes. Detailed video information is summarized in Appendix B.1.

To give users some note examples to start with their video mapping and comparison work, we employed extractive summarization algorithms to generate two key highlighting sentences for each video (e.g., as (7) in Figure 4.1, the note introduces a key position). We utilized a range of techniques, including TextRank, LexRank [71, 192], Latent Semantic Analysis (LSA) [85], and

BERT-based language models [225]. The top-rated two-sentence highlights for the videos as identified by the majority of these algorithms were chosen and presented in all study conditions.

4.5.4 Participants

We recruited 52 participants from our university campus, aged 18 to 45 (14 males, 35 females, 3 no-binary). Study conditions and video topics were randomly assigned. Participants assigned to social conditions were randomly grouped as three-person groups. The participants' level of education and background range from undergraduate students to Ph.D. holders. To ensure balanced prior knowledge across groups, we assessed participants' healthcare knowledge using a 7-point Likert scale [122], with a median level of 5.18. Given the potential that people with health or medical majors may possess superior knowledge and perceive connected with health-related topics, we checked and found that 1–2 people in each condition had health backgrounds, confirming even distribution. We may say that there was no group difference between participants' personal relevance with the topic. Their attitudes towards the video topics were also measured, yielding a mean score of 5.69 ($sd = 1.26$), where 1 stands for "strongly anti-[topic]" and 7 stands for "strongly pro-[topic]." Furthermore, as participants' education and their prior attitudes about the video topic largely influence their higher-order thinking skills and motivation [130], we ran a Kruskal-Wallis test to compare the means in different conditions since the values are not normally distributed. We found no significant differences across study conditions in education ($\chi^2(3) = 1.45$, $p = 0.69$) and prior attitudes about the video topic ($\chi^2(3) = 0.04$, $p = 0.99$).

4.5.5 Hypotheses

For Study 2, we proposed three hypotheses about how social nudging could affect participants' thinking engagement under different technical mediation tools, as well as thinking results from their notes and their attitude toward the video topic.

Social nudging can increase engagement in reflective thinking by accessing other peers' notes from the results of Study 1. In Study 2, participants will exchange their thoughts of video reviews more interactively and observe how others form their notes in real time. Therefore, our hypotheses are:

- ***H1a***: Using the same documentation tool, when participants are with social nudging, participants will have a higher level of higher-order thinking engagement than thinking alone.
- ***H1b***: For participants in the same social nudging availability (i.e., they all receive social nudging or they think alone), when participants document following the scaffold, they will have a higher level of higher-order thinking engagement than documenting on Google Docs.
- ***H1c***: With the synergistic effects of social nudging and thinking scaffold, participants with social nudging and DeepThinkingMap will demonstrate elevated levels of engagement in higher-order thinking compared to those who think alone with Google Docs.

Exposure to social norms may not only engage people in high-order thinking activities, but may also nudge them to change their attitudes and associated behavior as the thinking result, such as deciding to take vaccines [5, 226]. We hypothesize that:

- ***H2a***: Utilizing the same documentation tool, when participants are in the social nudging conditions with access to others' thoughts toward the topic, participants will be more likely to change their attitude to agree with the videos with higher confidence.
- ***H2b***: For participants with the same availability of social nudging, when participants are scaffolded with the documentation tool, they will be more likely to agree with the videos with higher confidence, compared to using Google Docs.
- ***H2c***: With the synergistic effects of social nudging and thinking scaffold, participants will develop more positive attitudes with greater certainty in group than those who think alone using Google Docs.

The notes participants documented during the video review and comparison activities may contain clues indicating their levels of higher-order thinking, as note-writing can help explicate and externalize the results of their thinking [131, 284]. We may use these writing results as the source to assess their thinking levels [284]. Thus, our last set of hypotheses are:

- ***H3a***: Utilizing the same documentation tool, when participants are in social conditions, participants will write notes that reveal more frequent higher-order thinking than thinking alone.

- ***H3b***: For participants with the same availability of social nudging, when participants document with scaffold, they will write more thoughtful notes, reflecting more frequent engagement in higher-order thinking, than writing on Google Docs.
- ***H3c***: With the synergistic effects of social nudging and thinking scaffold, participants will write more thoughtful notes after group video review, than those who think alone using Google Docs.

4.5.6 Data Collection and Analysis

We collected the engagement level of higher-order thinking, and also measured multiple higher-order thinking results, such as attitude and notes, to understand the cognitive levels and results of thinking activities. We interviewed participants as well to capture their thinking processes.

Engagement in Higher-order Thinking

We employed the same survey instrument as used in Study 1 (section 4.4.5) [130] for measuring participants' engagement in higher-order thinking during the study. To facilitate a more refined assessment of engagement in higher-order thinking – particularly in critical thinking – compared to Study 1, we reviewed the description to match it with the context of the current study design, and converted the survey to a 7-point Likert scale ranging from 1 ("Strongly disagreed") to 7 ("Strongly agreed"). Cronbach's alpha is 0.868.

Attitude toward Video Topic

As one important function of social nudging to encourage individuals to reflect over controversies, filter out irrelevant and incorrect information, and form decisions consistent with the correct information, we measured participants' attitudes toward the video topic, as well as their self-assessed confidence in their attitudes before and after the task. We developed single-question survey items to collect participants' attitudes about the video topic and their levels of confidence about those beliefs using a 7-point Likert scale. Survey questions asked include "To what extent do you agree that [topic] is currently safe?" and "How confident are you about your current opinion?".

Video-related Note

The video-related notes contributed by participants to the workspace (e.g., notes of video reviews) serve as repositories that capture people's externalized thoughts, which may help assess the depth of their thinking processes. First, we identified comments containing personal thoughts, which are notes with content that aren't directly overlapping with what's covered in the video, and applied the same coding method from Study 1 to categorize them into three cognitive levels [12], understanding (basic level thinking activity), analysis, and evaluation (higher-order thinking activity). We restricted the analysis to notes collected from conditions where DeepThinkingMap was used as notes are added and formatted in a consistent manner with the system.

Next, we conducted a language analysis to assess language use across distinct conditions on video-related notes. Prior research suggests that specific linguistic features can reveal insights into an individual's thoughts, emotions, and other psychological states [206, 213]. Hence, we utilized the Linguistic Inquiry and Word Count tool (LIWC 22) [34], a dictionary-based text analysis tool, to measure linguistic attributes including *analytical thinking* and *causation* words. In individual conditions, we scored participants' notes with LIWC. For social conditions, considering that participants often abstained from noting repeated ideas, as informed by the interview results, we analyzed the entire set of notes from one group as a single document and assigned a shared LIWC score to each member of the group.

Semi-structured Interview

To better grasp the role of social nudging in supporting participants' higher-order thinking and gather feedback on DeepThinkingMap, we conducted semi-structured interviews centered around participants' review process, attitude changes, collaboration experiences, and use of DeepThinkingMap. After transcribing the interviews, we utilized thematic analysis combining both open and axial coding [187]. In open coding, we pinpointed recurring codes tied to interview topics. These codes were later refined and merged into broader themes during axial coding, ensuring all themes were driven by consensus among the authors [187].

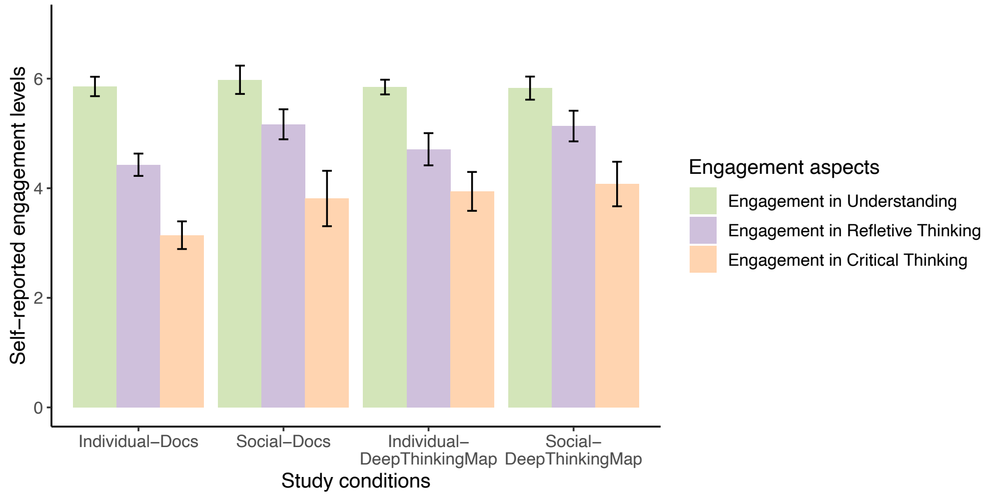


Figure 4.4: The self-reported engagement of understanding and higher-order thinking for all conditions in Study 2. Error bars indicate the standard error.

4.5.7 Results

Two-way mixed-effect ANOVAs were conducted to assess the influence of social nudging and DeepThinkingMap for different higher-order and lower-level thinking. Satterthwaite's method was used to approximate the degrees of freedom in their mixed-effects models, which may lead to non-integer degrees of freedom. The documentation tool (Google Docs vs. DeepThinkingMap) and the availability of social nudging (individual vs. social) are two independent variables. Video topic was included in the model as a random effect. The demographics, background, education, prior attitude, and self-reported healthcare knowledge were included as covariate variables. We also considered the contribution of members' educational backgrounds in the 3-person group and added the variance of education levels in each group as another covariate variable (the value for individual conditions was assigned as 0).

Engagement in Reflective Thinking

ANOVA results revealed significant differences in both the documentation tools and the availability of social nudging. As shown in Figure 4.4, individuals or groups using DeepThinkingMap engaged in significantly more reflective thinking than those using Google Docs ($F(1, 44) = 5.12$, $p = 0.03$, $\omega^2 = 0.08$). With the same documentation tool, there was a statistically significant effect ($F = (1, 44) = 15.21$, $p < 0.01$, $\omega^2 = 0.24$) such that participants in social conditions engaged more deeply in reflective thinking activities compared to solo participants.

Furthermore, a post-hoc Tukey comparison provided clarity on the differences between conditions. Regardless of documentation tools used, people with social nudging in social conditions all practiced more reflective thinking than those who thought alone (Social-Docs - Individual-Docs: $q = 3.63$, $p < 0.01$; Social-DeepThinkingMap - Individual-DeepThinkingMap: $q = 3.63$, $p < 0.01$). And we found that DeepThinkingMap further amplified this effect (Social-DeepThinkingMap - Individual-Docs: $q = 3.98$, $p < 0.01$).

Engagement in Critical Thinking

For the engagement of critical thinking, we found similar statistical results from ANOVA and post-hoc analysis. Social nudging significantly improved the engagement in critical thinking engagement ($F(1, 43.75) = 6.59$, $p = 0.01$, $\omega^2 = 0.11$), with participants in social conditions delving more profoundly into critical thinking engagement tasks compared to individual participants. Meanwhile, compared to participants using Google Docs, those who utilized DeepThinkingMap to generate and share thoughts exhibited significantly greater critical thinking engagement ($F(1, 43) = 6.60$, $p = 0.01$, $\omega^2 = 0.11$). Prior attitude plays a statistically significant role in the level of critical thinking ($\beta = -0.56$, $F(1, 38) = 21.28$, $p < 0.01$, $\omega^2 = 0.33$). Where participants hold more positive attitudes toward the topic, they engaged less in critical thinking since their existing attitude aligned with the videos.

Following ANOVA analysis, Tukey post-hoc comparisons found that critical thinking engagement is only significantly improved in Social-DeepThinkingMap as opposed to Individual-Docs ($q = 3.33$, $p < 0.01$), as shown in Figure 4.4. The joint impact of social nudging and DeepThinkingMap supports the efforts in critical thinking.

Engagement in Understanding

As shown in Figure 4.4, neither social nudging nor documentation tools exhibited a significant influence on the engagement of understanding. The documentation tool did not significantly impact engagement ($F(1, 44) = 0.32$, $p = 0.57$). And the effect of social nudging was marginally significant ($F(1, 44) = 3.48$, $p = 0.07$, $\omega^2 = 0.05$). Tukey comparisons further confirmed the lack of significant difference between any two conditions (all $p > 0.3$). The average engagement scores hovered around 6 out of 7 across all conditions, suggesting a consistently high level of understanding engagement, potentially nearing a saturation level for participants.

Video-related Notes

For the content analysis of video notes, we report two aspects: overall linguistic characteristics and cognitive levels of comments. Figure 4.5 demonstrates the number of comments and their cognitive levels across DeepThinkingMap-related conditions. One-way ANOVA model with one independent variable (social nudging) shows that triad participants created significantly more analysis comments ($F(1, 18.87) = 15.40$, $p < 0.01$, $\omega^2 = 0.41$) than individual participants who work in DeepThinkingMap alone. We also found there was a significant positive impact on the count of evaluation comments ($F(1, 19) = 15.02$, $p < 0.01$, $\omega^2 = 0.40$) when participants received social nudging in synchronous interactions.And ANOVA did not show evidence that there were significantly more understanding comments in the social conditions compared to the individual conditions ($F(1, 18.96) = 0.05$, $p = 0.83$). The results of comments in Study 2 imply new evidence about the positive impact on the engagement of higher-order thinking from the graph-based documentation prototype with social nudging cues.

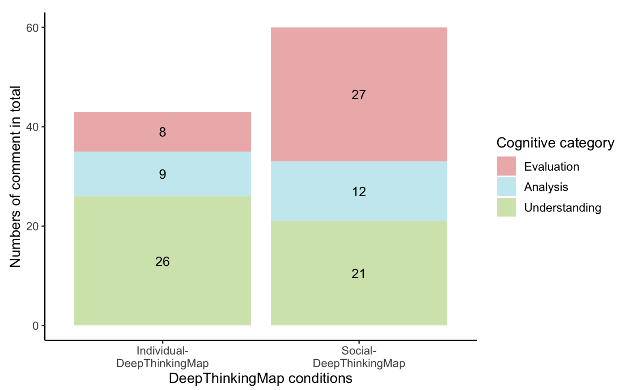


Figure 4.5: Cognitive levels of comments in DeepThinkingMap conditions of Study 2

For the linguistic results of all notes, including the summary of videos, we connect our findings with prior literature about writing and thinking levels. Analytics thinking score in LIWC, which is derived based on several categories of function words, captures the degree to which people use words that suggest logical and hierarchical thinking patterns [260]. Evidence shows that higher analytics scores positively correlated with good academic grades and application of reasoning skills [213]. Low scores imply that individuals write more personally from their word selection [213]. Table 4.3 shows that notes from social conditions had significantly higher analytics thinking scores than those from individual conditions ($F(1,44) = 5.16$, $p = 0.03$, $\omega^2 = 0.08$). Meanwhile, notes written using DeepThinkingMap also show marginally significant odds of higher analytics scores than notes using Google Doc ($F(1,43.16) = 3.76$, $p = 0.06$, $\omega^2 = 0.06$). Within words related to cognitive processes, we went through a full analysis and found that the frequency of cause keywords (e.g., "how," "because," and "make") has significant differences across conditions. As shown in Table 4.3, participants using DeepThinkingMap created more causal words in their video-related notes ($F(1,43) = 8.71$, $p < 0.01$, $\omega^2 = 0.15$), which may imply more critical thinking confirmed from prior studies [206]. However, social nudging didn't show a statistically significant impact on using casual words ($F(1,43) = 0.27$, $p = 0.61$).

Table 4.3: The means and standard deviations of the analytics thinking score and counts of causal words in video-related notes across conditions. Standard deviations are given in parentheses.

Condition	Analytics	Causes
Individual-Docs	72.26 (3.30)	3.84 (0.33)
Social-Docs	77.43 (2.56)	3.45 (0.41)
Individual-DeepThinkingMap	78.29 (3.07)	4.36 (0.47)
Social-DeepThinkingMap	81.36 (1.96)	4.46 (0.35)

Attitude Change

We are interested in how participants' attitude change in different conditions. Note that the change of attitudes is a proxy measure to understand the results of higher-order thinking in this work. We refined our description regarding attitude polarization, transitioning from a general pro-[topic] or anti-[topic] stance to a more specific pro-[video] or anti-[video] perspective, as the videos are all pro-[topic], in order to mitigate the potential confusion for readers about topic direction. Since our attitude measure is based on two variables, attitudes and self-assessed confidence in their attitudes, we conducted two separate mixed-effect ANOVA models. As a preliminary test, the correlation between attitude and confidence level is median ($r = 0.56$), implying that they are not strongly connected. For attitude, watching the videos significantly moved participants' attitudes toward aligning with the videos ($F(1,95) = 21.03$, $p < 0.01$, $\omega^2 = 0.17$). However, results revealed that neither social nudging ($F(1,95.37) = 0.66$, $p = 0.41$) nor DeepThinkingMap ($F(1,95) = 0.07$, $p = 0.79$) significantly impacted participants' attitude. For attitude confidence, watching the videos significantly strengthened participants' confidence in their attitudes ($F(1,95) = 14.27$, $p < 0.01$, $\omega^2 = 0.12$). And ANOVA confirmed the significant effects of both DeepThinkingMap and social nudging on participants' confidence about their attitudes. DeepThinkingMap significantly strengthened attitude confidence ($F(1,95.01) = 5.04$, $p = 0.03$, $\omega^2 = 0.04$). Similarly, social nudging had a significant impact ($F(1,95.59) = 8.91$, $p < 0.01$, $\omega^2 = 0.08$). From post-hoc Tukey tests, social nudging consistently improved attitude confidence significantly across documentation tools, compared to thinking alone (Social-Docs vs. Individual-Docs: $q = 2.96$, $p = 0.02$; Social-DeepThinkingMap vs. Individual-DeepThinkingMap: $q = 2.96$, $p = 0.02$). Moreover, we observed that social nudging with DeepThinkingMap was significantly more effective in strengthening confidence than solitary work with Google Docs ($q = 3.50$, $p < 0.01$). Figure 4.6 and 4.7 illustrate the change in attitudes and their confidence over the task.

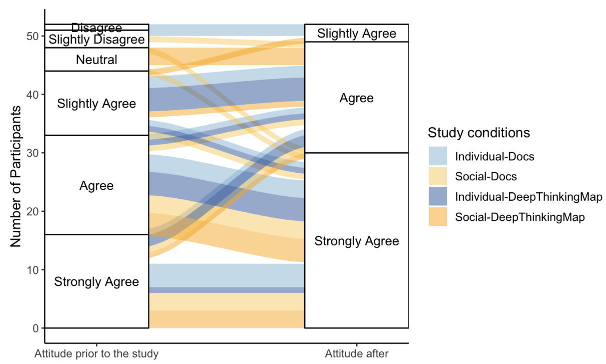


Figure 4.6: Pre- and post-study attitude shift on a 7-point Likert scale.

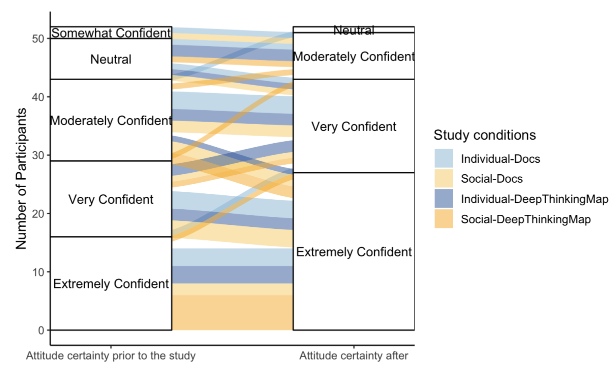


Figure 4.7: Pre- and post-study attitude confidence shift

Interview Results

Influences of Social Nudge Participants in social conditions appreciated the accessibility of their peers' video notes. They indicated that the diversity of viewpoints and focal points in their peers' notes contributed to a multifaceted understanding of the video content. Furthermore, participants engaged with the Social-DeepThinkingMap condition disclosed an active practice of integrating their notes with those of others within the concept map. This behavior was identified as a collaborative strategy aimed at synthesizing a more comprehensive overview.

"I think that we are touching upon different aspects sometimes. Some teammates that read the clips first, and then put them and some of them their notes was more like last year and had more details than mine." - P406

"I actually use the line connections to draw lines between the notes, because I just thought it was a very helpful tool. And I think drawing that connection between like, different points helps to draw my eye to like, figure out like, oh, this point leads into this, or this point leads into this. [...] if I saw a collaborators' note, and I thought my note was important to that as well or like lead into that I would draw a connection between that." - P404

Throughout the collaborative review, participants reported an increased confidence regarding their perspectives on the video topics. Engaging with peers not only acted as a validation mechanism by reinforcing their viewpoints, but also fostered an environment wherein participants felt compelled to actively and critically engage, rather than merely fulfilling study requirements.

"I think reading the opinions of my teammates definitely influenced a little bit, but it seemed to go hand in hand with the ideas that I had as well. So it kinda of confirmed in your brain that you're on the right track. Your opinion about this is agreed upon with other people" - P403

"I think just watching the videos, and taking notes with them (group members) in a more serious sense, rather than just like reading an article in passing. So having like a designated time to actually focus on it helps me like pay attention and think" - P212

Moreover, they reported that this collaborative approach aided in bridging existing gaps in their knowledge. In Study 2, the real-time social interaction further enriched the experience. Participants reported a heightened sense of social presence when collaborating on shared document platforms and communicating through Discord channels. This increased perception of social presence acted as a catalyst, encouraging participants to more freely share their thoughts and contribute video notes.

"Some information that my teammates felt was important like that helps to fill in some of the gaps about, like what was the context of this? Or what does this mean? Or why might this be important?" - P210

"Because it was more of like a group project. I felt like sort of the social pressure to contribute and do my part so that I wouldn't burden my teammates versus when I do it individually, I would procrastinate and not care about like the time limit as much" – P207

Conversely, the study revealed a strategic alteration in participants' behavior when confronted with repetitious notes or actions executed by peers within the same team. In such instances, they chose to avoid redundancy and contribute uniquely by taking alternative roles such as note organization, summarization, or establishing links between notes.

"I didn't do much notes, because my teammates were much faster than me. So what I did instead was I basically edited their points to add more, more detail that I felt like the point was missing some information." – P208

Our findings illustrate the nuanced effects of social nudging on participants' individual and collaborative behaviors in the context of video reviewing. On the positive side, collaborative note-taking serves to create a more comprehensive understanding of the video topic, while also increasing participants' attitude confidence, and sense of social presence. However, there are negative ramifications as well: social nudging could sometimes lead to behavioral shifts, such as the deletion or avoidance of note-taking when participants observe similar contributions from peers. This phenomenon aligns with previous literature on social blocking, underscoring the double-edged nature of social influences in collaborative settings [61].

Feedback for DeepThinkingMap Participants identified multiple advantages conferred by DeepThinkingMap. For example, they reported that the capacity for linking notes using DeepThinkingMap can visually represent interrelated concepts. This feature was observed to catalyze higher-order thinking and facilitate a more organized approach to addressing questions and obtaining a comprehensive view of the video, in contrast to conventional documentation tools, which typically compel users to relate ideas linearly.

"It definitely helps me to see how people think that things are related together. Also, it helps me to watch critically, while I check how I can connect my posts to the video." – P406

"It was really helpful because you can visualize things better. And the fact that we could connect concepts to other concepts made it easier to follow connections within concepts. And overall, help me answer questions, like better compared to like, regular note taking where sometimes like, it's just all messed up, and you don't know where one thing is. It's just hard to find information where you want to find it." – P304

Despite the positive feedback, participants suggested improving DeepThinkingMap's user interface and interactivity. Common recommendations included the incorporation of hotkeys for specific functions and the option to personalize color schemes for the concept map's components. Collectively, the interview results underscore the efficacy of DeepThinkingMap in video reviewing tasks, particularly its feature allowing for the meaningful interconnection of ideas.

4.6 Discussion

For high-stakes online content, higher-order thinking is required but needs external support. We propose leveraging social behavioral transparency and knowledge as nudge to promote higher-order thinking processes. And we compared the impact of social nudging under different documentation tools with or without thinking support. Ultimately, we found that social nudging augmented general users' critical and reflective thinking engagement and led to thoughtful notes and attitude changes. These results are summarized in Table 4.4. In the subsequent discussion, we discuss how social nudging shapes participants' cognitive levels in different social interaction scenarios. We also elaborate on the opportunities of social nudging and thinking scaffolding, from the framework of the dual-process theory. Furthermore, we explore the design space of cues for nudging users toward deeper engagement in higher-order thinking.

4.6.1 Social Nudge through Cognitive Heuristics

Aligning with previous research on reflection nudging [219], our work identifies the potential benefits of applying social nudging to empower general users in the context of high-stakes information processing. Our results provide insights into how sharing peers' behaviors and thoughts encourages individuals to think critically and reflectively, produce thoughtful notes, and change their attitude confidence.

Table 4.4: Summary of Findings in Study 1 and 2

Impact of	On	Results
Social nudging	Engagement in higher-order thinking	With static social nudge cues, reflective thinking engagement is improved, while critical thinking is not. With synchronous social nudges, engagement in both higher-order thinking activities is higher. (H1a supported)
	Attitude toward the topic	Social nudging strengthens attitude confidence, while no significant difference is found for attitude positivity (H2a partially supported)
	Written thinking results	With static social nudge cues, there was no difference in thoughtfulness of comments. With synchronous social nudges, there are more comments with higher-order thinking levels; and group notes are more analytical. (H3a partially supported)
Thinking scaffold, e.g., DeepThinkingMap	Engagement in Higher-order Thinking	Engagement of both higher-order thinking activities is higher with DeepThinkingMap. (H1b supported)
	Attitude toward the topic	Attitude confidence was stronger with DeepThinkingMap, while no difference is found for attitude positivity. (H2b partially supported)
	written thinking results	With synchronous social nudging, notes include more causal words, but are not more analytical. (H3b partially supported)
Interaction effect	Engagement in Higher-order Thinking	Engagement of both higher-order thinking activities is significantly higher with both social nudging and cognitive scaffold. (H1c supported)
	Attitude toward the topic	Attitude confidence was stronger with both social nudging and cognitive scaffold. (H2c partially supported)
	written thinking results	H3c not supported

Disclosing peers' thoughts can nudge people to engage in more reflective thinking processes in both asynchronous and synchronous scenarios; critical thinking when interacting with peers synchronously in a triad. The confidence in participants' attitudes was also strengthened with social nudging. Interview results suggested possible confirmation bias in social nudging, in which peers' thoughts could have socially validated participants' personal opinions. They tend to be more confident about their attitudes confirmed by social validation. In both studies, we controlled the quality of nudging content, such as selecting average-quality content as the nudge in Study 1 and controlling peers' level of education statistically to examine the behavioral impact associated with thoughts they produced. The interview results further confirmed that exposing them to peers' thoughts could encourage them to overcome their evaluation apprehension (i.e., fear of sharing one's own thoughts publicly as they may be subject to evaluation) and actively express their thoughts. The findings reiterate the values of utilizing crowd intelligence and civic engagement, even for complex tasks that may typically require domain expertise.

Consistent with the effort of engaging in high-order thinking, for the thinking result, we also found that participants tended to produce more analytical notes with social nudging. However, the transition from thinking engagement to writing thoughtfully is not always successful. While social nudging had a significant influence on critical and reflective thinking, people might not write down more thoughtful notes during video processing according to both linguistic aspects and manual coding evaluation. One potential explanation is that users may lack the necessary support to capture the nuances of their thoughts and separate thoughts from emotional reactions [107].

Although our experiments didn't request teamwork directly, we still observed contribution redundancy and some team role adjustments, illustrating the complexities inherent in social interactions and cooperations. In Study 1, we observed that the notes of average quality include about 30 video keywords, which may discourage subsequent participants from generating new notes and cause fewer map interactions. In Study 2, slower-paced participants further explained that they instead monitored others' work and made editorial modifications when they identified gaps or redundancies. Thus, our findings not only validate the potential of social nudging but also underscore the need for nuanced mechanisms to address the emergent complexities to preserve individuals' thinking opportunities and independence.

When comparing the two studies, we conjecture that two factors that we haven't fully investigated may be important. First, simultaneous interactions in Study 2 may provide a better sense of social presence so that individuals can clearly perceive others' real-time engagement on the video content and details. Also, interactions in Study 2 enabled immediate feedback from others and can accelerate the influence of social nudging. Participants may adjust their thoughts and activities more quickly and more frequently.

Contrasting our focus and results of social nudging with other social support ideas, we summarize the connection in ideas, and link to findings in this work. Social facilitation, in psychology, mainly focuses on the influences based on the direct presence or awareness of others, including the spotlight effect and co-author effect. Study 2 (with synchronous interactions) and many social nudging designs [40] utilize these effects to influence and guide user behavior indirectly. In Study 1, we stimulated an asynchronous online cooperation scenario where participants were not told that they were seen by others and participants in the nudging condition didn't have active interaction with others. The positive results in both Study 1 and 2 indicate the flexibility of cognitive heuristics nudging may adopt, which makes it possible to promote thinking under different interaction scenarios. As social nudging influences individuals through social factors, however, it can still have negative effects as in other collaborative cases. For instance, individuals may conform to the group majority [114], lead to groupthink, or cause social loafing harming group performance. For the attitude change, we did not further report the results of the participants whose attitudes didn't align with the videos in quantitative results, because the data size was not sufficient to draw any findings. Our interviews still gave us some evidence that individuals stayed open-minded about the topic. As discussed above, we also see the contribution redundancy from the aspect of the group performance.

4.6.2 Potential of Social Nudging Coupled with Thinking Scaffolds: Nudge Plus

Consistent with previous findings for concept mapping [39], we found the graphical structure of documentation method engaged people in not only better understanding the video contents, but deeply reflecting and thinking critically. It corroborates the benefits of concept mapping to bolster collaborative higher-order thinking in the context of high-stakes information processing. Our post hoc analysis revealed a synergistic effect when social nudging was incorporated into

DeepThinkingMap, a thinking facilitation intervention example. Especially in the engagement of critical thinking, while we didn't observe group-level differences from the post-hoc comparison for social nudging, the joint effect of social nudging and cognitive scaffold resulted in an amplification of effects when compared to individual thinking with Google Docs.

Looking into the nudging theory in recent behavioral public policy work, the boosting effect of social nudging and scaffolding can find its theoretical underpinning in the dual-process theory [73] and resonates with a new line of nudging strategy – *Nudge Plus*. Nudging is designed to work on fast and automatic type 1 processes that utilize the biases, leaving system 2 unengaged. Studies within HCI domains have expanded the nudging to more conscious nudging designs [40], while authors generally still discussed them within the traditional nudging scope. Concept mapping as cognitive scaffolding, on the other hand, encompasses system 2 thinking, prompting users to slow down and engage in a more structured, deliberate form of thinking. When we combine nudging with deliberative thinking scaffolding, we invite both thinking processes into activation. Behavior scientists recently coined this as "nudge plus", referring to "interventions that have a reflective strategy embedded into the design of a nudge." [18, 127] There has emerged more cognitive psychology and neuroscience evidence about how the two systems process together, such as parallel-competitive and default-interventionist conflict resolution strategies [18, 73, 244]. Following more findings from cognitive theories and behavior science, the nudge plus can be a promising avenue to design mixed tools with reflexive and conscious thinking support for video review and other reflection-needed tasks.

4.6.3 Design Implications for Higher-order Thinking Nudge

Various tools have been proposed to add and integrate information and increase the interaction between general users with other objects, to lower higher-order thinking threshold and support various tasks [27]. We utilize the conceptual understanding and notes from non-expert users as nudging content and identify the potential mechanisms of social nudging from transparent social behaviors and thoughts. This opens more possible roles socially generated and shared content may serve. As shown in the video review task in this work, we may design deliberately incomplete thoughts as nudging cues, to encourage users to critically engage with the information encountered, do their own research, and draw reflective conclusions. This approach aligns with similar ideas on scaffolding concept mapping itself using parts of expert templates, which suggests that promoting users to generate high-quality map with kit-build map can activate thinking [204].

However, our findings also reveal the potential pitfall. Unverified peer views may become sources of misinformation, especially when amplified by influencers or privileged voices. This dilemma underscores the potential limitations of facilitating the prominent display of targeted user-generated content for nudging and other purposes. To navigate this double-edged sword, researchers may carefully design thinking interventions that stimulate both intuitive (System 1) and reflective (System 2) thinking processes.

Another promising opportunities as quality of nudge content may not be a priority lies in the incorporation of AI-generated content on reflection support. LLMs have made it possible to generate human-like relevant content to given topics and understand human language to a certain extent [37]. Thus, when quality is not required, tailored AI-generated content could possibly be used for higher-order thinking support with caution, potentially in real-time and adaptive to users' cognitive status, with appropriate disclosure of AI generation and ethical regulation. As people may not always trust AI, we may also leverage people's hesitation toward AI technology to slow down their information processing process and invest in extra higher-order thinking efforts [195].

Social nudging and nudge plus are general tooltiks for encouraging users to invest more cognitive efforts in higher-order thinking for any specific high-stakes task. In our experiments, we applied it in a video review task which relates to fundamental health knowledge, healthy food purchase, health decisions, etc. Within health decision making domain, there are other applications, such as pursuing healthy lifestyles [174] and adherence to treatment plans. Out of health domain, social nudging can be integrated in the domains, including productivity and educations. For instance, the team awareness and the appreciation could be reminded by social nudging, which may lead to individual and group productivity [87]. And they could be looped with other scaffolding techniques, framing and timing designs, etc, to amplify the influence. When it comes to the implementation of nudge plus in the real world, there are multiple design details that need to be taken into consideration, some of them we controlled or discussed in the experiments. Nudging may tailor with the levels and directions of users' intrinsic and extrinsic motivations (e.g., internal goals and social rewards), besides the trigger and users' abilities. Moreover, nuanced measurement in the process of thinking task may be designed to capture subtle changes in user behavior and cognition. Similar to the distinction between critical thinking processes and thinking notes in our results, a detailed analysis may help researchers understand the underlying processes and the potential gaps in between.

4.7 Limitations and Future Work

Our experiments have several limitations. First, we used four specific health-related topics, rather than surveying what participants may be interested in beforehand. We tried to mitigate it in the study setup by choosing topics from prevalent social debates like GMO [184], and choosing the popular videos on YouTube, so that the selected videos simulate videos that participants typically encounter in everyday life. We did statistical tests to make sure the personal relevance of the topic is not statistically different; we used participants' education level to mitigate the potential confound of personal relevance to a extent.

We ensured that all the videos shown are credible and consistent with current scientific consensus, due to ethical concerns. This can lead to a lack of observations of how participants may react and discuss misinformation in the videos. Moreover, our participant sample may not represent diverse demographics, health backgrounds, or attitudes globally. When we recruited participants, the contentious nature of topics like vaccination deterred some from participating. It led to a skewed distribution of prior attitudes in participants, so the findings didn't capture much interaction and thinking results between a triad of participants with different group opinion dynamics. Although attitude change is not the sole focus of social nudging here, further studies could benefit from considering these factors when working on persuasion and nudging for tasks such as misinformation detection and correction.

Furthermore, our study primarily examined the effects of social nudging in contexts where group members were unfamiliar with each other. Prior research, however, has suggested that the source of information plays a pivotal role in determining its perceived credibility. For instance, information originating from a reliable friend is often deemed more trustworthy than that from a stranger [94]. As a direction for future research, we aim to incorporate both credible sources and misinformation. We also plan to engage participants from varied backgrounds and relationships to comprehensively discern the nuances of social nudges across different demographic groups.

Given the limited study duration and lab setting, it is difficult to predict the long-term effect of social nudging. We would like to further address these limitations in the future by implementing DeepThinkingMap as a website plugin for daily usage. We plan to conduct longitudinal studies where users may use DeepThinkingMap to watch videos of diverse topics and mixed quality and gauge the long-term effects of our designs and social nudging for general users. Another limitation, due to our lab setting, is that we pertain the user motivation in the experiments. We didn't design the levels of motivation among participants nor did we measure them as participants were instructed in the task. In our longitudinal study, we will also consider the motivational states users have when receiving the social nudging or cognitive scaffold to have a more nuanced analysis.

Chapter 5

Highlighting Non-verbal Gestures for Knowledge Co-creation

For knowledge co-creation, such as group brainstorming, communication between collaborators is essential and the information shared via communication is not limited to verbal content. When people communicate verbally, they may also co-produce hand movements as part of the communication process, but the potential of nonverbal behavior is easily overlooked. Going beyond using external visualization support for concepts and smooth interactions, we examine prioritizing users' in-conversation gesture use as an intrinsic mechanism to bolster group creativity. In this chapter, we aim to gain a deeper understanding of the usage and function of hand gestures in computer-mediated group brainstorming. Through a secondary analysis of laboratory study data, we verified that metaphoric gestures, or producing spatial cues with hands to convey intended concepts, can best influence self's and partner's idea generation. Also, the positive effect of metaphoric gesture is independent of media richness of communication medium (e.g., whether there's visibility or not).

5.1 Introduction

Understanding and supporting the interaction processes of group creativity continue to be a topic of importance in the Group and HCI communities, especially when considering how to support design activities that demand novel and useful design ideas [241]. For example, there has been much work exploring the design space of creativity support tools by employing multimodal interfaces and information representations to assist the sharing, exchange, and integration of in-

This chapter is modified from the following paper: Liao, J., & Wang, H. C. (2019). Gestures as intrinsic creativity support: Understanding the usage and function of hand gestures in computer-mediated group brainstorming. Proceedings of the ACM on Human-Computer Interaction, 3(GROUP), 1-16. [159]

formation in groups for purposes of creative design. Group brainstorming as a common practice to generate ideas through collaborative teamwork is in particular of relevance and interest in this regard [161, 274].

Prior work has provided theoretical frameworks and empirical understandings about the mechanisms and characteristics of group brainstorming. Group brainstorming can be modeled as a socio-cognitive process that involves both social exchange of information among team members, as well as cognitive operations on information received by individuals (e.g., ideas shared by other team members during a brainstorm) [62]. Research also showed that the social process of information exchange, as well as the cognitive process of creative perception and thinking, are interdependent. Effective brainstorming relies on using the ideas shared through team communication as stimuli for creative thinking [211]. Facilitating a smooth exchange of concepts and ideas in brainstorming groups is an important aspect in group brainstorming. Specifically, the styles of communication and collaboration can have significant effects on creativity tasks [224, 247]. The joint development and selection of ideas in group brainstorming would also require social interactions among group members inherently.

To improve communication and idea-sharing in brainstorming groups, previous work focused on extending group discussion with external support. One line of work investigated using shared large displays to achieve unhampered social interactions during brainstorming activities [38, 101]. As the physical properties and functions of such interactive displays allow co-located workers to more easily see and make use of ideas shared by peers, the approach enables interface designs for coordinating multiple participants' attentions and collaborative behaviors. Another line of work explored using visual representations of concepts, such as photos and drawings, to elevate the stimulating utility of shared ideas in creative thinking [105, 239, 274]. In Wang et al. [274], ideas shared verbally by workers were used as queries to retrieve relevant pictures from an image database to represent the ideas. These language-retrieved pictures serve as multimodal representations that help to expand the space of concepts accessible to workers. For the purpose of concept expansion, affinity diagram and similar representations highlighting inter-concept connections were introduced to structure collective ideas and links between them [280].

In summary, most of the previous work focuses exclusively on external support. The supportive mechanisms, including social displays and visual representations of verbal content, are extraneous to individuals, requiring workers to use specific tools (e.g., [239, 274]) or work at specific locations (e.g., [91, 101, 105]) to improve brainstorming outputs. Existing brainstorming tools are mostly designed to expose individuals to diverse stimuli provided by other group members and to organize ideas that are fragmentary and unstructured [105]. However, when team members don't have access to these specialized tools, brainstorming falls back to an unaided mode, raising needs to investigate supportive mechanism relatively intrinsic to individuals.

We note that the use of gesture in conversation is *intrinsic* to individuals, typically requiring no specific external tool to perform, which has the potential to serve as the basis of intrinsic support for brainstorming. However, little is known regarding how nonverbal hand gestures co-produced by people during conversations may affect group brainstorming. In this chapter, we aim to understand how worker's nonverbal behaviors, such as hand gestures, related to group brainstorming when collaborating in face-to-face as well as in computer-mediated settings. Specifically, rather than targeting the general process of hand gesture use in generic communication, we instead focus on the self-supporting properties of gestures for idea generation and creativity.

We conducted a secondary analysis of a laboratory experiment focused on group brainstorming with different communication mediums (i.e., properties of the communication tool used to communicate) from Lin et al [164]. The dataset collects verbal and non-verbal behaviors of participants in multiple brainstorming sessions, with the communication medium, language fluency, and brainstorming topics. It only includes coded spontaneous gestures from hand movement data, including iconic gestures, metaphoric gestures, and pointing gestures, proposed previously by the literature [95, 139, 189]. We re-examined the relationship between hand gestures and movements, communication medium, and creativity performance, using statistical analyses.

The results revealed that not all hand movements play the same role in group brainstorming. We found that metaphoric gestures had the most positive correlation with participants' idea-generation outputs, regardless of which communication medium was used. Medium, on the other hand, did not impact individual's creativity production. We also verified that the visibility of partners helped participants producing more hand movements and semantic-rich gestures. The results provided a proof-of-concept, suggesting that nonverbal gestures have the potential to work as an intrinsic mechanism for supporting group brainstorming, and this mechanism is mostly independent of communication media. Based on the results, we recommend researchers and designers to further assess various gesture-based solutions, including skill development programs for training productive gesture use or visualization designs for visualizing nonverbal behaviors to amplify individual's and group's creative potential.

5.2 Background and Related Work

5.2.1 Communication Medium and Creativity

Creativity is an essential part of everyday life and work and has been commonly recognized as the main vehicle of a project, an organization, and even the whole society. Scholars spanning across different fields, including sciences and engineering, have been studying creativity in diverse contexts and attempting to support it over the centuries [241]. One common definition of creativity focuses on the production aspect, specifically the generation of novel and useful ideas/products for meeting practical needs [10]. Creativity can also be seen as relatively stable behavioral traits and cognitive abilities of creative people [100]. Creativity could also be a personal or a social and cultural phenomenon [250]. The current scope is limited to the first angle, focusing on creativity as the generation of creative ideas in teamwork.

Prior work has also looked into how different factors influence creativity at the individual and the collective levels, as well as in different contexts of work [57, 250]. In the individual level, one focus is on the enhancement of individual expertise, such as trying to figure how people could be more creative during the work, or to improve individual's productivity on specific tasks [57]. From a collective perspective, they pay attention instead on the interaction between group members [224] and how different social factors impact creativity, including cultural difference, language difference, evaluation process, etc. Specific factors identified in previous work include but are not limited to group creative style [224], collaboration atmosphere [247], group size [88], and individual cognition [58]. However, little attention has been allocated to examine the influence of the properties of communication media (e.g., video or audio) on group creativity in the literature. As computer-mediated creative work in online settings is increasingly commonplace, there's emerging need to fill this gap in research.

The relationship between communication medium and creativity roots in information exchange and interpersonal communication from the perspective of collective creativity. In management, scholars have explored the effects of specialized computer-based collaboration support tools on brainstorming tasks, resulting in the research paradigm of "electronic brainstorming"(EBS) [59]. Studies have shown that EBS improves productivity, especially in small groups [59]. Previous work also verifies that these support tools may minimize the negative process of production block due to turn-taking by allowing group members to discuss/work simultaneously to produce ideas [203]. On the other hand, these researches of system-regulated brainstorming participation aren't fully compatible with the norm of free-form, naturalistic social communication that's common in typical teamwork [55].

In regular face-to-face communication, there is no limit on verbal or non-verbal expressions, and turn-taking is typical for conversations. While in computer-mediated remote teamwork, some aspects of non-verbal communication may be restrained due to technological limitations. For example, in video conferencing the field of view of the camera used is limited even though it may still provide some visibility of team workers' facial expressions (but possibly no visibility of hand movements). On the other side, audio conferencing or telephone call can only support audibility, and there will be no visibility at all. In group creativity tasks, medium properties may affect how people express and receive ideas in the group, and may also affect what are to be shared when the expression of certain concepts is hindered by medium constraints.

5.2.2 Gesture and Creativity

Speech-accompanying hand movements, or hand gestures, are produced spontaneously during people's thinking and speaking. Non-verbal communication behavior, including hand gestures, accounts for about two-thirds of human interaction [117]. Gestures can be categorized into a couple of types based on interpreted meaning: emblems, iconic gestures, metaphoric gestures, beat gestures, and some others [139]. Research has established that different types of gestures can exert different functions in communication and social interactions. As group creativity depends on smooth exchange of concepts and ideas among team workers, it's essential to investigate the roles of hand gesture used in group brainstorming tasks.

The general functions of gesture used in conversations can be of two folds, communicative versus non-communicative [52]. Gestures produced by individuals may convey information to other participants to supplement communication goals, but may also directly associate with gesturer's own cognitive processes, such as memory retrieval and speech production [51, 139].

Gestures, with its non-communicative and self-oriented functions, are interdependent of gesturer's thinking as the inner cognitive representations. And mechanisms for thinking and gesturing may be identical. Work in embodied cognition demonstrates that cognitive processes can be shaped by the activation of other parts of the human body beyond just the brain, such as hand movement and facial expression [70]. On the other hand, gestures also complement speech in communication. Some gestures provide clear social functions substituting language itself, such as finger gestures that represent numbers, hand-waving gestures that convey social intentions (e.g., welcoming or agreement), etc. When people describe the size and shape of an object, they may use their hand gestures to show the dimensions of the object in the air. If the object is nearby, people will directly point to it. The usage of gestures is inherent to individuals and is an intrinsic device that people may use to express and demonstrate intents. Both non-communicative and communicative functions of gestures may bolster communication of concepts that's essential to group creativity [51, 139]. For example, with appropriate gesture use, people may be able to produce or share ideas that are otherwise difficult to obtain in verbal means.

Now we shift the focus to the technological aspect. In prior technical work of gesture recognition and communication applications, there has been a substantial effort into computer vision techniques to perform gesture recognition [117]. One important decision to make in these systems and applications resides on determining what characteristics or features of static hand gesture and dynamic hand movement are captured and modeled [23, 216, 292]. Another challenge resides in the source and modality of annotated data. The state of art technical research tends to focus on using multimodal sensing to derive rich representations of gestures (e.g., using RGB and depth camera like Microsoft's Kinect), and to build community-shared gesture datasets, like NATOPS Aircraft Handling Signals Database (Naval Air Training and Operating Procedures Standardization) [246], Kinect gesture dataset [165], NYU Hand Pose Dataset [265], for modeling and benchmarking purposes. While vision-based gesture recognition has been increasingly mature in certain application scenarios (e.g., content analysis of political communication), in the arena of group creativity, we still need to obtain an understanding regarding how different types of gestures affect brainstorming communication between people to determine how we may shape intrinsic brainstorming support with gesture recognition to bolster group creativity.

5.2.3 Research Questions

Here, we investigated the roles of hand movements in cross-lingual group brainstorming across different media conditions (face-to-face, video, audio). The reason we manipulated communication medium is to examine the variability of gesture use patterns in face-to-face (offline) and in computer-mediated communication (online) settings. Based on the related work and discussion above, we proposed three hypotheses to examine:

- ***H1: Medium and Group Creativity*** In group brainstorming, individuals and their partners will be more creative in rich media (e.g., face-to-face) than lean media (e.g., audio). Richer media provides more communication channels and better visibility for people, therefore people may make use of the great bandwidth to exchange concepts and ideas better with one another through verbal and non-verbal expressions. We expect people will produce more ideas face-to-face than through video, and bring up more ideas via video than audio.
- ***H2: Medium and Gesture*** Individuals will make more gestures in rich media than lean media. Since rich media affords visibility and gestures of one person will then become visible to another one, people can make use of it to use gestures to convey information. We expect that individuals will gesture more face-to-face than in the video. Similarly, compared to video, they will gesture less when communicating only via audio.
- ***H3: Gesture and Group Creativity*** Individuals who use more gestures during the brainstorming tasks will produce more ideas. Gestures are shown to be potentially functional to depict or complement the verbal contents in communication [52], through communicative and non-communicative mechanisms. Overall, we expect that individuals (and groups) who produce more hand gestures will also produce more ideas in the brainstorming session.

To clarify the roles of hand gestures in group creativity, we also performed exploratory analyses to examine the specifics of the relationship between different types of hand gestures and brainstorming performance across conditions.

5.3 Method

5.3.1 Experiment Data

We answered the hypotheses utilizing data from an experiment conducted with 72 participants (33 females) from Lin et al [164].

Participants and Procedure

In the experiment, Lin et al. invited non-English-fluent and English-fluent participants to be paired to form two kinds of language groups: *L1-L2 group* and *L2-L2 group*. Group L1-L2 consists of one English fluent speaker (L1) and one non-fluent speaker (L2); participants in group L2-L2 are both non-fluent speakers. Fluent English speakers (L1) are considered to be those who live in an English-speaking country for more than 7 years and speak English for at least half of their time there. For fulfilling the basic requirement in using English as the common language in the collaborative work, all non-fluent speakers were required to show that they scored at least 80/120 in TOEFL iBT test or an equivalent language test. Participants also didn't know the team member paired with them prior to the experiment.

During the experiment, participants were asked to discuss three brainstorming topics with the goal to generate as many ideas as possible under different medium settings. The three brainstorming tasks share a similar structure and ask participants to brainstorm about "How do you think if all humans having an extra thumb (an extra eye / a pair of wings) in the year 2020? List the possible benefits, drawbacks, and changes to people if this were to happen" [275].

Experiment Design

The experiment was designed as a within-subject experiment. The presentation orders of medium conditions and task were counterbalanced. In the three tasks, participants was set to use different communication medium conditions: face-to-face conversation (F2F), audio-only conversation (Audio) and video-mediated conversation (Video or V-Handvis). In the video condition, researchers adopted the HandVis video conferencing system to highlight trace visualizations of hand movement as communication support [163, 164].

This HandVis tool highlights the positions of participant's left and right hands with circles on the screen. HandVis updates the visual highlights dynamically in real-time in order to reduce the participant's cognitive burden in tracing and understanding hand gestures. HandVis serves in our analysis as the intervention to bserve whether the design of extrinsic communication support (e.g., visualization) has the potential to influence idea generation, as well as to affect people's intrinsic use of gestures.In terms of media richness, HandVis affords a similar level of visibility as a regular video conferencing tool. Details about the design of this HandVis system were described in [163, 164].

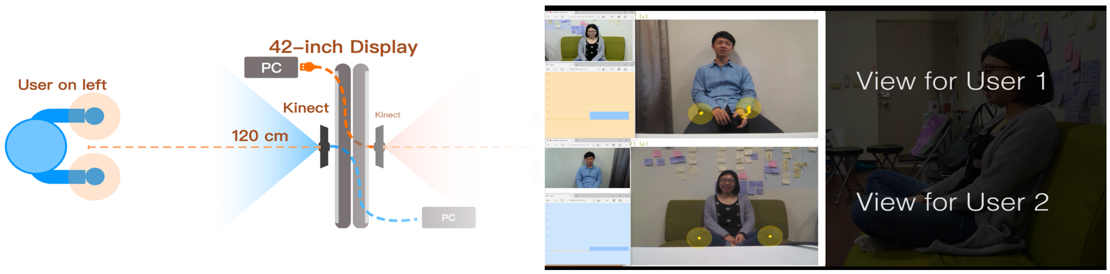
For all experiment sessions, Lin et al. used Kinect-taping and motion sensors to capture and analyze participants' hand movements during conversations [275]. Two Microsoft Kinect sensors were used to simultaneously capture the movements of all upper body joints of two participants working in the same group over the course of the brainstorming discussion.

In all conditions, participants were sitting in a fixed chair which was 120cm away from sensors, as illustrated in Figure 5.1(A) from the original experiment [164]. Figure 5.1(A) provides the bird's eye view of half of the experimental space. 42-inch displays were put on a table, locating in the middle of the room, and Kinect sensors were settled in front of the displays. There were two regular RGB cameras placed on top of the displays to record videos of the whole discussion.

In F2F condition, participants could see each other directly in the same room with the displays removed. For Audio, two participants were separated by the displays and displays were turned off to simply block the visibility. So they could only hear each other and communicate over speech, but there's no access to images. For Video or V-Handvis condition, the 42-inch displays were showing the HandVis interface designed by another project [163]. HandVis interface displayed live streaming videos of participants in real-time [163]. A screenshot of the Video condition is shown in Figure 5.1(B).

5.3.2 Measures in Experimental Data

With the experiments, the experimental dataset also includes the processed creativity performance and hand gesture measure. For instance, they code and count the number of ideas generated individually and in the group from the transcripts. And we have the count of different types of semantic gestures from the video logs and count the frequency of each type of gesture. There-



(A) (B)

Figure 5.1: Overview of the HandVis video conferencing system used in the study: (A) physical configuration of the video setting [164]; (B) user interfaces seen by both users of a video chat [163]

```
00000000.026
      head 0.07  0.3   1.31
      shoulder center  0.08  0.13  1.3
      shoulder left    -0.07 0.03  1.33
      shoulder right   0.22  0.04  1.3
      elbow left -0.11 -0.15 1.25
      elbow right0.27  -0.15 1.23
      wrist left -0.08 -0.27 1.15
      wrist right0.22  -0.26 1.12
      hand left  -0.12 -0.28 1.13
      hand right 0.2   -0.3  1.09
====================================
```

Figure 5.2: Sample data of joint positions captured by the Kinect sensor used in the study.

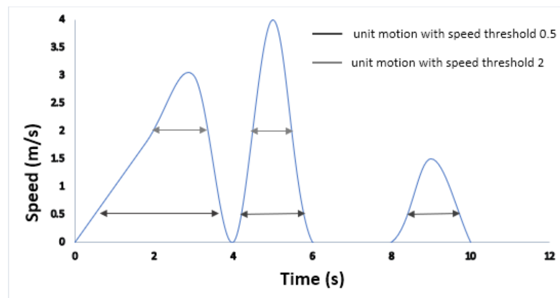


Figure 5.3: Using speed thresholds to define unit motions from the sensor-captured motion data.

fore, the independent variables involved in the experiment include task, media condition, gender, English fluency, and group language composition. All of them are categorical variables. In the rest of this section, we introduce the detailed data process they did to help readers understand the measures.

Unit Motion

Following the analysis and definition of Yabe & Tanaka [287], a gestural unit motion is defined as being static at the beginning of the motion, and then the speed increases gradually. Once the speed reaches the maximum, it starts to slow down until the movement becomes static again. Microsoft Kinect sensors were applied to track upper body joints of participants and record the skeleton data they had at the frequency of 30 frames per second, which was proposed and used previously by Wang and Lai [275]. Figure 5.2 shows a slice of the sample data we collected. The dataset consists of three-dimensional spatial positions of 10 different points of the upper body over time. Four joints related to hand movements were chosen (left hand, right hand, left elbow, right elbow). Then Euclidean speed of every joint was computed using two successive positions in the space at time t and $t+1$. And then we could derive the time/count of unit motion through speed data by applying the filtering threshold of speed. Figure 5.3 shows an example of a speed-based time series plot. If we set 2 (meter/s) as the filtering threshold of speed for coding and counting unit motions, there are 2 motions identified and the total amount of time of hand movement is about 3 seconds. They chose 0.15 m/s for the following analysis through trials to create a representative cleaned dataset consisting of multiple unit motions. More details of the data processing technique can be found in [275].

Gesture Count

Lin et al. also counted the ideas generated in the video while coding the gestures. Ideas are defined as conceptually relevant contributions relevant to the given brainstorming topics. When an idea was co-constructed by two people through the conversation, the speaker who first initiated the conversational thread of the idea was identified as the contributor for data analysis.

According to the literature of gesture analysis, four categories of gestures were coded: *Iconic Gesture* (IG), *Metaphoric Gesture* (MG), *Pointing Gesture* (PG) and *Other Gesture* (OG) [95, 139, 189]. Iconic gesture refers to hand gestures that represent meanings relevant to the concurrent semantic content. (e.g., it may look in some way like what it's intended to mean). Metaphoric gesture is similar to iconic gesture in that they exhibit or strengthen the meaning relevant to speech, but the gesture is used to present images of abstract concepts instead of concrete objects. Pointing gestures are those gestures which serve as a spatial reference. For the rest of the gestures that are not classified as IG, MG or PG, they coded them as other gestures.

5.3.3 Methodology of Data Analysis

With the dataset, a Mixed effect analysis of variance model (Mixed-effect ANOVA) is introduced to conduct data analyses. Mixed-effect ANOVA is applicable when some of the factors have fixed factor levels while others have random factor levels [197]. A classic mixed ANOVA model for two factors without interaction term, where factor A is fixed and factor B is random can be encoded as following:

$$Y_{ijk} = \mu_{...} + \alpha_i + \beta_j + \epsilon_{ijk}$$

Here, we assume $\mu_{...}$ is constant. α_i represents constant level effect for A, with the restriction that $\sum \alpha_i = 0$. β_j follows $N(0, \sigma_\beta^2)$, where σ_β^2 is the variance of factor B. And ϵ_{ijk}s are independent $N(0, \sigma^2)$, the random error term. β_j and ϵ_{ijk} are independent.

Our experiment used a randomized block design with each participant as a subject. The model was run as individual-level analysis. Every participant was assigned to one language group (L1-L2 or L2-L2) and the assignment stayed constant. Therefore, each participant was nested in one language group in modeling. By considering the possible individual difference in brainstorming performance, we implemented a random effect to account for variance between individuals. Therefore, the final model is a mixed-effect ANOVA with multiple fixed factors, such as group language composition; individual was included as a nested random factor within language group. The dependent variables varied according to hypotheses and research questions, and we took necessary pre-processing steps to ensure the premises and assumptions for the aforementioned statistical models were met. For post-hoc analyses for mixed-effect ANOVA models, we applied Tukey-Kramer's method to perform multiple comparisons of the means of different levels. It's similar to Tukey's test, but it can be applied to datasets with unequal sample sizes, such as ours. The modeling and analysis were conducted by R 3.0.

5.4 Results and Discussion

5.4.1 Analysis of H1: Medium and Group Creativity

To explore the relationship between communication medium and idea count, figure 5.4 shows the distributions of idea counts by the three medium conditions. The black dots are raw idea counts, light triangles are means for each group and boxes represent the quarters. The means and quantiles of idea count in V-Handvis ($mean = 6.40, median = 6$) are slightly greater than F2F ($mean = 5.99, median = 5.5$) and audio ($mean = 5.91, median = 5$) conditions. To complement the descriptive statistics and visual pattern shown in Figure 5.4, we used mixed-effect ANOVA to see whether they are statistically correlated shown in Table 5.1. In Model 1 of the table, both V-Handvis and F2F conditions show no statistical difference on idea count: $p = 0.15$ for V-Handvis, while $p = 0.82$ for F2F, with audio condition as a baseline. H1 is thus not supported.

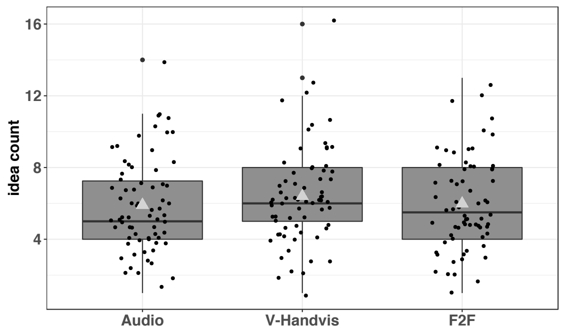


Figure 5.4: Scatter plot and box plot of idea count by medium conditions

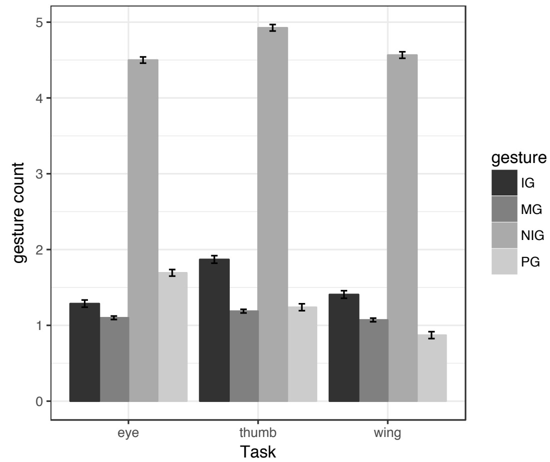


Figure 5.5: Gesture count by medium condition as estimated by mixed-effect ANOVA models

In Figure 5.4, the black dots are raw idea count data while light triangle points are means on each group.
In the Figure 5.5, bars represent the estimates of variants of gesture count. And small intervals shown at the top of bars are their 95% confidence intervals. Mean and standard deviation are estimated by mixed ANOVA model.

Table 5.1: Regression coefficients of mixed-effect models 1 and 2

	Predictors	Model 1 $\sqrt{\text{Idea-self}}$	Model 2 $\sqrt{\text{Hand-unit-motion}}$
Fixed factor	Intercept	2.3705***	7.3241***
	V-HandVis	0.1013	1.6532***
	F2F	0.0160	0.7166**
	Extra thumb	—	0.5963*
	Extra wing	—	-0.4201
standard deviation of random effect residuals	Group	0.2854	0.5011
	Indidivual	0.4072	1.9472

The intercept represents the baseline situation for audio medium condition and topic of having extra eye.
Significance code: $p < 0.01 : ***; p < 0.05 : **; p < 0.1 *$.

5.4.2 Analysis of H2: Medium and Gesture

The analysis between medium and gesture is split into two parts: the analysis for unit motion count, and the analysis for gesture subtypes. Note that not every unit motion counted here corresponds to a semantic-rich gesture (e.g., iconic or metaphoric gesture), and the gross unit motion count mainly serves as a proxy measure of the frequency of gesture use.

We applied a mixed-effect ANOVA to explore how medium influences gesture use by using the count of unit motions as the dependent variable (denoted as Model 2). Model 2 detects significant effects of communication medium and task (as a control variable) on unit motion count (V-HandVis: $p = 0.000$, F2F: $p = 0.034$; Task-thumb: $p = 0.076$, Task-wing: $p = 0.211$). It shows that participants in medium conditions with visibility, especially in V-Handvis condition ($mean = 9.03$, $sd = 2.41$), tend to produce more unit hand motions than in the audio condition ($mean = 7.39$, $sd = 2.68$).

A further detailed analysis was conducted to examine the effect of medium on each gesture subtype. We re-ran the same model (Model 2) with the frequency of gesture performed in each gesture category (IG, MG, PG, OG) as the dependent variable each time. Their results are consistent across different gestural categories as shown in Figure 5.5. In F2F and V-Handvis conditions (i.e., conditions with visibility), all kinds of gestures are more frequent than that in audio-only conversations (all p-values are $0.000 < 0.05$). The exposure to visual images in communication may encourage speakers to use gestures more frequently during their social interactions.

H2 is supported in part by results from the mixed-model ANOVAs. Through the analysis of medium effects on hand unit motions and the frequencies of different gesture subtypes, there is a main effect of media condition on gesture usage. However, we didn't see statistically significant differences between Face-to-Face and V-HandVis conditions. The effects of medium only appears when we compare audio to another medium condition that affords visibility. It's more important for people to see one another during the conversation, rather than visual quality. Restricted visibility afforded by our video conferencing tool is sufficient to trigger more frequent gesture use by participants.

Table 5.2: Correlation between gesture subtypes, hand unit motion, and idea count

	Gesture				Hand	Idea
	Iconic	Metaphoric	Other	Pointing	unit motion	
Iconic gesture	1.0000					
Metaphoric gesture	0.6051	1.0000				
Other gesture	0.0008	-0.0873	1.0000			
Pointing gesture	0.1243	0.0899	0.1337	1.0000		
Hand unit motion	0.0751	0.0386	0.1780	0.1250	1.0000	
Idea	0.2777	0.3942	0.0081	0.1625	-0.1439	1.0000

5.4.3 Analysis of H3: Gesture and Group Creativity

Gestures Effect on Self's and Partner's Ideation

To analyze the associations between idea generation and gesture use, we conducted a list of parallel analyses, including an exploratory correlation, ANOVAs for examining factors influencing self's and partner's idea count, and the impact of visibility on gesture and idea count.

Table 5.2 shows an exploratory analysis that looks into the correlations between different gesture subtypes, hand unit motion and individual's idea count. We see a negative correlation between general hand unit motion and idea count ($r = -0.14, p = 0.040$). Higher hand unit motion level implies a less idea count in return. On the other hand, there exist significant positive yet moderate correlations between idea count and iconic gesture ($r = 0.28, p = 0.000$), as well as between idea count and metaphoric gesture ($r = 0.39, p = 0.000$). More frequent metaphoric and iconic gesture usages appear to connect to greater idea productions by individuals.

With preliminary findings from the correlational analyses shown above, mixed-effect ANOVAs were applied to examine factors underlying individuals' idea generation, by including gesture subtypes as predictors (see Table 5.3 for model construction) and other control variables. As shown in Model 3 in Table 5.3, individual's idea generation is positively and significantly associated with subject's metaphoric gestures performance ($F(1, 135) = 3.91$, $p = 0.000$) as well as proficiency in English ($F(33, 135) = 1.52$, $p = 0.01$). Frequent metaphoric gesture usage predicts greater productivity in idea generation. However, we confirm that iconic gesture has no effect on idea count. Neither motion count nor motion time is statistically significant in predicting the count of idea generation.

On the other side, we aim to examine the influence of individual's gesture on his partner.

Table 5.3: Regression coefficients of idea models

	Predictors	Model 3 $\sqrt{\text{Idea-self}}$	Model 4 $\sqrt{\text{Idea-other}}$
Fixed factor	Intercept	2.5189***	2.4567***
	L2-L2 group	-0.2622**	—
	Motion count	—	0.0023**
	Metaphoric gesture	0.0246***	0.0151**
	other gesture	—	-0.0110***
Standard deviation of random effect residuals	Group	0.1973	0.2673
	Individual	0.2028	0.1593

The intercept represents the baseline situation for L1-L2 language group.
Significance code: $p < 0.01 : ***; p < 0.05 : **; p < 0.1 *$.

Model 4 in Table 5.3 uses idea counts from participants' partners as the dependent variable, and all the other settings are the same as Model 3. The result indicates that speaker's other gesture (OG) use has a significant negative effect on partner's idea productivity ($F(1, 133) = 3.91$, $p = 0.000$). Note the coefficient for other gesture is negative, which means that the more other gestures performed, the less productive the partner was in idea generation. It's again noteworthy that speaker's metaphoric gestures have a significant positive effect on partner's productivity in idea generation ($F(1, 133) = 3.91$, $p = 0.016$).

Visibility and Idea Generation

To further assess the influences of different types of gestures by medium, we created a new variable called *visibility* to split the dataset, inspired by the earlier analysis of medium and gesture reported in section 5.4.3. The two medium conditions with visibility, V-Handvis, and F2F, were combined into one level of the new dummy variable named visibility, where andio was another level. Steiger Test [249] was applied to compare correlation values between idea count and the counts of multiple types of gestures in conditions of visibility and invisibility. From table 5.4, the only significant gesture subtype is iconic gesture. Its positive correlation with idea count is significantly larger in the visible condition than in the invisible. However, for metaphoric gestures, there's no difference in its correlation with idea count in both visibility conditions.

As discussed earlier in the chapter, gestures can have communicative (other-oriented) and non-communicative (self-oriented) functions. Results of Model 3 and Model 4 reflect that the two

Table 5.4: Conditional correlation tests

	Visible	Invisible	Correlation Test p-value
Data size	136	68	
Iconic gesture	0.3449	0.1076	0.0963*
Metaphoric gesture	0.4109	0.3649	0.7209
other gesture	0.0457	-0.1014	0.3297
Pointing gesture	0.1628	0.2012	0.7931

functions are not mutually exclusive, which is consistent with the findings in previous work [139]. Metaphoric gestures contribute to the process of creation of both the speaker and his/her partner since it is a significant factor for both speaker's and her/his partner's idea generation. Also in assessing medium influence, we may differentiate visible and invisible conditions by media properties naturally. For audio, participants can't see one another and thus gestures can't convey information to their partners. In this case, gestures can only serve the non-communicative or self-oriented function. In contrast, through media with visibility (V-Handvis and F2F), gestures can convey information when people communicate, and there may also exist communicative or other-oriented function with gesture use. In Table 5.4, iconic gestures are of greater correlation with idea count in visible conditions than in the invisible condition, but we didn't see a significant impact of iconic gestures on self's or partner's idea count.

To summarize, we found that metaphoric gestures have a positive impact on self's and partner's idea productivity. The effects of metaphoric gestures on ideation are independent of the property of communication media.

5.5 Discussion

Collaborative idea generation is an intricate process that involves both individual ideation and social interaction. We did a secondary analysis to closely examine the function of one kind of nonverbal communication behavior - hand gestures - during group brainstorming under different medium conditions. We seek to understand the potential influence of different types of hand gestures on creativity performance. In summary, the results don't support H1 (rich media improve group creativity outcomes), but they are partially in favor of H2 (media richness increases the presence of hand motion/gesture) and H3 (gestures have beneficial effects on people's creativity).

We note that brainstorming performance is independent of medium condition. The lack of medium effect on creative performance stays constant regardless of whether a group involves non-native speakers or not. Previous work has mixed findings and arguments regarding whether the choice of communication medium affects work outcomes, and whether language and cultural differences further interact with media properties [33, 227]. This contributes another set of empirical results to this line of work, and we didn't find that some simple manipulation of media properties can provide effective external support for group brainstorming, either in intralingual or interlingual groups.

5.5.1 Roles of Gestures in Computer-Mediated Group Brainstorming

In testing whether and how medium properties affect the process of communication, the results show that the frequency of gestures is influenced by the visibility of the communication media. When participants are visible to each other, the amount of hand motions and gestures is also amplified. The results support the hypothesis that a person's gesture usage will be influenced by whether his gestures can be seen by the partner. What's noteworthy is that the video conferencing tool we used in the study can achieve the same effect as face-to-face communication in enhancing the amount of gesture used in conversations.

We also see interesting patterns in how general hand movement as well as the use of specific types of gestures associated with idea generation performance. The general usage of hand movement (quantified as unit motions) correlated negatively with idea counts. Team members tend to produce fewer ideas if they have a greater amount of general hand movements. Nevertheless, when examining the details of how specific types of gestures affect idea generation, we found that participants who have more metaphoric gestures are more productive in creative outputs. Other types of gestures and general hand movement don't contribute to the outcome. Note that these specific types of gestures are systematically coded according to the definitions established from previous work [95, 139, 189]. More importantly, the impact of metaphoric gesture is both self-oriented (i.e., impacting self's idea generation) and other-oriented (i.e., impacting partner's idea generation). This finding offers a novel contribution to the understanding of how speech-accompanying gestures may play a role in group creativity work.

From the semantic gestures, we found that only metaphoric gestures showed a significant association with self's and partner's idea counts. Different from iconic gestures that are for gesturally representing physical objects, metaphoric gestures are about the presentation of abstract concepts with hand movements. While possible ideas on the three brainstorming topics would arguably involve a lot of physical objects (which may be represented with iconic gestures), it appears that the formation of ideas would still depend on the conceptual parts – the ideas that one express and share to their partner that are represented by metaphoric gestures. We believe that the finding is consistent with the theoretical perspective of creative cognition, as creativity requires associative reasoning that connects remote concepts for novelty. Metaphoric gestures are likely to be an effective device to foster connectivity between concepts through embodied expressions, which appears to be a component critical to the conceptual processing behind creativity.

5.5.2 Toward Intrinsic Group Creativity Support

Throughout the analysis, we have identified that the use of metaphoric gestures may play a pivotal role in shaping group brainstorming outcomes, including both self's and partner's idea generation. Also, we found that the utility of different communication media didn't change this result. Both findings suggest that it can be an ideal "intrinsic support" to use metaphoric gestures for group creativity. The choice of communication tool (external support) may become irrelevant for supporting group brainstorming. Rather, metaphoric gestures can be seen as an intrinsic tool that's applicable widely to both offline and online settings.

Another implication from the study is the viability of gesture training as a way to endow individuals with skills for productive gesture use. Individual training of semantic gesture use could be helpful for individuals and the whole group. There are potential benefits to individuals to use own hand gesture during brainstorming conversations and to specifically increase the attention to metaphoric gesture use. Real-time interventions and computer-based tutoring may be devised to improve people's ability in perceiving and producing these gestures in accompanying their idea generation.

Although our focus in this work is on using gestures as an intrinsic mechanism, the results also imply new directions for the design and application of data-driven machine learning techniques to facilitate brainstorming. With the accumulation of modality-rich datasets of hand gesture usage in conversations, data-driven models may be trained to distinguish metaphoric gestures, which involve presentations of abstract concepts, from other types of hand movement. These models can be applied to annotate or highlight these meaningful hand movements in ways that facilitate team worker's self-gesture production and training.

5.5.3 Limitations and Future Work

Our analysis has some limitations in the applications of these findings due to the experiment topics. For example, some topics may involve more abstract concepts, and some others may involve interaction with physical objects, which may consequently affect what gestures people produce during the interaction. Therefore, it could be a limitation that we only examine limited brainstorming data.

Considering the marked difference between different individuals on personal communication styles, prior training, preferences, and constraints in expression, and cultural inclination in hand movement during conversations, the current analysis didn't examine whether and how semantic-specific gestures developed and employed by different populations (e.g., thumb-up as a gesture to represent affirmation) could be used as intrinsic brainstorming support. Rather, we may only speak to the utilities of functional gesture types like metaphoric and iconic gestures. We believe that including and accommodating individuals' gestural differences is also an important direction to pursue in future studies and design work.

Chapter 6

Lessons Learned and Future Work

The systems and user studies in this dissertation elucidate the potential technical and social support mechanisms of knowledge structuring and prioritization within the realm of online collaborative knowledge generation. Through the designing of systems, empirical evidence including interview insights and data analysis results, we attain a series of nuanced design implications in terms of designing and understanding knowledge structuring and prioritization. Next, we discuss the knowledge visibility for online knowledge collaboration in a broader scope, and future work.

6.1 Design Implications for Knowledge Structuring and Prioritizing

6.1.1 Guiding Knowledge Articulation through Social Mechanisms

Research on online communities consistently reveals that a vast majority of participants, estimated at up to 90%, remain in attendance level – engaging with content like reading blogs or using open-source projects without contributing [200]. This phenomenon, partially attributed to motivation lack or social concerns, was corroborated by our user study in Chapter 4, which also identified similar patterns of reluctance among participants to share their personal thoughts.

In response, Chapter 4 and 5 introduce and evaluate approaches aimed to *elevating the visibility of others' mental conceptions* to encourage the sharing and related thinking process. DeepThinkingMap explores exhibiting peers' understanding and reflections on a concept mapping canvas next to the video player. This approach promotes individual reflection and critical thinking, thereby encouraging the sharing of analytical and evaluative knowledge. Similarly, our findings in chapter 5 demonstrate that participants used metaphoric gestures similarly frequently as they did in face-to-face conversations when the visibility of interlocutors' hand gestures was emphasized. Crucially, these metaphoric gestures supported creative thinking intrinsically. In both studies, we present the opportunities to explicit and amplify implicit social cues that can leverage social mechanisms, such as nudging or mimicry, to improve participation and sharing between weak-tied community members. In the meanwhile, these experiments also remind the importance of social transparency via carefully designed technology affordances, such as nonverbal cues and connections in others' thoughts. However, one pitfall of this approach is that the increased visibility of user-generated contributions may inadvertently amplify potential misinformation, misleading users due to the heightened influence of visible content. One proposed mitigation strategy involves bolstering social presence alone through interface designs, though it would hinder the informational influences of shared knowledge itself.

6.1.2 Structuring and Semantic Linking

In the exploration of knowledge structuring, both Kentaurus and DeepThinkingMap employ concept maps as a basic approach to link individual contributions or curate connections to existing content. The essence of these concept maps lies in their capacity for semantic linking and the flexibility of associations. This visualization support is not unique to our systems; other innovations for discussion and document organization also utilize similar visualization strategies, such as bubble treemaps [293, 295], and layered hierarchical trees [264], to articulate semantic relationships. These methodologies also align closely with mental representation formation that also involves association and categorization [229].

These structures and links can be beneficial both in the creation and interpretation stages. From the perspective of reception and interpretation, semantically categorized knowledge as a cohesive system enhances user experiences by guiding navigation and comprehension. For instance, in our observation, using concepts as categorization tags allowed participants to explore sub-topics sequentially with Kentaurus, and the associations between tags were helpful for intuitive navigation. It is particularly beneficial for users with limited background knowledge, providing them with an accessible entry point into the topic with top-down grounding. Conversely, when contributors apply these structures during the sharing process, they may engage more deeply with the free-form associations between pieces of information. This flexibility encourages contributors to explore more connections from a wide spectrum of ideas. As detailed in Chapter 4, participants appreciated the flexibility to add new associations, without the constraints of maintaining consistency typically required in conventional documentation [282]. Along with our results, this dual-faceted benefit underscores the value of semantic structuring and linking in enhancing both the comprehension and contribution phases in collaborative settings.

6.2 Knowledge Visibility In Knowledge Collaboration

Connecting my studies, we envision a dynamic framework where knowledge flows through various knowledge work phases, undergoing continuous structuring and restructuring collectively for appropriate visibility for different goals and users. During the journey, our studies design and investigate highlight the *traceability and context* of knowledge during structuring and prioritization. For example, in Kentaurus, we link abstract concepts and individual videos or comments through semantic similarity and occurrences, to help users track between source materials and summaries. Moreover, Kentaurus provides various contexts of video content in the interface interactions. Similarly, DeepThinkingMap associates video timestamps with personal understanding, enabling efficient revisitation and understanding of content from multiple perspectives.

From the approaches to design visibility of priority knowledge during synchronous and asynchronous interactions, this dissertation explores two main lines: social computing designs and automatic computational methods. To promote social-oriented knowledge visibility, Kentaurus points out the informational potential of audience comments to categorize videos and influence video exploration processes. In the case of DeepThinkingMap, participants prioritized and confirmed key takeaways from original content, fostering discussions and personal contributions atop. On the technical front, we have leveraged advanced technologies and contextual data to enhance the visibility of crucial semantic information. Kentaurus designs an NLP computational pipeline to extract the structures and semantic associations between individual knowledge. From the experiment data in Chapter 5, sensor data, and video augmentation techniques were utilized to draw the traces of hand movements. A summary and comparison of studies is in Table 6.1. These examples, together with the rapid development of generative AI and human-AI collaboration, highlight a promising direction to analyze and curate pivotal semantic information to the forefront: by integrating both approaches, we may harness the strengths for human-centric stan-

Chapter	Methodology	Cognitive process	System example
Chapter 3	Social & Computational	Understanding	Kentaurus
Chapter 4	Social	Analyzing & Evaluating	DeepThinkingMap
Chapter 5	Computational	Creating	–

Table 6.1: Methodology and Cognitive Processes Breakdown by Chapters

dards of social computing and scalable computational methods, to reduce the cognitive burdens and align with collective human-centric goals.

6.3 General Limitations and Future Work

This dissertation introduces our investigation of knowledge structuring and prioritizing. However, it's important to acknowledge many other methods and collaboration scenarios remain unexplored. In this section, we describe some general limitations and promising avenues for future research.

6.3.1 Designing for Adaptive and Explainable Knowledge Structuring

In online communities, where resources and members continuously evolve, the clarity and self-explanatory nature of knowledge structures become crucial. Presently, these studies focus on enhancing collaboration among current community members, focusing on their immediate cooperative endeavors. Compared to the original content, however, some knowledge structures tend to be abstract and not inherently self-explanatory, such as the meaning of links in concept maps [202]. As we envisage a continuum of structured knowledge through various work phases, it's also important to re-use or re-structure knowledge, requiring evaluation and interpretation of the knowledge and its context and goals [168]. Therefore, future work may consider how to support explainability, addressing the challenges associated with knowledge transfer. For example, one starting point could be providing the rationale of links within concept maps, which could be "evidence of", "opposite to", or "lead to" in the example of video analysis and review.

Meanwhile, the process of structuring and prioritizing pivotal knowledge is dynamic and iterative, necessitating adaptations to the evolving goals of both individuals and communities [140]. Due to lab study settings, our findings are limited to short-term effects and didn't observe and evaluate the adaption of our ideas of knowledge structuring and prioritizing. Future studies can develop more adaptable designs to afford dynamic archiving and re-structuring options over time and conduct longitudinal studies. In another line, subsequent research could explore proactively detecting and responding to user needs in real time. For instance, following the recent interactive designs to support iterative knowledge exploration [167, 296], we may explore introducing semi-automatic techniques to capture user behaviors and explicit feedback to refine the structures as users process.

6.3.2 Multimodal Opportunities

This dissertation mainly works on textual and pre-recorded video resources online, and touches on real-time video meetings. With the advance of communication media and personal devices, there emerges new communication media including but not limited to audio (e.g., clubhouse), live-streaming (e.g., TikTok), Augmented Reality (AR), and Virtual Reality (VR). These media provide new information from diverse communication channel combinations, demanding greater cognitive bandwidth for effective processing. Similar NLP computational support to represent the video content could be applied to organize audio and other video-based materials when the main content is transformed into texts. Furthermore, future studies could leverage Computer Vision (CV) and multimodal generative AI to extract and merge knowledge from multiple media sources to expand the knowledge scope in consideration.

Conversely, these novel communication approaches pave another direction for designing interactions for structuring and prioritizing information. In this dissertation, most user operations and commands are through conventional interfaces, such as tablets. Our interview from Kentaurus also observes that users prefer straightforward interfaces rather than systems that have a steep learning curve. To reduce the physical operations and entry barrier of interfaces, we can imagine future opportunities that cater to the evolving needs and preferences of users, to organize, integrate, and re-structure knowledge using audio, gestures, or even thoughts.

6.3.3 Potential and Implications of Human-AI Collaborative Knowledge Visibility Design

Following the design opportunities to create adaptive and multimodal knowledge visibility support, this era also introduces significant prospects while demanding judicious design decisions for future human-AI collaborative models, for visibility of pivotal knowledge and beyond. As we harness the potential of generative AI-based support, it raises a crucial research question: How can we orchestrate human and AI intelligence to achieve shared objectives [289]. One line of work that relates to this dissertation focuses on task delegation and sharing agency, which needs research and understanding of limitations for humans and machines, such as in [147]. Drawing from theoretical frameworks such as dual process theory, for example, future studies may explore strategies to either leverage cognitive heuristics to design explainable AI that include humans in the design loop, or to bolster rational thinking by providing support for credibility assessment and contextual enhancement, as evidenced by our findings in Chapter 3.

Meanwhile, scholars have recognized the challenges of hallucination and bias within generative AI systems [297]. As an example, Large Language Models (LLM) could exhibit self-contradictions during multi-turn conversations, such as in chatGPT. A recent evaluation of Copilot, integrated into Bing and Office 365, revealed that nearly half of the references it generated lacked full support for the search engine's answers [169]. Future research needs to overcome these obstacles through additional designs carefully. For instance, researchers may develop visualization strategies to downgrade the credibility of AI-based results by associating them with alternative views or social-generated content, similar to thinking nudging designs in [209].

As we actively explore the potential of generative AI to reduce human workload and streamline group communication, we can't ignore the tensions between AI automation and human manipulation. In Chapter 3, we discussed this issue within the context of online learning and proposed the imperative for a balanced approach that fosters active engagement and exploration underpinned by automatic guidance. Drawing on Constructive psychology [245] and learning theories [232], it's evident that achieving mastery in knowledge and skills both needs ample practice and decision-making opportunities, supported by access to well-structured information. As dependence on AI-driven aids grows, it's crucial to examine their long-term effects on individual and community development, particularly concerning skill acquisition, interpersonal connections, and overall growth.

Chapter 7

Conclusion

With the advanced distributed collaboration tools and open collaboration frameworks, there emerges fruitful knowledge collaboration in online communities and a wealth of open sourced knowledge products. However, they frequently exhibit a lack of structure and organization, rendering them suboptimal for knowledge collaboration and the process of structuring and prioritizing takes significant effort, skills, and time.

In this dissertation, I demonstrate the potential of computational and social scaffolds in restructuring how knowledge is explored, analyzed, and co-created, to mitigate cognitive limitations and achieve different tasks. In Chapter 3, I explored the feasibility of blending and structuring the semantically structured overviews of audience comments and videos to support active video exploration. Chapter 4 showcased how reflective and critical thinking can be promoted through the synergy of social nudging via structured notes and conscious thinking intervention. In Chapter 5, my analysis revealed that metaphoric gestures could predict the performance of creative knowledge co-creation on both interlocutors, promising for nonverbal knowledge organization.

This work highlights the significance of socio-cognitive and technological interplay in knowledge collaboration in fluid online communities. It marks the potential toward a self-reinforcing knowledge collaboration cycle, wherein individuals and groups gather, structure, integrate, and collectively create knowledge for communal benefit with joint human and AI supports.

Bibliography

[1] GitHub: Where the World Builds Software. https://github.com/. Accessed: January, 2024.

[2] Wikipedia: The Free Encyclopedia. https://www.wikipedia.org/. Accessed: January, 2024.

[3] Semantic scholar. https://www.semanticscholar.org, 2024. Accessed: January, 2024.

[4] Z. Aghajari, E. P. Baumer, and D. DiFranzo. Reviewing interventions to address misinformation: The need to expand our vision beyond an individualistic focus. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1):1–34, 2023.

[5] D. Albarracin and S. Shavitt. Attitudes and attitude change. *Annual review of psychology*, 69:299–327, 2018.

[6] C. Albers and D. Lakens. When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of experimental social psychology*, 74:187–195, 2018.

[7] M. Almaliki. Online misinformation spread: A systematic literature map. In *Proceedings of the 2019 3rd international conference on information system and data mining*, pages 171–178, 2019.

[8] S. Alshamrani, A. Abusnaina, M. Abuhamad, D. Nyang, and D. Mohaisen. Hate, obscenity, and insults: Measuring the exposure of children to inappropriate comments in youtube. In *Companion Proceedings of the Web Conference 2021*, pages 508–515, 2021.

[9] H. Alves and A. Mata. The redundancy in cumulative information and how it biases impressions. *Journal of personality and social psychology*, 117(6):1035, 2019.

[10] T. M. Amabile. *Creativity in context: Update to the social psychology of creativity*. Hachette UK, 1996.

[11] Amazon Science. From structured search to learning to rank and retrieve. https://www.amazon.science/blog/from-structured-search-to-learning-to-rank-and-retrieve, 2023. Available online: https://www.amazon.science/blog/from-structured-search-to-learning-to-rank-and-retrieve.

[12] L. W. Anderson and D. R. Krathwohl. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman, 2001.

[13] M. Antoniak, M. Walsh, and D. Mimno. Tags, borders, and catalogs: social re-working of genre on librarything. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–29, 2021.

[14] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives. Dbpedia: A nucleus for a web of open data. In *international semantic web conference*, pages 722–735. Springer, 2007.

[15] A. Ayobi, P. Marshall, and A. L. Cox. Trackly: A customisable and pictorial self-tracking app to support agency in multiple sclerosis self-care. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–15, New York, NY, USA, 2020. Association for Computing Machinery.

[16] L. Azzopardi. Cognitive biases in search: a review and reflection of cognitive biases in information retrieval. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, pages 27–37, 2021.

[17] J. N. Bailenson. Nonverbal overload: A theoretical argument for the causes of zoom fatigue. 2021.

[18] S. Banerjee and P. John. Nudge plus: incorporating reflection into behavioral public policy. *Behavioural Public Policy*, pages 1–16, 2021.

[19] H. Bar and L. Mentch. R-cmap—an open-source software for concept mapping. *Evaluation and Program Planning*, 60:284–292, 2017.

[20] K. U. Barthel, N. Hezel, and R. Mackowiak. Graph-based browsing for large video collections. In *MultiMedia Modeling: 21st International Conference, MMM 2015, Sydney, NSW, Australia, January 5-7, 2015, Proceedings, Part II 21*, pages 237–242. Springer, 2015.

[21] E. P. Baumer. Reflective informatics: Conceptual dimensions for designing technologies of reflection. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 585–594, New York, NY, USA, 2015. Association for Computing Machinery.

[22] D. Bawden, C. Holtham, and N. Courtney. Perspectives on information overload. In *Aslib proceedings*, volume 51, pages 249–255. MCB UP Ltd, 1999.

[23] J. Beh, D. K. Han, R. Durasiwami, and H. Ko. Hidden markov model on a unit hypersphere space for gesture trajectory recognition. *Pattern Recognition Letters*, 36:144–153, 2014.

[24] I. Beltagy, K. Lo, and A. Cohan. Scibert: Pretrained language model for scientific text. In *EMNLP*, 2019.

[25] Y. Benkler and H. Nissenbaum. Commons-based peer production and virtue. *Journal of political philosophy*, 14(4), 2006.

[26] Y. Benkler, A. Shaw, and B. M. Hill. Peer production: A form of collective intelligence. *Handbook of collective intelligence*, 175, 2015.

[27] M. Bentvelzen, P. W. Woźniak, P. S. Herbes, E. Stefanidi, and J. Niess. Revisiting reflection in hci: Four design resources for technologies that support reflection. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 6(1), mar 2022.

[28] K. Bergram, M. Djokovic, V. Bezençon, and A. Holzer. The digital landscape of nudging: A systematic literature review of empirical research on digital nudges. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–16, 2022.

[29] M. M. Bhuiyan, M. Horning, S. W. Lee, and T. Mitra. Nudgecred: Supporting news credibility assessment on social media through nudges. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–30, 2021.

[30] C. Bicchieri and E. Dimant. Nudging with care: The risks and benefits of social information. *Public choice*, 191(3-4):443–464, 2022.

[31] M. V. Birk, C. Atkins, J. T. Bowey, and R. L. Mandryk. Fostering intrinsic motivation through avatar identification in digital games. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 2982–2995, 2016.

[32] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[33] P. Bordia. Face-to-face versus computer-mediated communication: A synthesis of the experimental literature. *The Journal of Business Communication (1973)*, 34(1):99–118, 1997.

[34] R. L. Boyd, A. Ashokkumar, S. Seraj, and J. W. Pennebaker. The development and psychometric properties of liwc-22. *Austin, TX: University of Texas at Austin*, pages 1–47, 2022.

[35] C. J. Brame. Effective educational videos: Principles and guidelines for maximizing student learning from video content. *CBE—Life Sciences Education*, 15(4):es6, 2016.

[36] R. N. Brewer and J. Jones. Pinteresce: exploring reminiscence as an incentive to digital reciprocity for older adults. In *Proceedings of the 18th ACM conference companion on computer supported cooperative work & social computing*, pages 243–246, 2015.

[37] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg, et al. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*, 2023.

[38] S. Buisine, G. Besacier, A. Aoussat, and F. Vernier. How do interactive tabletop systems influence collaboration? *Computers in human behavior*, 28(1):49–59, 2012.

[39] A. J. Cañas, G. Hill, R. Carff, N. Suri, J. Lott, G. Gómez, T. C. Eskridge, M. Arroyo, and R. Carvajal. Cmaptools: A knowledge modeling and sharing environment. 2004.

[40] A. Caraban, E. Karapanos, D. Gonçalves, and P. Campos. 23 ways to nudge: A review of technology-mediated nudging in human-computer interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2019.

[41] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 1–14, 2007.

[42] K. W. Chan and S. Y. Li. Understanding consumer-to-consumer interactions in virtual communities: The salience of reciprocity. *Journal of Business Research*, 63(9-10):1033–1040, 2010.

[43] T. L. Chartrand and R. Van Baaren. Human mimicry. *Advances in experimental social psychology*, 41:219–274, 2009.

[44] D. H. Chau, A. Kittur, J. I. Hong, and C. Faloutsos. Apolo: making sense of large network data by combining rich user interaction and machine learning. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 167–176, 2011.

[45] S.-L. Chen, T. Liang, M.-L. Lee, and I.-C. Liao. Effects of concept map teaching on students' critical thinking and approach to learning and studying. *Journal of Nursing Education*, 50(8):466–469, 2011.

[46] X. Cheng, T. Nolan, and L. A. Macaulay. Don't give up the community: a viewpoint of trust development in online collaboration. *Inf. Technol. People*, 26:298–318, 2013.

[47] M. T. Chi and R. Wylie. The icap framework: Linking cognitive engagement to active learning outcomes. *Educational psychologist*, 49(4):219–243, 2014.

[48] C.-M. Chiu, M.-H. Hsu, and E. T. Wang. Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories. *Decision support systems*, 42(3):1872–1888, 2006.

[49] J. Cho, T. Xu, A. Zimmermann-Niefield, and S. Voida. Reflection in theory and reflection in practice: An exploration of the gaps in reflection support among personal informatics apps. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–23, 2022.

[50] S. Choi and A. Segev. Finding informative comments for video viewing. *SN Computer Science*, 1:1–14, 2020.

[51] M. Chu and S. Kita. Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General*, 137(4):706, 2008.

[52] M. Chu and S. Kita. Co-thought and co-speech gestures are generated by the same action generation process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(2):257, 2016.

[53] R. B. Cialdini. *Influence: Science and practice*, volume 4. Pearson education Boston, MA, 2009.

[54] R. B. Cialdini, W. Wosinska, D. W. Barrett, J. Butner, and M. Gornik-Durose. Compliance with a request in two cultures: The differential influence of social proof and commitment/consistency on collectivists and individualists. *Personality and Social Psychology Bulletin*, 25(10):1242–1253, 1999.

[55] H. H. Clark and S. E. Brennan. Grounding in communication. 1991.

[56] M. Corrente and I. Bourgeault. *Innovation in transcribing data: meet otter. ai*. SAGE Publications, Ltd., 2022.

[57] F. Damanpour and D. Aravind. Organizational structure and innovation revisited: From organic to ambidextrous structure. In *Handbook of organizational creativity*, pages 483–513. Elsevier, 2012.

[58] A. R. Dennis, R. K. Minas, and A. P. Bhagwatwar. Sparking creativity: Improving electronic brainstorming with individual cognitive priming. *Journal of Management Information Systems*, 29(4):195–216, 2013.

[59] A. R. Dennis and B. H. Wixom. Investigating the moderators of the group support systems use with meta-analysis. *Journal of Management Information Systems*, 18(3):235–257, 2002.

[60] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[61] M. Diehl and W. Stroebe. Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of personality and social psychology*, 53(3):497, 1987.

[62] M. Diehl and W. Stroebe. Productivity loss in idea-generating groups: Tracking down the blocking effect. *Journal of personality and social psychology*, 61(3):392, 1991.

[63] V. Dimitrova, A. Mitrovic, A. Piotrkowicz, L. Lau, and A. Weerasinghe. Using learning analytics to devise interactive personalised nudges for active video watching. In *Proceedings of the 25th conference on user modeling, adaptation and personalization*, pages 22–31, 2017.

[64] N. Diviani, B. Van den Putte, C. S. Meppelink, and J. C. van Weert. Exploring the role of health literacy in the evaluation of online health information: insights from a mixed-methods study. *Patient education and counseling*, 99(6):1017–1025, 2016.

[65] S. Dodson, I. Roll, M. Fong, D. Yoon, N. M. Harandi, and S. Fels. An active viewing framework for video-based learning. In *Proceedings of the fifth annual ACM conference on learning at scale*, pages 1–4, 2018.

[66] V. dos Santos, É. F. de Souza, K. R. Felizardo, W. M. Watanabe, N. L. Vijaykumar, S. M. Aluizio, and A. C. Júnior. Conceptual map creation from natural language processing: a systematic mapping study. *Revista Brasileira de Informática na Educação*, 27(03):150–176, 2019.

[67] P. Drucker. *The practice of management*. Routledge, 2012.

[68] I. Dubovi and I. Tabak. An empirical analysis of knowledge co-construction in youtube comments. *Computers & Education*, 156:103939, 2020.

[69] A. Dutceac Segesten, M. Bossetta, N. Holmberg, and D. Niehorster. The cueing power of comments on social media: how disagreement in facebook comments affects user engagement with news. *Information, Communication & Society*, 25(8):1115–1134, 2022.

[70] P. Ekman and W. V. Friesen. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2):124, 1971.

[71] G. Erkan and D. R. Radev. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of artificial intelligence research*, 22:457–479, 2004.

[72] J. S. B. Evans, D. E. Over, and K. I. Manktelow. Reasoning, decision making and rationality. *Cognition*, 49(1-2):165–187, 1993.

[73] J. S. B. Evans and K. E. Stanovich. Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, 8(3):223–241, 2013.

[74] Facebook. CrowdTangle. `https://www.crowdtangle.com/`, Year of Access or Publication. Accessed: January, 2024.

[75] A. Falk and U. Fischbacher. A theory of reciprocity. *Games and economic behavior*, 54(2):293–315, 2006.

[76] J. Fang, Y. Wang, C.-L. Yang, C. Liu, and H.-C. Wang. Understanding the effects of structured note-taking systems for video-based learners in individual and social learning contexts. *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP):1–21, 2022.

[77] S. Faraj, S. L. Jarvenpaa, and A. Majchrzak. Knowledge collaboration in online communities. *Organization science*, 22(5):1224–1239, 2011.

[78] M. Fernandez and H. Alani. Online misinformation: Challenges and future directions. In *Companion Proceedings of the The Web Conference 2018*, WWW '18, page 595–602, Republic and Canton of Geneva, CHE, 2018. International World Wide Web Conferences Steering Committee.

[79] L. Finlay. Reflecting on 'reflective practice'. 2008.

[80] S. T. Fiske. *Social beings: Core motives in social psychology*. John Wiley & Sons, 2018.

[81] S. T. Fiske and S. L. Neuberg. A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In *Advances in experimental social psychology*, volume 23, pages 1–74. Elsevier, 1990.

[82] R. Fleck and G. Fitzpatrick. Reflecting on reflection: Framing a design landscape. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, OZCHI '10, page 216–223, New York, NY, USA, 2010. Association for Computing Machinery.

[83] M. Flintham, C. Karner, K. Bachour, H. Creswick, N. Gupta, and S. Moran. Falling for fake news: investigating the consumption of news via social media. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–10, 2018.

[84] B. J. Fogg. A behavior model for persuasive design. In *Proceedings of the 4th international Conference on Persuasive Technology*, pages 1–7, 2009.

[85] P. W. Foltz. Latent semantic analysis for text-based research. *Behavior Research Methods, Instruments, & Computers*, 28:197–202, 1996.

[86] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris. School engagement: Potential of the concept, state of the evidence. *Review of educational research*, 74(1):59–109, 2004.

[87] T. Fritz, A. Lill, A. N. Meyer, G. C. Murphy, and L. Howe. Cultivating a team mindset about productivity with a nudge: A field study in hybrid development teams. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW2):1–21, 2023.

[88] R. B. Gallupe, A. R. Dennis, W. H. Cooper, J. S. Valacich, L. M. Bastianutti, and J. F. Nunamaker Jr. Electronic brainstorming and group size. *Academy of Management Journal*, 35(2):350–369, 1992.

[89] C. Geeng, S. Yee, and F. Roesner. Fake news on facebook and twitter: Investigating how people (don't) investigate. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–14, 2020.

[90] H. R. Geertsen. Rethinking thinking about higher-level thinking. *Teaching Sociology*, pages 1–19, 2003.

[91] F. Geyer, J. Budzinski, and H. Reiterer. Ideavis: a hybrid workspace and interactive visualization for paper-based collaborative sketching sessions. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, pages 331–340. ACM, 2012.

[92] A. Ghazawneh and O. Henfridsson. A paradigmatic analysis of digital application marketplaces. *Journal of Information Technology*, 30:198–208, 2015.

[93] J. J. Gibson. The theory of affordances. *Hilldale, USA*, 1(2):67–82, 1977.

[94] K. Giffin. The contribution of studies of source credibility to a theory of interpersonal trust in the communication process. *Psychological bulletin*, 68(2):104, 1967.

[95] S. Goldin-Meadow. *Hearing gesture: How our hands help us think*. Harvard University Press, 2005.

[96] Google. Google docs. `https://docs.google.com`. Accessed: April, 2023.

[97] I. Goswami and O. Urminsky. When should the ask be a nudge? the effect of default amounts on charitable donations. *Journal of Marketing Research*, 53(5):829–846, 2016.

[98] R. Gouveia, F. Pereira, E. Karapanos, S. A. Munson, and M. Hassenzahl. Exploring the design space of glanceable feedback for physical activity trackers. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '16, page 144–155, New York, NY, USA, 2016. Association for Computing Machinery.

[99] M. Grootendorst. Keybert: Minimal keyword extraction with bert., 2020.

[100] J. P. Guilford. Creativity. *American Psychologist*, 5(9):444–454, 1950.

[101] F. Guimbretière, M. Stone, and T. Winograd. Fluid interaction with high-resolution wall-size displays. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 21–30. ACM, 2001.

[102] H. Halbert and L. P. Nathan. Designing for discomfort: Supporting critical reflection through interactive tools. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pages 349–360, 2015.

[103] H. Hamzah, M. I. Hamzah, and H. Zulkifli. Systematic literature review on the elements of metacognition-based higher order thinking skills (hots) teaching and learning modules. *Sustainability*, 14(2):813, 2022.

[104] M. J. Hannafin. Emerging technologies, isd, and learning environments: Critical perspectives. *Educational Technology Research and Development*, 40(1):49–63, 1992.

[105] G. Harboe and E. M. Huang. Real-world affinity diagramming practices: Bridging the paper-digital gap. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 95–104, New York, NY, USA, 2015. ACM.

[106] B. Hare. Survival of the friendliest: Homo sapiens evolved via selection for prosociality. *Annual review of psychology*, 68:155–186, 2017.

[107] P. A. Harrison and J. L. Fopma-Loy. Reflective journal prompts: A vehicle for stimulating emotional competence in nursing. *Journal of Nursing Education*, 49(11):644–652, 2010.

[108] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

[109] J. P. Hausberg and S. Spaeth. Why makers make what they make: motivations to contribute to open source hardware development. *R&D Management*, 50(1):75–95, 2020.

[110] C. Haythornthwaite. Crowds and communities: Light and heavyweight models of peer production. In *2009 42nd Hawaii international conference on system sciences*, pages 1–10. IEEE, 2009.

[111] J. He, Q. Ping, W. Lou, and C. Chen. Paperpoles: Facilitating adaptive visual exploration of scientific publications by citation links. *Journal of the Association for Information Science and Technology*, 70, 2019.

[112] J. Heer. Agency plus automation: Designing artificial intelligence into interactive systems. *Proceedings of the National Academy of Sciences*, 116(6):1844–1850, 2019.

[113] N. Helberger, K. Karppinen, and L. D'acunto. Exposure diversity as a design principle for recommender systems. *Information, Communication & Society*, 21(2):191–207, 2018.

[114] D. D. Henningsen, M. L. M. Henningsen, J. Eden, and M. G. Cruz. Examining the symptoms of groupthink and retrospective sensemaking. *Small Group Research*, 37(1):36–64, 2006.

[115] T. Hirashima, K. Yamasaki, H. Fukuda, and H. Funaoi. Framework of kit-build concept map for automatic diagnosis and its preliminary use. *Research and Practice in Technology Enhanced Learning*, 10:1–21, 2015.

[116] B. Hjørland. What is knowledge organization (ko). *Knowledge Organization*, 35:86–101, 2008.

[117] K. Hogan. *Can't get through: eight barriers to communication*. Pelican Publishing, 2003.

[118] J. Högberg, J. Hamari, and E. Wästlund. Gameful experience questionnaire (gamefulquest): an instrument for measuring the perceived gamefulness of system use. *User Modeling and User-Adapted Interaction*, 29(3):619–660, 2019.

[119] J. L. Howell and K. A. Ratliff. Not your average bigot: The better-than-average effect and defensive responding to implicit association test feedback. *British Journal of Social Psychology*, 56(1):125–145, 2017.

[120] S. Hrastinski. Asynchronous and synchronous e-learning. *Educause quarterly*, 31(4):51–55, 2008.

[121] S. Hrastinski. A theory of online learning as online participation. *Computers & Education*, 52(1):78–82, 2009.

[122] C. Hu. A new measure for health consciousness: Development of a health consciousness conceptual model. In *Unpublished paper presented at National Communication Association 99th Annual Convention, Washington, DC (November 2013)*, 2013.

[123] E. Hussein, P. Juneja, and T. Mitra. Measuring misinformation in video search platforms: An audit study on youtube. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1):1–27, 2020.

[124] G. Inc. Google cloud natural language, 2019.

[125] G. Inc. Google trends: Energy crisis european, 2023. [Online; accessed 08-Aug-2023].

[126] H. Jeong, U. Cress, J. Moskaliuk, and J. Kimmerle. Joint interactions in large online knowledge communities: The a3c framework. *International Journal of Computer-Supported Collaborative Learning*, 12:133–151, 2017.

[127] P. John and G. Stoker. Rethinking the role of experts and expertise in behavioural public policy. *Policy & Politics*, 47(2):209–225, 2019.

[128] T. Jones and K. Cuthrell. Youtube: Educational potentials and pitfalls. *Computers in the Schools*, 28(1):75–85, 2011.

[129] D. Kahneman. *Thinking, fast and slow*. Macmillan, 2011.

[130] D. Kember, D. Y. Leung, A. Jones, A. Y. Loke, J. McKay, K. Sinclair, H. Tse, C. Webb, F. K. Yuet Wong, M. Wong, et al. Development of a questionnaire to measure the level of reflective thinking. *Assessment & evaluation in higher education*, 25(4):381–395, 2000.

[131] D. Kember, J. McKay, K. Sinclair, and F. K. Y. Wong. A four-category scheme for coding and assessing the level of reflection in written work. *Assessment & evaluation in higher education*, 33(4):369–379, 2008.

[132] S. Kernan Freire, C. Wang, S. Ruiz-Arenas, and E. Niforatos. Tacit knowledge elicitation for shop-floor workers with an intelligent assistant. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–7, 2023.

[133] M. L. Khan. Social media engagement: What motivates user participation and consumption on youtube? *Computers in human behavior*, 66:236–247, 2017.

[134] Z. N. Khlaif and S. Salha. Using tiktok in education: a form of micro-learning or nano-learning? *Interdisciplinary Journal of Virtual Learning in Medical Sciences*, 12(3):213–218, 2021.

[135] J. Kim, D. Choi, N. Lee, M. Beane, and J. Kim. Surch: Enabling structural search and comparison for surgical videos. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2023.

[136] J. Kim, P. T. Nguyen, S. Weir, P. J. Guo, R. C. Miller, and K. Z. Gajos. Crowdsourcing step-by-step information extraction to enhance existing how-to videos. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 4017–4026, 2014.

[137] J. Kim, B. Tabibian, A. Oh, B. Schölkopf, and M. Gomez-Rodriguez. Leveraging the crowd to detect and reduce the spread of fake news and misinformation. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 324–332. ACM, 2018.

[138] P. A. Kirschner, J. Sweller, and R. E. Clark. Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist*, 41(2):75–86, 2006.

[139] S. Kita, M. W. Alibali, and M. Chu. How do gestures influence thinking and speaking? the gesture-for-conceptualization hypothesis. *Psychological review*, 124(3):245, 2017.

[140] A. Kittur, A. M. Peters, A. Diriye, T. Telang, and M. R. Bove. Costs and benefits of structured information foraging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2989–2998, 2013.

[141] T. Klingberg. *The overflowing brain: Information overload and the limits of working memory*. Oxford University Press, 2009.

[142] E.-Y. Ko, E. Choi, J.-w. Jang, and J. Kim. Capturing diverse and precise reactions to a comment with user-generated labels. In *Proceedings of the ACM Web Conference 2022*, pages 1731–1740, 2022.

[143] D. R. Krathwohl. A revision of bloom's taxonomy: An overview. *Theory into practice*, 41(4):212–218, 2002.

[144] A. Kuznetsov, J. C. Chang, N. Hahn, N. Rachatasumrit, B. Breneisen, J. Coupland, and A. Kittur. Fuse: In-situ sensemaking support in the browser. *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, 2022.

[145] B. H. Kwasnik. The role of classification in knowledge representation and discovery. 1999.

[146] E. R. Lai. Critical thinking: A literature review. *Pearson's Research Reports*, 6(1):40–41, 2011.

[147] V. Lai, S. Carton, R. Bhatnagar, Q. V. Liao, Y. Zhang, and C. Tan. Human-ai collaboration via conditional delegation: A case study of content moderation. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2022.

[148] D. Laniado, R. Tasso, Y. Volkovich, and A. Kaltenbrunner. When the wikipedians talk: Network and tree structure of wikipedia discussion pages. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 5, pages 177–184, 2011.

[149] P. Lara-Navarra, H. Falciani, E. A. Sánchez-Pérez, and A. Ferrer-Sapena. Information management in healthcare and environment: Towards an automatic system for fake news detection. *International Journal of Environmental Research and Public Health*, 17, 2020.

[150] C. Lawless, P. Smee, and T. O'Shea. Using concept sorting and concept mapping in business and public administration, and in education: an overview. *Educational Research*, 40(2):219–235, 1998.

[151] C. S. Lee, H. Osop, D. H.-L. Goh, and G. Kelni. Making sense of comments on youtube educational videos: A self-directed learning perspective. *Online information review*, 41(5):611–625, 2017.

[152] K. J. Lee, A. Davila, H. Cheng, J. Goh, E. Nilsen, and E. Law. "we need to do more... i need to do more": Augmenting digital media consumption via critical reflection to increase compassion and promote prosocial attitudes and behaviors. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–20, 2023.

[153] Y.-C. Lee, W.-C. Lin, F.-Y. Cherng, H.-C. Wang, C.-Y. Sung, and J.-T. King. Using time-anchored peer comments to enhance social interaction in online educational videos. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 689–698, 2015.

[154] D. Leonard and S. Sensiper. The role of tacit knowledge in group innovation. *California management review*, 40(3):112–132, 1998.

[155] A. Lewis and D. Smith. Defining higher order thinking. *Theory into practice*, 32(3):131–137, 1993.

[156] J. Li, V. Vinayagamoorthy, J. Williamson, D. A. Shamma, and P. Cesar. Social vr: A new medium for remote communication and collaboration. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2021.

[157] J. Liao, M. Singh, Y.-T. Hung, W.-C. Lin, and H.-C. Wang. Conceptcombo: Assisting online video access with concept mapping and social commenting visualizations. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*, pages 362–364, 2023.

[158] J. Liao, M. Singh, and H.-C. Wang. Deepthinkingmap: Collaborative video reflection system with graph-based summarizing and commenting. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*, pages 369–371, 2023.

[159] J. Liao and H.-C. Wang. Gestures as intrinsic creativity support: Understanding the usage and function of hand gestures in computer-mediated group brainstorming. *Proceedings of the ACM on Human-Computer Interaction*, 3(GROUP):1–16, 2019.

[160] J. Liao and H.-C. Wang. Nudge for reflective mind: Understanding how accessing peer concept mapping and commenting affects reflection of high-stakes information. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pages 1–7, 2022.

[161] L. A. Liikkanen, K. Kuikkaniemi, P. Lievonen, and P. Ojala. Next step in electronic brainstorming: Collaborative creativity with the web. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 2029–2034, New York, NY, USA, 2011. ACM.

[162] M. Limón. On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and instruction*, 11(4-5):357–380, 2001.

[163] K.-Y. Lin. A multidimensionally codified dataset of speech-accompanying gesture in cross-lingual communication. Master's thesis, Department of Computer Science, National Tsing Hua University, 2017.

[164] K.-Y. Lin, S. Yong, S.-P. Wang, C.-T. Lai, and H.-C. Wang. Handvis: Visualized gesture support for remote cross-lingual communication. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 1236–1242, New York, NY, USA, 2016. ACM.

[165] Z. Lin, Z. Jiang, and L. S. Davis. Recognizing actions by shape-motion prototype trees. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 444–451. IEEE, 2009.

[166] C. Liu, J. Kim, and H.-C. Wang. Conceptscape: Collaborative concept mapping for video learning. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2018.

[167] M. X. Liu, J. Hsieh, N. Hahn, A. Zhou, E. Deng, S. Burley, C. B. Taylor, A. Kittur, and B. A. Myers. Unakite: Scaffolding developers' decision-making using the web. *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, 2019.

[168] M. X. Liu, A. Kittur, and B. A. Myers. To reuse or not to reuse? a framework and system for evaluating summarized knowledge. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–35, 2021.

[169] N. F. Liu, T. Zhang, and P. Liang. Evaluating verifiability in generative search engines. *arXiv preprint arXiv:2304.09848*, 2023.

[170] C. Lopezosa, E. Orduña-Malea, and M. Pérez-Montoro. Making video news visible: Identifying the optimization strategies of the cybermedia on youtube using web metrics. *Journalism practice*, 14(4):465–482, 2020.

[171] S. Loria et al. textblob documentation. *Release 0.15*, 2(8):269, 2018.

[172] Z. Lu, S. Heo, and D. J. Wigdor. Streamwiki: Enabling viewers of knowledge sharing live streams to collaboratively generate archival documentation for effective in-stream and post hoc learning. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–26, 2018.

[173] S. Ludvigsen, G. Stahl, N. Law, and U. Cress. From the editors: Collaboration and the formation of new knowledge artifacts. *International Journal of Computer-Supported Collaborative Learning*, 10:1–6, 2015.

[174] B. M. DiCosola III and G. Neff. Nudging behavior change: Using in-group and out-group social comparisons to encourage healthier choices. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2022.

[175] B. M. DiCosola III and G. Neff. Nudging behavior change: Using in-group and out-group social comparisons to encourage healthier choices. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2022.

[176] X. Ma and N. Cao. Video-based evanescent, anonymous, asynchronous social interaction: Motivation and adaption to medium. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, pages 770–782, 2017.

[177] W. E. Mackay and A.-L. Fayard. Hci, natural science and design: a framework for triangulation across disciplines. In *Proceedings of the 2nd conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 223–234, 1997.

[178] K. C. Madathil, A. J. Rivera-Rodriguez, J. S. Greenstein, and A. K. Gramopadhye. Healthcare information on youtube: a systematic review. *Health informatics journal*, 21(3):173–194, 2015.

[179] A. Madden, I. Ruthven, and D. McMenemy. A classification scheme for content analyses of youtube video comments. *Journal of documentation*, 69(5):693–714, 2013.

[180] A. Majchrzak, S. Faraj, G. C. Kane, and B. Azad. The contradictory influence of social media affordances on online communal knowledge sharing. *Journal of Computer-Mediated Communication*, 19(1):38–55, 2013.

[181] L. Mamykina, B. Manoim, M. Mittal, G. Hripcsak, and B. Hartmann. Design lessons from the fastest q&a site in the west. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 2857–2866, 2011.

[182] G. Marchionini. *Information seeking in electronic environments*. Number 9. Cambridge university press, 1995.

[183] G. Marchionini. Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4):41–46, 2006.

[184] C. Marris. Public views on gmos: deconstructing the myths. *EMBO reports*, 2(7):545–548, 2001.

[185] J. Matejka, T. Grossman, and G. Fitzmaurice. Swifter: improved online video scrubbing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1159–1168, 2013.

[186] R. McCreadie, C. Macdonald, and I. Ounis. Crowdsourced rumour identification during emergencies. In *Proceedings of the 24th international conference on world wide web*, pages 965–970, 2015.

[187] N. McDonald, S. Schoenebeck, and A. Forte. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for cscw and hci practice. *Proceedings of the ACM on human-computer interaction*, 3(CSCW):1–23, 2019.

[188] M. W. McElroy. *The new knowledge management*. Routledge, 2010.

[189] D. McNeill. *Hand and mind: What gestures reveal about thought*. University of Chicago press, 1992.

[190] J. Mezirow. How critical reflection triggers transformative learning. *Adult and Continuing Education: Teaching, learning and research*, 4:199–213, 2003.

[191] A. Michalovich and A. Hershkovitz. Assessing youtube science news' credibility: The impact of web-search on the role of video, source, and user attributes. *Public Understanding of Science*, 29(4):376–391, 2020.

[192] R. Mihalcea and P. Tarau. Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*, pages 404–411, 2004.

[193] X. Mu, K. Xu, Q. Chen, F. Du, Y. Wang, and H. Qu. Moocad: Visual analysis of anomalous learning activities in massive open online courses. In *EuroVis (Short Papers)*, pages 91–95, 2019.

[194] A. J. Munro, K. Höök, and D. Benyon. *Social navigation of information space*. Springer Science & Business Media, 2012.

[195] T. Nadarzynski, O. Miles, A. Cowie, and D. Ridge. Acceptability of artificial intelligence (ai)-led chatbot services in healthcare: A mixed-methods study. *Digital health*, 5:2055207619871808, 2019.

[196] N. J. Navimipour and Y. Charband. Knowledge sharing mechanisms and techniques in project teams: Literature review, classification, and current trends. *Computers in Human Behavior*, 62:730–742, 2016.

[197] J. Neter, M. H. Kutner, C. J. Nachtsheim, and W. Wasserman. *Applied linear statistical models*, volume 4. Irwin Chicago, 1996.

[198] I. Nonaka. A dynamic theory of organizational knowledge creation. *Organization science*, 5(1):14–37, 1994.

[199] I. Nonaka and H. Takeuchi. The knowledge-creating company. *Harvard business review*, 85(7/8):162, 2007.

[200] B. Nonnecke and J. Preece. Lurker demographics: Counting the silent. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 73–80, 2000.

[201] D. A. Norman. *The psychology of everyday things.* Basic books, 1988.

[202] J. D. Novak. *Learning, creating, and using knowledge: Concept maps as facilitative tools in schools and corporations*. Routledge, 2010.

[203] J. F. Nunamaker, A. R. Dennis, J. S. Valacich, D. Vogel, and J. F. George. Electronic meeting systems. *Communications of the ACM*, 34(7):40–61, 1991.

[204] Nurmaya, A. Pinandito, Y. Hayashi, and T. Hirashima. Investigation of the relationship between map quality and higher-order thinking in kit-build concept map. In *International Conference on Human-Computer Interaction*, pages 363–377. Springer, 2023.

[205] M. Olson. The logic of collective action [1965]. *Contemporary Sociological Theory*, 124, 2012.

[206] T. O'Riordan, D. E. Millard, and J. Schulz. Is critical thinking happening? testing content analysis schemes applied to mooc discussion forums. *Computer Applications in Engineering Education*, 29(4):690–709, 2021.

[207] F. Paas, A. Renkl, and J. Sweller. Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional science*, 32(1/2):1–8, 2004.

[208] L. Pan, C. Li, J. Li, and J. Tang. Prerequisite relation learning for concepts in moocs. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1447–1456, 2017.

[209] S. Park, S. Kang, S. Chung, and J. Song. Newscube: delivering multiple aspects of news to mitigate media bias. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 443–452, 2009.

[210] S. A. Paul and M. C. Reddy. Understanding together: sensemaking in collaborative information seeking. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 321–330, 2010.

[211] P. B. Paulus and V. R. Brown. Toward more creative and innovative group idea generation: A cognitive-social-motivational perspective of brainstorming. *Social and Personality Psychology Compass*, 1(1):248–265, 2007.

[212] A. Pavel, D. B. Goldman, B. Hartmann, and M. Agrawala. Sceneskim: Searching and browsing movies using synchronized captions, scripts and plot summaries. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 181–190, 2015.

[213] J. W. Pennebaker, C. K. Chung, J. Frazee, G. M. Lavergne, and D. I. Beaver. When small words foretell academic success: The case of college admissions essays. *PloS one*, 9(12):e115844, 2014.

[214] M. Pérez Ortiz, S. Bulathwela, C. Dormann, M. Verma, S. Kreitmayer, R. Noss, J. Shawe-Taylor, Y. Rogers, and E. Yilmaz. Watch less and uncover more: Could navigation tools help users search and explore videos? In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval*, pages 90–101, 2022.

[215] P. Pirolli and S. Card. Information foraging. *Psychological review*, 106(4):643, 1999.

[216] P. K. Pisharady and M. Saerbeck. Recent methods and databases in vision-based hand gesture recognition: A review. *Computer Vision and Image Understanding*, 141:152–165, 2015.

[217] M. M. Plack, M. Driscoll, M. Marquez, L. Cuppernull, J. Maring, and L. Greenberg. Assessing reflective writing on a pediatric clerkship by using a modified bloom's taxonomy. *Ambulatory Pediatrics*, 7(4):285–291, 2007.

[218] M. Polanyi. The tacit dimension. In *Knowledge in organisations*, pages 135–146. Routledge, 2009.

[219] M. Prilla and B. Renner. Supporting collaborative reflection at work: A comparative case analysis. In *Proceedings of the 2014 ACM International Conference on Supporting Group Work*, pages 182–193, 2014.

[220] P. Pyöriä. The concept of knowledge work revisited. *Journal of knowledge management*, 9(3):116–127, 2005.

[221] N. Rachatasumrit, G. A. Ramos, J. Suh, R. Ng, and C. Meek. Forsense: Accelerating online research through sensemaking integration and machine research support. *26th International Conference on Intelligent User Interfaces*, 2021.

[222] V. Rajaraman and V. Bush. As we may think. *Resonance*, 5:94–103, 1945.

[223] A. K. Rapp, M. G. Healy, M. E. Charlton, J. N. Keith, M. E. Rosenbaum, and M. R. Kapadia. Youtube is the most frequently used educational video source for surgical preparation. *Journal of surgical education*, 73 6:1072–1076, 2016.

[224] D. K. Ray and N. C. Romano. Creative problem solving in gss groups: Do creative styles matter? *Group Decision and Negotiation*, 22(6):1129–1157, Nov 2013.

[225] N. Reimers and I. Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 11 2019.

[226] M. D. C. Reñosa, J. Landicho, J. Wachinger, S. L. Dalglish, K. Bärnighausen, T. Bärnighausen, and S. A. McMahon. Nudging toward vaccination: a systematic review. *BMJ global health*, 6(9):e006237, 2021.

[227] R. E. Rice. Task analyzability, use of new media, and effectiveness: A multi-site exploration of media richness. *Organization science*, 3(4):475–500, 1992.

[228] C. M. Ridings, D. Gefen, and B. Arinze. Some antecedents and effects of trust in virtual communities. *The journal of strategic information systems*, 11(3-4):271–295, 2002.

[229] L. J. Rips, E. E. Smith, and D. L. Medin. 11 concepts and categories: Memory, meaning, and metaphysics. *The Oxford handbook of thinking and reasoning*, page 177, 2012.

[230] A. Ruangrotsakun, D. Oh, T.-V. Nguyen, K. Lee, M. Ser, A. Hiew, R. Ngo, Z. Shureih, R. Khanna, and M. Kahng. Viva: Visual exploration and analysis of videos with interactive annotation. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces*, pages 162–165, 2023.

[231] H. Saksono, C. Castaneda-Sceppa, J. Hoffman, V. Morris, M. Seif El-Nasr, and A. G. Parker. Storywell: designing for family fitness app motivation by using social rewards and reflection. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–13, 2020.

[232] R. C. Schank, T. R. Berman, and K. A. Macpherson. Learning by doing. In *Instructional-design theories and models*, pages 161–181. Routledge, 2013.

[233] R. Schmitt-Beck. Bandwagon effect. *The international encyclopedia of political communication*, pages 1–5, 2015.

[234] F. Schmitz-Justen and A. F. X. Wilhelm. An empirical study of factors impacting on knowledge processes in online forums: factors of interest and model outline. *Int. J. Web Based Communities*, 2:318–338, 2006.

[235] P. Schultes, V. Dorner, and F. Lehner. Leave a comment! an in-depth analysis of user comments on youtube. 2013.

[236] M. Schwab, H. Strobelt, J. Tompkin, C. Fredericks, C. Huff, D. Higgins, A. Strezhnev, M. Komisarchik, G. King, and H. Pfister. booc. io: An education system with hierarchical concept maps and dynamic non-linear learning plans. *IEEE transactions on visualization and computer graphics*, 23(1):571–580, 2016.

[237] S. Sen, S. K. Lam, A. M. Rashid, D. Cosley, D. Frankowski, J. Osterhouse, F. M. Harper, and J. Riedl. Tagging, communities, vocabulary, evolution. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 181–190, 2006.

[238] C. Shao, G. L. Ciampaglia, A. Flammini, and F. Menczer. Hoaxy: A platform for tracking online misinformation. In *Proceedings of the 25th international conference companion on world wide web*, pages 745–750, 2016.

[239] Y. Shi, Y. Wang, Y. Qi, J. Chen, X. Xu, and K.-L. Ma. Ideawall: Improving creative collaboration through combinatorial visual stimuli. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 594–603. ACM, 2017.

[240] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE symposium on visual languages*, pages 336–343. IEEE, 1996.

[241] B. Shneiderman. Creativity support tools: A grand challenge for hci researchers. In *Engineering the User Interface*, pages 1–9. Springer, 2009.

[242] A. Shoufan and F. Mohamed. Youtube and education: A scoping review. *IEEE Access*, 2022.

[243] K. Shu, L. Cui, S. Wang, D. Lee, and H. Liu. defend: Explainable fake news detection. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 395–405, 2019.

[244] S. A. Sloman. The empirical case for two systems of reasoning. *Psychological bulletin*, 119(1):3, 1996.

[245] I. So. Cognitive development in children: Piaget development and learning. *Journal, of Research in Science Teaching*, 2:176–186, 1964.

[246] Y. Song, D. Demirdjian, and R. Davis. Tracking body and hands for gesture recognition: Natops aircraft handling signals database. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 500–506. IEEE, 2011.

[247] M. Sorli and D. Stokic. *Innovating in product/process development: Gaining pace in new product development*. Springer Science & Business Media, 2009.

[248] S. Spaeth and S. Niederhöfer. User motivations in peer production. *The Handbook of Peer Production*, pages 123–136, 2020.

[249] J. Steiger. Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87:245–251, 01 1980.

[250] R. J. Sternberg and R. J. Sternberg. *Handbook of creativity*. Cambridge University Press, 1999.

[251] Stveshawn. Contextual topic identification. `https://github.com/Stveshawn/contextual_topic_identification`, 2020.

[252] Y. Sun, S. Wang, S. Feng, S. Ding, C. Pang, J. Shang, J. Liu, X. Chen, Y. Zhao, Y. Lu, et al. Ernie 3.0: Large-scale knowledge enhanced pre-training for language understanding and generation. *arXiv preprint arXiv:2107.02137*, 2021.

[253] C.-Y. Sung, X.-Y. Huang, Y. Shen, F.-Y. Cherng, W.-C. Lin, and H.-C. Wang. Exploring online learners' interactive dynamics by visually analyzing their time-anchored comments. In *Computer Graphics Forum*, volume 36, pages 145–155. Wiley Online Library, 2017.

[254] C. R. Sunstein. *Why nudge?* Yale university press, 2014.

[255] C. R. Sunstein. Nudges that fail. *Behavioural public policy*, 1(1):4–25, 2017.

[256] J. Sweller. Cognitive load theory. In *Psychology of learning and motivation*, volume 55, pages 37–76. Elsevier, 2011.

[257] E. Tan. Informal learning on youtube: Exploring digital literacy in independent online learning. *Learning, media and technology*, 38(4):463–477, 2013.

[258] C.-L. Tang, J. Liao, H.-C. Wang, C.-Y. Sung, and W.-C. Lin. Conceptguide: Supporting online video learning with concept map-based recommendation of learning path. In *Proceedings of the Web Conference 2021*, pages 2757–2768, 2021.

[259] T. Tanprasert, S. S. Fels, L. Sinnamon, and D. Yoon. Scripted vicarious dialogues: Educational video augmentation method for increasing isolated students' engagement. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–25, 2023.

[260] Y. R. Tausczik and J. W. Pennebaker. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54, 2010.

[261] Y. A. S. Team. Youtube data api, 2020. `"https://developers.google.com/youtube/v3/"`.

[262] R. H. Thaler and C. R. Sunstein. *Nudge: Improving decisions about health, wealth, and happiness*. Penguin, 2009.

[263] J. D. Thompson. *Organizations in action: Social science bases of administrative theory*. Routledge, 2017.

[264] S. Tian, A. X. Zhang, and D. Karger. A system for interleaving discussion and summarization in online collaboration. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3):1–27, 2021.

[265] J. Tompson, M. Stein, Y. Lecun, and K. Perlin. Real-time continuous pose recovery of human hands using convolutional networks. *ACM Transactions on Graphics (ToG)*, 33(5):169, 2014.

[266] M. Tortoriello, R. Reagans, and B. McEvily. Bridging the knowledge gap: The influence of strong ties, network cohesion, and network range on the transfer of knowledge between organizational units. *Organization science*, 23(4):1024–1039, 2012.

[267] W. B. Towne, C. P. Rosé, and J. D. Herbsleb. Conflict in comments: Learning but lowering perceptions, with limits. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 655–666, 2017.

[268] W. M. Trochim. An introduction to concept mapping for planning and evaluation. *Evaluation and program planning*, 12(1):1–16, 1989.

[269] A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, 185(4157):1124–1131, 1974.

[270] `https://www.notion.so`. Notion. Software tool, 2024.

[271] B. Vasilescu, A. Serebrenik, P. Devanbu, and V. Filkov. How social q&a sites are changing knowledge sharing in open source software communities. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 342–354, 2014.

[272] T. Wambsganss, A. Janson, and J. M. Leimeister. Enhancing argumentative writing with automated feedback and social comparison nudging. *Computers & Education*, 191:104644, 2022.

[273] D. Wang, Q. Yang, A. Abdul, and B. Y. Lim. Designing theory-driven user-centric explainable ai. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–15, 2019.

[274] H.-C. Wang, D. Cosley, and S. R. Fussell. Idea expander: Supporting group brainstorming with conversationally triggered visual thinking stimuli. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, CSCW '10, pages 103–106, New York, NY, USA, 2010. ACM.

[275] H.-C. Wang and C.-T. Lai. Kinect-taped communication: using motion sensing to study gesture use and similarity in face-to-face and computer-mediated brainstorming. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3205–3214. ACM, 2014.

[276] J.-K. Wang, M. Ashleigh, and E. Meyer. Knowledge sharing and team trustworthiness: it's all about social ties! *Knowledge Management Research & Practice*, 4(3):175–186, 2006.

[277] M. Wang, B. Cheng, J. Chen, N. Mercer, and P. A. Kirschner. The use of web-based collaborative concept mapping to support group learning and interaction in an online environment. *Internet High. Educ.*, 34:28–40, 2017.

[278] S. Wang and E. Walker. Providing adaptive feedback in concept mapping to improve reading comprehension. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2021.

[279] R. W. White, S. T. Dumais, and J. Teevan. Characterizing the influence of domain expertise on web search behavior. In *Proceedings of the second ACM international conference on web search and data mining*, pages 132–141, 2009.

[280] W. Widjaja and M. Takahashi. Distributed interface for group affinity-diagram brainstorming. *Concurrent Engineering*, 24(4):344–358, 2016.

[281] Wikipedia. Title of the article — Wikipedia, The Free Encyclopedia, 2023. [Online].

[282] J. Wilson, A. Mandich, and L. Magalhães. Concept mapping: A dynamic, individualized and qualitative method for eliciting meaning. *Qualitative Health Research*, 26(8):1151–1161, 2016.

[283] M. L. Wilson, B. Kules, B. Shneiderman, et al. From keyword search to exploration: Designing future search interfaces for the web. *Foundations and Trends® in Web Science*, 2(1):1–97, 2010.

[284] F. K. Wong, D. Kember, L. Y. Chung, and L. Y. CertEd. Assessing the level of student reflection from reflective journals. *Journal of advanced nursing*, 22(1):48–57, 1995.

[285] Q. Wu, Y. Sang, S. Zhang, and Y. Huang. Danmaku vs. forum comments: understanding user participation and knowledge sharing in online videos. In *Proceedings of the 2018 ACM International Conference on Supporting Group Work*, pages 209–218, 2018.

[286] B. Xu, D. R. Jones, and B. Shao. Volunteers' involvement in online community based software development. *Information & management*, 46(3):151–158, 2009.

[287] T. Yabe and K. Tanaka. Similarity retrieval of human motion as multi-stream time series data. In *Database Applications in Non-Traditional Environments, 1999.(DANTE'99) Proceedings. 1999 International Symposium on*, pages 279–286. IEEE, 1999.

[288] C.-L. Yang. Understanding and reducing perception gaps with mediated social cues when building workplace relationships through cmc. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*, pages 303–306, 2021.

[289] Q. Yang, A. Steinfeld, C. Rosé, and J. Zimmerman. Re-examining whether, why, and how human-ai interaction is uniquely difficult to design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.

[290] K. Yatani. Effect sizes and power analysis in hci. *Modern statistical methods for hci*, pages 87–110, 2016.

[291] Y. Ye and K. Kishida. Toward an understanding of the motivation of open source software developers. In *25th International Conference on Software Engineering, 2003. Proceedings.*, pages 419–429. IEEE, 2003.

[292] H.-S. Yoon, J. Soh, Y. J. Bae, and H. S. Yang. Hand gesture recognition using combined features of location, angle and velocity. *Pattern recognition*, 34(7):1491–1501, 2001.

[293] A. X. Zhang, L. Verou, and D. Karger. Wikum: Bridging discussion forums and wikis using recursive summarization. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 2082–2096, 2017.

[294] T. Zhang, R. Ramakrishnan, and M. Livny. Birch: an efficient data clustering method for very large databases. *ACM sigmod record*, 25(2):103–114, 1996.

[295] X. Zhang, S. Chandrasegaran, and K.-L. Ma. Conceptscope: Organizing and visualizing knowledge in documents based on domain ontology. In *Proceedings of the 2021 Chi Conference on human factors in computing systems*, pages 1–13, 2021.

[296] X. Zhang, J. Li, P.-W. Chi, S. Chandrasegaran, and K.-L. Ma. Concepteva: Concept-based interactive exploration and customization of document summaries. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–16, 2023.

[297] Y. Zhang, Y. Li, L. Cui, D. Cai, L. Liu, T. Fu, X. Huang, E. Zhao, Y. Zhang, Y. Chen, et al. Siren's song in the ai ocean: A survey on hallucination in large language models. *arXiv preprint arXiv:2309.01219*, 2023.

[298] J. Zhao, C. Bhatt, M. Cooper, and D. A. Shamma. Flexible learning with semantic visual exploration and sequence-based recommendation of mooc videos. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–13, 2018.

Appendix A

Appendix for Chapter 3

A.1 Calculation of Video reference prerequisite and Wikipedia reference prerequisite

Video reference prerequisite (Vrp) between concepts a and b is calculated from their term frequency in the videos.As shown in Equation A.1. $Vrw(a,b)$ quantifies the frequency with which concept b is referenced in videos containing concept a. And $GVrw(a,b)$ djusts for potential concept sparsity in videos by weighting $Vrw(a,b)$ in relation to videos that are pertinent to a, rather than videos that explicitly mention a. When $Vrp(a,b) > 0$, a is a prerequisite of b from the Vrr aspect. Following the same method on Wikipedia pages, we can get the value of Wikipedia reference prerequisite (Wrf). This is adapted from Tang et al.'s system, ConceptGuide [258].

$$
\begin{aligned}
Vrp(a,b) &= GVrw(b,a) - GVrw(a,b), \\
GVrw(a,b) &= \frac{\sum_{i=1}^{M} Vrw(a_i,b) \cdot w(a_i,b)}{\sum_{i=1}^{M} w(a_i,b)}, \\
Vrw(a,b) &= \frac{\sum_{v \in V} f(a,v) \cdot r(b,v)}{\sum_{v \in V} f(a,v)},
\end{aligned}
\tag{A.1}
$$

Let $M = 10$ and $a_1, \ldots, a_10 \in C$ represent the top 10 concepts most semantically similar to concept a, where C is the universal set of all concepts. The function $w(a,b)$ computes the cosine distance between the BERT-embeddings of a and b [60]. Given that V is the set of searched videos, $f(a,v)$ indicates the term frequency of concept a in video v, while $r(b,v)$ is 1 if concept b appears in video v and 0 otherwise.

Appendix B

Appendix for Chapter 4

B.1 Videos used in Study 1 and 2

Table B.1: The meta information of the two videos used in Study 1.

Title	Video type	View count [1]
Immune System Explained I – Bacteria Infection	Theory explanation	53,195,265
The Truth About the Turmeric in Your Golden Latte	Application	510,020

1. The view count was collected in July 2023.

Table B.2: The meta information of videos used in Study 2

Title	Channel	View count (in millions) [1]	Topic [2]
Are GMOs Good or Bad? Genetic Engineering & Our Food	Kurzgesagt – In a Nutshell	12m	GMO
Why are GMOs Bad?	SciShow	3m	GMO
The Truth About GMOs	Real Science	0.5m	GMO
The Side Effects of Vaccines - How High is the Risk?	TKurzgesagt – In a Nutshell	16m	Vaccine
The Science of Anti-Vaccination	SciShow	2.9m	Vaccine
COVID-19 Vaccine for 5-to-11-Year-Old Children Explained	Public university channel	0.02m	Vaccine

[1] The view count was collected in June 2023.

[2] The videos are ordered the same as in the experiment.

B.2 Questionnaires

Please answer the questions based on your own experiences during the task on a 7-Likert scale, 1 as strongly disagree, and 7 as strongly agree. The items are adopted from [130].

Understanding

- The videos and task required me to understand the main points mentioned in the video and the general topic.
- To succeed in the task and communicate with others in the task or the future, I needed to comprehend the videos and the general topic.
- I needed to understand the videos in order to perform the task, e.g., communicate with the group or friends, finish the mapping, etc.
- During the experiment, I had to continually think about the videos that I just watched and

the general topic.

Reflective Thinking

- I questioned what others may understand and review video content, and tried to think of a better way.
- I thought over what I had been thinking and considered alternatives during the experiment.
- I reflected on video reviews and my actions to see whether I could have improved what I did and said during the experiment.
- I re-appraised my experiences in the task and learned from it, during the experiment.

Critical Thinking

- As a result of this task, I changed my attitude toward this topic.
- The task, including the processes, videos, and notes, has challenged some of my firmly held beliefs.
- As a result of this task, I found there were better ways than my normal way of consuming the videos and reviewing.
- During this experiment, I discovered some faults in what I had previously believed to be right.