Towards

PREDICTING MEMORY

in multimodal group interactions

maria tsfasman

Towards predicting memory in multimodal group interactions

Maria TSFASMAN

Towards predicting memory in multimodal group interactions

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus prof. dr. ir. H. Bijl,
chair of the Board for Doctorates,
to be defended publicly on
Monday, 23 February 2026, at 10:00 o'clock

by

Maria TSFASMAN

Master of Science in Artificial Intelligence,
Radboud University, the Netherlands,
born in Moscow, Russia.

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. C.M. Jonker	Delft University of Technology, *promotor*
Dr. B.J.W. Dudzik	Delft University of Technology, *copromotor*
Dr. C.R.M.M. Oertel	Delft University of Technology, *copromotor*
Independent members:	
Prof. dr. D. J. K. Heylen	University of Twente
Prof. dr. M. A. Neerincx	Delft University of Technology
Prof. dr. A. A. Salah	Utrecht University
Prof. dr. O. E. Scharenborg	Delft University of Technology
Prof. dr. A. Hanjalic	Delft University of Technology, *reserve member*

SIKS Dissertation Series No. 2026-17.

The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

Keywords: Conversational memory; Social signal processing; Memory encoding; Memory retention; Group interaction; First-party memory annotation; Multimodal interaction; Affect; Affective computing; Emotional saliency; Personal relevance; Automatic meeting support systems; Non-verbal signals; Multimodal corpora; Multi-party interaction

Printed by: www.proefschriftmaken.nl

Cover by: Jay Patel

ISBN 978-94-6518-249-0

An electronic copy of this dissertation is available at https://repository.tudelft.nl/.

If you talk all the time you will never hear what anybody else has to say and therefore, all you will have to talk about is your own conversation. The same is true for people who think all the time that means, when I use the word, "Think", talking to yourself, sub-vocal conversation, the constant chit-chat of symbols and images and words inside your skull. Now, if you do that all the time you'll find that you have nothing to think about except thinking and just as you have to stop talking to hear what I have to say - you have to stop thinking to find out what life is about. And the moment you stop thinking you'd come into immediate contact with what Korzybski called, so delightfully, "The unspeakable world".

Alan Watts

CONTENTS

Scientific Abstract xiii

Summary xv

Samenvatting xvii

1. Introduction 1
1.1. Research scope . . . 3
1.2. Research questions and aim . . . 6
1.2.1. Constructing a dataset for conversational memory prediction . . . 6
1.2.2. Identifying the relationship between group affect and memory annotations . . . 7
1.2.3. Identifying multimodal predictors of conversational memorability . . . 9
1.3. Societal relevance . . . 11
1.4. Dissertation outline . . . 12

2. MeMo corpus 13
2.1. INTRODUCTION . . . 15
2.2. MOTIVATION FOR CREATING MEMO . . . 16
2.3. Guiding principles for collecting MeMo . . . 19
2.3.1. P1: Maximising ecological validity . . . 19
2.3.2. P2: Maximising the construct validity of conversational memory measure . . . 21
2.3.3. P3: Considering Context-sensitivity of Memory Processes . . . 22
2.4. RELATED WORK . . . 22
2.4.1. Facilitated Modelling Perspective . . . 22
2.4.2. Memory process . . . 24
2.4.3. Recorded behaviour & measures . . . 25
2.4.4. Task context & memory task . . . 26
2.4.5. Sample representation . . . 26
2.4.6. Data quantity . . . 27
2.5. METHOD: DATA COLLECTION PROCEDURE . . . 27
2.5.1. Overall procedure . . . 27
2.5.2. Memory measures . . . 31
2.6. METHOD: DATASET PROCESSING AND CURATION . . . 34
2.6.1. Pseudo-anonymisation . . . 34

2.6.2. Data processing . . . 36
2.6.3. Data curation . . . 36
2.6.4. Extracting multimodal features . . . 37
2.7. DATASET CONTENTS . . . 39
2.7.1. Dataset subjects . . . 39
2.7.2. Multi-modal Recordings . . . 41
2.7.3. Questionnaire-based data . . . 42
2.8. VALIDATION: USING MEMO CORPUS FOR COMPUTATIONAL MODELLING . . . 45
2.8.1. Dependency Analyses . . . 45
2.8.2. Example: Group-level Conversational Memory Prediction . . . 47
2.9. DISCUSSION: POTENTIAL FUTURE TASKS . . . 48
2.9.1. Conversational memory encoding modelling . . . 49
2.9.2. Conversational memory retention modelling . . . 50
2.9.3. Perceived reason for retention modelling . . . 51
2.10. CONCLUSIONS . . . 51
2.11. Limitations . . . 52
2.12. DATASET AND CODE AVAILABILITY . . . 54

3. The relationship between memory and affect 57
3.1. Introduction . . . 59
3.2. Background and motivation . . . 60
3.2.1. Choice of annotation perspective . . . 61
3.2.2. Time-continuous operationalisation of concepts . . . 62
3.2.3. Group-based analysis . . . 63
3.2.4. Related work: Memorability prediction . . . 64
3.2.5. Conversational memory modelling . . . 65
3.3. Methods: Dataset . . . 65
3.3.1. Data source . . . 65
3.3.2. Data preparation . . . 65
3.3.3. Annotation collection . . . 66
3.3.4. Processing and derived measures . . . 67
3.4. Methods: Analysis . . . 68
3.4.1. Metrics . . . 68
3.4.2. Statistical testing procedure . . . 70
3.5. Empirical investigation . . . 73
3.5.1. Experiment 1: Random uniform . . . 73
3.5.2. Experiment 2: Random with observed range . . . 74
3.5.3. Experiment 3: Temporal shuffle . . . 76
3.6. Discussion . . . 77
3.7. Conclusions . . . 79

4. Predicting conversational memory 81
4.1. Introduction . . . 83

4.2. Background and motivation 84
4.2.1. Conversational setting 86
4.2.2. Group-level analysis 87
4.2.3. continuous operationalisation 88
4.3. The overall approach 89
4.3.1. Multimodal analysis 89
4.3.2. Qualitative analysis of memorable moments 90
4.4. Dataset 92
4.4.1. General description 92
4.4.2. Individual memorability annotation 92
4.4.3. Memory reason self-reports 93
4.5. Dataset pre-processing 94
4.5.1. Memory preprocessing 94
4.5.2. Multimodal features 97
4.5.3. MeMo subset used in this study 99
4.6. Classification methods 99
4.6.1. Model architectures 99
4.6.2. Features 101
4.6.3. Training samples 101
4.6.4. Random baseline 102
4.6.5. Feature ablation study 102
4.7. RQ1(a): Empirical results 102
4.7.1. Memory level analysis 103
4.7.2. Memory-level analysis across time 105
4.8. RQ1(b): Computational results (classification) 106
4.9. RQ2 results: Memory reason analysis 107
4.10. Discussion 109
4.11. Conclusions 113

5. Discussion & Conclusion **115**
5.1. Contributions and findings 116
5.2. Potential applications and societal implications 120
5.2.1. User modelling 120
5.2.2. Facilitation 121
5.2.3. Conversational agents 121
5.2.4. Summarisation and note-keeping 123
5.2.5. Memory augmentation 123
5.3. Limitations 125
5.4. Ethical considerations 127
5.5. Future research 128
5.5.1. Further modelling conversational encoding and retention from social signals 128
5.5.2. Further empirical research of conversational memory and inter/intra-personal factors 129

A. MeMo corpus **131**
A.1. Questionnaires used in MeMo corpus 132
A.2. Formulations of original questionnaire 134
A.3. Additional recruitment criteria 136
A.4. Separated audio synchronisation procedure 136
A.5. Longitudinal questionnaire completeness 137
A.6. Dataset version statistics 138
A.7. Questionnaire descriptive statistics 139

B. The relationship between memory and affect **141**
B.1. Visualisation of metric comparison for all experiments 142
B.1.1. DTW 142
B.1.2. PATE F1 143
B.1.3. PATE 144
B.1.4. Eucledian distance 145

Acknowledgements **171**

Curriculum Vitae **175**

List of Publications **177**

SIKS Dissertations **179**

SCIENTIFIC ABSTRACT

Conversational memory (the process by which individuals encode, retain, and retrieve information from social interactions) plays a critical role in shaping long-term social connections, guiding decision-making, and determining the impact of interpersonal communication over time. While computational systems have been developed to track short-term engagement and affective states in group interactions, no prior research has investigated how specific conversational moments are encoded into memory in real-time, nor how such memory might be modelled computationally. This dissertation addresses this gap by exploring whether group-level multimodal behaviour can predict conversational memorability in free-flowing, multi-party discussions.

To enable such modelling, this dissertation presents the MeMo corpus, the first multimodal conversational dataset annotated with first-party memory reports. The data collection approach prioritises ecological validity, participant diversity, and construct reliability, using repeated group video calls and first-party time-aligned memory annotations.

Drawing on this dataset, the dissertation then investigates whether third-party affective annotations, often used in intelligent systems, can act as a proxy for conversational memory. The analysis demonstrates that while affect and memory are conceptually related, observed affect labels do not reliably predict what participants encode in memory, underscoring the need for conversational memory-specific computational research.

The dissertation proceeds to identify behavioural indicators of memorability by conducting both empirical and computational analyses of non-verbal signals. It shows that group eye-gaze and speaker activity patterns are dependent on how likely conversational segments are to be encoded in participants' memory. A set of standard classifiers trained on these signals has shown to predict memorability at the group level with above-chance likelihood, creating a baseline for future research. In addition, a qualitative analysis of self-reported reasons for remembering reveals that people retain conversational moments that support self-image and foster interpersonal connections.

Overall, this dissertation establishes the viability and value of modelling conversational memory in multiparty settings. It lays a foundation for the development of intelligent systems capable of recognising and responding to the long-term relevance of interactions, with implications for user modelling, meeting facilitation, and memory augmentation.

SUMMARY

People often remember parts of conversations that are important to them, such as something personal, useful, or emotionally engaging. These memories help shape relationships, guide decisions, and influence how we communicate in the future. While many computer systems can already track emotions or attention in group settings, no previous research has looked at how specific moments in conversations are stored in memory or how this process could be predicted using technology.

On the path towards training computer systems to predict such memorable moments, this dissertation first introduces a new dataset called the MeMo corpus (Chapter 2). It includes group video conversations along with direct reports from participants about which moments they remembered. The data was collected in a way that reflects real-life conversations, using repeated video calls and memory reports that are linked to specific moments in time.

The study in chapter 3 then asks whether affective signals, such as emotional tone or energy in a conversation, could help predict what people will remember. These kinds of emotional signals are often used in artificial intelligence systems. However, the results show that emotional signals alone are not enough to explain what people remember from a conversation.

Next, in chapter 4, the dissertation looks at other behavioural signs, such as where people were looking and who was speaking. These signals were found to be significantly linked with memory: for example, people tend to remember parts of a conversation where there was shared attention or dynamic speaking patterns. Using these signals, simple computer models were able to predict which parts of the conversation were more likely to be remembered. The study also looked at why people remembered certain moments and found that many of them were related to personal relevance or social connection.

This work shows that it is possible to build systems that recognise which parts of a conversation are more memorable. This can be useful for improving automatic meeting tools, personal assistants, and other technologies that support communication and augment memory.

SAMENVATTING

Mensen onthouden vaak delen van gesprekken die voor hen belangrijk zijn, zoals iets persoonlijks, nuttigs of emotioneel boeiends. Deze herinneringen helpen bij het vormgeven van relaties, het nemen van beslissingen en het beïnvloeden van hoe we in de toekomst communiceren. Hoewel veel computersystemen al emoties of aandacht in groepsverband kunnen volgen, is er nog geen onderzoek gedaan naar hoe specifieke momenten in gesprekken in het geheugen worden opgeslagen of hoe dit proces met behulp van technologie kan worden voorspeld.

Op weg naar het trainen van computersystemen om dergelijke memorabele momenten te voorspellen, introduceert dit proefschrift eerst een nieuwe dataset, het MeMo-corpus (hoofdstuk 2). Deze bevat groepsvideogesprekken en directe verslagen van deelnemers over welke momenten zij zich herinnerden. De gegevens zijn verzameld op een manier die echte gesprekken weerspiegelt, met behulp van herhaalde videogesprekken en geheugenverslagen die aan specifieke momenten in de tijd zijn gekoppeld.

In hoofdstuk 3 wordt vervolgens onderzocht of affectieve signalen, zoals de emotionele toon of energie in een gesprek, kunnen helpen voorspellen wat mensen zich zullen herinneren. Dit soort emotionele signalen wordt vaak gebruikt in kunstmatige-intelligentiesystemen. De resultaten laten echter zien dat emotionele signalen alleen niet voldoende zijn om te verklaren wat mensen zich van een gesprek herinneren.

Vervolgens wordt in hoofdstuk 4 gekeken naar andere gedragssignalen, zoals waar mensen naar keken en wie er aan het woord was. Deze signalen bleken significant verband te houden met het geheugen: mensen onthouden bijvoorbeeld vaker delen van een gesprek waarin er gedeelde aandacht was of dynamische spreekpatronen. Met behulp van deze signalen konden eenvoudige computermodellen voorspellen welke delen van het gesprek waarschijnlijk beter zouden worden onthouden. Het onderzoek keek ook naar waarom mensen bepaalde momenten onthielden en ontdekte dat veel daarvan verband hielden met persoonlijke relevantie of sociale verbondenheid.

Dit werk laat zien dat het mogelijk is om systemen te bouwen die herkennen welke delen van een gesprek beter te onthouden zijn. Dit kan nuttig zijn voor het verbeteren van automatische vergadertools, persoonlijke assistenten en andere technologieën die communicatie ondersteunen en het geheugen versterken.

1

INTRODUCTION

1

Group video calls have become a prominent part of our work and personal lives, propelled by recent technological advances and Covid-19 pandemic. The fact that they happen online and have more than two participants can lead to challenges, with some group members leaving the conversation feeling unheard, misunderstood and disconnected from others [1, 2]. A promising way to address these challenges lies in computational systems that can support online meetings and enhance interactions by fostering inclusivity, resolving conflicts, and encouraging deeper connections [3, 4]. These systems track and interpret verbal and non-verbal signals to provide real-time feedback. This way, they have already shown to promote balanced participation and improve social interaction outcomes [5, 6]. These systems monitor the attentional patterns within the interaction to determine how engaged participants are at each time stamp, aiming to show participants for real-time meeting statistics, to improve meeting summarisation, or to support the understanding of how relevant the meeting is at each time stamp for the participants [7–9]. However, while pointing towards the immediate relevance of an event, these signals of relevance might not hold in the long term. In other words, they do not necessarily indicate which moments will be remembered or become socially meaningful in the upcoming interactions (in such contexts as repeated meetings within a professional team). Internal states can shift in retrospect to the event, for example, an emotion felt in the moment may be reinterpreted later [10]. It is, therefore, the remembered experience, not just the real-time signal (e.g. speaker activity or speech dynamics), that shapes relational dynamics and longer-term social outcomes [11].

Memory for conversations plays a crucial role in social bonding and behavioural prediction: while real-time attentional patterns reflect immediate relevance, it is memory (i.e., what participants encode and retain from the interaction) that shapes long-term relationships, guides future decisions, and determines the lasting impact of conversations [11]. Despite its importance, little is known about how conversational moments are remembered over time, especially from a computational perspective, where modelling which moments are likely to be remembered could support socially aware systems and memory-sensitive interaction design.

1.1. RESEARCH SCOPE

While research has identified various factors influencing conversational memory (e.g., relationships, linguistic features, participant characteristics [12–17]), it remains unclear how these factors interact to determine which moments are remembered or forgotten. Computational modelling has proven effective in inferring internal states such as emotion or engagement from non-linear patterns in multimodal behavioural data (e.g., combining facial expressions, gestures, and speech to model affect [18]). In this thesis, 'modelling' refers to internal state modelling, as commonly used in affective computing [18] and social signal processing [19], where continuous behavioural and perceptual data streams are analysed to infer underlying cognitive or affective states. Rather than relying solely on participants' reports, this approach captures the dynamic and often subtle variations in internal states in question. While previously applied to memorability of media segments [20, 21], there have been no previous attempts to computationally model memory for conversations. Some findings from media memorability may provide relevant insights for conversational context: for instance, the role of multimodal alignment in supporting encoding, or the importance of attentional allocation during naturalistic tasks in predicting incidental memory performance [20, 21]. However, live conversations represent a setting different from media consumption in several ways: they are co-constructed, interactive, temporally contingent, and variable in structure and content. These characteristics suggest that memory for conversation may rely on different cognitive and affective mechanisms and thus requires tailored modelling approaches.

Human memory is understood through three sub-processes: memory encoding (processing an experience), memory retention (preserving the experience), and memory retrieval (accessing the preserved experience) [22]. These processes are inherently linked. For instance, any study of memory involves measuring participants' retrieval (so far, this is the only known way of measuring which events were encoded and retained in participants' memory). While the three memory subprocesses cannot be entirely separated, research typically focuses on one as the primary subject of study, with methodologies tailored to the specific process under investigation. For example, studies investigating memory retrieval focus on moments when memories are accessed, often triggered by contextual relevance, such as during a collaborative task [23]. While studies of retrieval focus on the moment a memory is being accessed, studies of memory encoding examine the initial processing of a stimulus (e.g., a person's cognitive, emotional, or behavioural response at the time the experience is occurring) to understand what makes certain moments more likely to be remembered later. This means that while the memory task itself might involve retention and retrieval (e.g. free-recall or recognition [24]), the focus of

memory encoding studies is the specific stimulus viewed or behaviour displayed during the event mentioned in the reported memory.

While modelling all three subprocesses could be useful for intelligent systems applications, this thesis primarily focuses on modelling **memory encoding**. This means focusing on the moments of conversation when the remembered event occurred. This gives us the opportunity to investigate if there are properties of a specific event that make a conversational segment memorable. This is useful for intelligent systems applications, since the system would be able to monitor the potential memorability of events online as they occur, enabling the system to use this information to support the users in real-time.

Most internal state modelling, such as affect recognition, focuses on individuals, despite many intelligent system applications operating in group contexts (e.g., online meetings, public deliberations, and collaborative educational settings). However, group settings differ fundamentally from individual ones, as group-level states emerge from collective interactions and do not correspond to a mere combination of individual participants' states [25–27]. Research on conversational memory encoding has primarily focused on dyadic interactions [12–17], leaving a gap in understanding memory processes in larger group settings. This said, a recent study has explored conversational recall in a group setting, showing differences between recall quality depending on participants' roles in the conversation [28]. Although exploring conversational encoding in group settings, Brown et al. 2024, similar to other conversational memory researchers, operate on the level of one recall measure per conversation session (such as recall quality metrics). Yet, for real-time computational prediction for such applications as meeting facilitation, continuous operationalisation is needed - for example, time-aligned ground truth labels of whether or not each conversational segment was remembered. In addition, current conversational memory studies investigate speech features correlating with conversational recall, widely ignoring other modalities, such as non-verbal signals, that have shown promise in predicting related cognitive states (e.g. affect [18] or attention [9]). To summarise, this thesis aims to study the following gaps:

- **Conversational memory modelling:** While memorability of media has been modelled before, memory of conversations has not been computationally modelled before.
- **Group-level modelling:** Most existing work focuses on individuals (media context) or dyads (conversational contexts), while group-level states in conversations remain underexplored (although recently approached from a cognitive science perspective by Brown-Schmidt et al. [28]).
- **Continuous operationalisation:** Current conversational memory

studies often use one recall metric per session, but real-time applications require time-aligned, segment-level memory labels.

- **From unimodal to multimodal analysis:** Conversational memory research primarily uses speech features, neglecting non-verbal modalities.

This thesis aims to address these gaps by investigating the possibility of group-level, continuous, multimodal prediction of conversational memory encoding.

1

1.2. RESEARCH QUESTIONS AND AIM

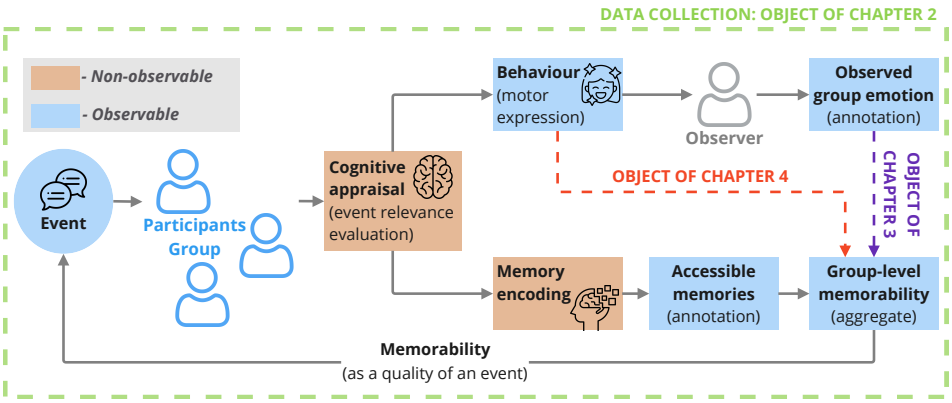


Figure 1.1.: The research questions and main outcomes of the content chapters of the thesis

This thesis aims to pave the way towards predicting the likelihood of a conversational moment being encoded in participants' memory in the context of free-flowing multi-party conversations. On the path towards this aim, we conduct three studies, visualised in *Figure 1.1*. The first contribution of this thesis (*Chapter 2*, green frame in *Figure 1.1*) lays down the foundation for the other two studies. It presents a corpus aimed at conversational memorability modelling. The second study (*Chapter 3*, purple arrow in *Figure 1.1*) investigates the utility of memorability modelling by analysing whether memorability annotations can be uniquely derived from emotion annotations. The final study (*Chapter 4*, orange arrow in *Figure 1.1*) dives into the memorability prediction, investigating what behaviours can signal memorability and the distributions underlying reasons for remembering a moment. We describe each chapter's research gaps and research questions in more detail in the next subsections.

1.2.1. CONSTRUCTING A DATASET FOR CONVERSATIONAL MEMORY PREDICTION

The first step on the path towards conversational memory modelling is creating a dataset that is specifically crafted for this purpose. Computational modelling of socio-cognitive states usually involves training a predictive model on data annotated with ground truth labels of an investigated state, so that, given new data, the model is able to identify the labels based on the observed behavioural patterns [29]. Such computational modelling relies on high-quality, ecologically valid

datasets with reliable ground-truth labels [30–32]. Data collection is time- and resource-consuming, and sometimes it is possible to adopt an existing dataset by adding a required annotation to it. However, in the case of memory, that is not possible for two reasons. First, the annotations need to be done by the participants themselves since the third-party observers' memory of conversations is quantitatively and qualitatively different from that of the participants [13, 33]. Second, it's important that all annotators use a consistent time frame for their annotations, because the longer it has been since the event, the more likely it is that participants will be at different stages of remembering or forgetting the information [34]. While there are some datasets annotated with memory in the context of media perception [21, 35], no such dataset exists in a conversational context. This is why our first research question is as follows:

RQ1: Conversational memory dataset

How can a multimodal conversational dataset be designed to validly capture first-party memory reports to support computational modelling of memory processes in multi-party meetings?

To answer this research question, we have collected the MeMo corpus: the first conversational corpus continuously annotated with participants' memory reports, aimed at multimodal modelling of encoding and retention in multi-party conversations [36]. The MeMo corpus is designed to facilitate research with two primary goals: identifying verbal and non-verbal signals associated with conversational memory and developing models that aid in meeting support by predicting memory outcomes in the context of repeated interactions. In *Chapter 2* (published as Tsfasman et al. [36]), we present this dataset, describe the principles for its design, discuss its validity, present proof of its usefulness, and introduce the problems that can be addressed using this corpus.

1.2.2. IDENTIFYING THE RELATIONSHIP BETWEEN GROUP AFFECT AND MEMORY ANNOTATIONS

Affective Computing has advanced techniques for recognising human emotions, often through multimodal signals such as facial expressions and speech, to improve user interaction with intelligent systems [18]. Emotions are known to influence cognitive processes, especially memory, with emotional arousal and valence shown to enhance memory encoding and recall [37–40]. It is logical to hypothesise, based on these previous findings, that perceived affect could potentially act as a proxy for memory, a premise that has motivated the integration

of emotional components into computational memory models across various intelligent systems [41–43].

Despite these promising conceptual links, there remain key gaps that limit our understanding of how affective annotations relate to memory within practical Multimodal Emotion Recognition (MER) settings. First, while behavioural science studies typically rely on first-person self-reports and physiological measures of experienced emotion, MER often uses third-party observers' annotations of visible behaviour, which may not accurately reflect internal emotional states or memory relevance [44, 45]. Second, research on the emotion-memory relationship generally treats affect and memory as static states, whereas MER systems increasingly adopt time-continuous annotations to capture dynamic emotional changes [46, 47], yet how these continuous measures relate to memory remains unclear. Third, most empirical work has focused on individuals, despite many MER applications operating in social groups where collective emotions and group-level memory dynamics play crucial roles [25, 48]. Addressing these gaps, *Chapter 3* empirically investigates the association between time-continuous perceived group emotions (arousal and valence) and group memorability in naturalistic conversational settings. This leads to our second research question approached in *Chapter 3*:

RQ2: Group affect as a proxy to memory

To what extent do third-party, time-continuous annotations of perceived group emotions (arousal and valence) predict group-level memorability in unstructured, naturalistic conversational interactions?

This paper investigates whether group affect annotations can predict conversational memory, using data from the MeMo corpus (described in *Chapter 2*). We investigate the relationship between perceived group affect (measured continuously through third-party annotations of arousal, valence, and intensity) and group memorability. To assess this relationship, we employ metrics sensitive to temporal dynamics, including the Proximity-Aware Time series Evaluation (PATE), PATE F1, Euclidean distance, and Dynamic Time Warping (DTW). These metrics allow us to capture both categorical and continuous similarities while accounting for temporal shifts typical in human behavioural data. We further validate our findings by comparing real affect-memory alignments against multiple null hypotheses generated via synthetic data, testing whether observed associations exceed what could be expected by chance or temporal misalignment.

1.2.3. IDENTIFYING MULTIMODAL PREDICTORS OF CONVERSATIONAL MEMORABILITY

Chapter 3 concluded that observed affect annotations do not convey the same information as memorability labels, further justifying the importance of memorability modelling. In *Chapter 4* we, therefore, dive into conversational memorability modelling.

Humans are social creatures that continuously express their internal states not only through speech, but also with their body language. These non-verbal signals serve to communicate one's needs, intentions, and state of mind [49]. Therefore, computational models are developed to predict a user's internal state from their face expressions, hand gestures, body pose, and eye gaze [29]. While some computational models are trained to predict emotions, dominance, and involvement [44, 50, 51], only a few are developed to predict how likely a human is to encode and retain an event in their memory. For example, there have been a few models trained to predict how likely a person is to encode and later retain an event from their brain signals [20, 52]. There are also some models that have been trained to predict the memorability of videos and images based on the stimulus characteristics [35, 53]. However, brain signals are rarely accessible to intelligent systems (due to difficulties in collecting those as well as privacy concerns) and there is no clear stimulus/perceiver boundary in the context of free-flowing conversations. For example, when one person is talking, their speech is an auditory stimulus for another, but it is unclear whether one's own speech is also a part of a stimulus. Therefore, for intelligent system applications (see *Section 5.2*) that have access to the videos of the participants, it would be useful to employ a computational model that predicts the likelihood of an event being encoded in the user's memory based on participants' verbal and non-verbal behaviour. Although such multimodal models have been previously developed to predict autobiographical recall in media contexts [54], social signals have never been used to predict conversational encoding and retention in conversation. This leads to our third research question approached in *Chapter 4*:

RQ3: Non-verbal signals and memory

Can non-verbal behaviours, such as group eye gaze and speaker activity, serve as indicators of which conversational moments are more likely to be encoded in participants' memory? If so, what specific patterns in these signals predict conversational memory?

Group eye-gaze behaviours have been previously shown to be predictive of affect [55], involvement [56] and attention [57] in social interactions. Since affect, attention and involvement are closely related

to memory [58, 59], it is logical to hypothesise that similar non-verbal signals would be predictive of conversational memory. Therefore, the hypothesis was that group eye-gaze behaviour can be used to predict how memorable the moment is for a group. In *Chapter 4*, published as Tsfasman et al. [60], we investigated this question via two types of analysis: machine learning-based modelling and traditional statistical analysis. From a modelling perspective, we trained a computational model on group eye-gaze signals and speaker activity features. From a traditional statistical perspective, we analysed what kind of eye-gaze and speaker activity patterns are more common within moments of different memorability levels. We also investigated what kind of non-verbal signals are more likely to occur right before or after a memorable segment.

Human memory is selective, and only personally relevant events get encoded and retained [61]. Conversational memory can, therefore, be seen as an indication of conversational event relevance per participant or the group as a whole. As personal relevance depends on personal motivation, memorable moments can be categorised by the type of motivation of the moment's relevance. The second research question of this chapter is, therefore, the following:

RQ4: Reasons for remembering

What are the common types of self-reported reasons for remembering a conversational moment?

Previous research hypothesises that humans tend to retain experiences that (1) enrich or confirm their self-image [15], (2) connect them to other individuals [16, 62, 63], or (3) guide their future actions, thoughts and responses [11]. If that is true, the most common motivations behind remembering a moment would be connected to these functions.

1.3. SOCIETAL RELEVANCE

Conversational memory modelling potential benefits society in various applications. In this section, we briefly introduce those.

User modelling for long-term interaction. The success of intelligent systems for long-term interaction depends heavily on their ability to adapt to individual users by building rich and dynamic user profiles [64]. While commonly modelled internal states such as affect, engagement, and mood can personalise short-term interactions, effective long-term personalisation requires systems to account for the selective nature of human memory, i.e. the ability to recall only personally relevant experiences from past interactions [61, 65]. Memory modelling addresses the gap between what users retain versus what systems assume they remember, enabling intelligent systems to align more closely with human cognitive processes. This alignment is critical, as systems designed for sustained engagement must adapt not only to individual user preferences but also to group-level behaviour when functioning in multi-user settings [27].

Facilitation of social interaction. Loneliness and poor relationship quality are pressing societal issues that affect mental and physical health [66–68]. The field of affective computing [69] and the development of conversational facilitation tools [70] have made efforts to enhance the quality of human connections by fostering understanding and reducing misunderstandings in group interactions. By integrating memory modelling, such systems can identify shared conversational moments and reinforce these as common ground, ultimately strengthening social bonds [63, 71]. The research in this thesis is motivated by the idea that meeting facilitation systems with memory-aware capabilities could hold the potential to deepen conversations and foster lasting relationships, addressing societal needs for meaningful connections and improved collaboration in personal and professional settings.

Conversational agents and memory. Memory plays a central role in human conversations, enabling individuals to maintain social bonds and demonstrate social intelligence [16, 62]. For conversational agents to succeed in long-term interactions, they must replicate this ability, remembering and appropriately referencing past exchanges to build rapport and sustain user engagement [72, 73]. However, existing agents often misinterpret the concept of shared memory, assuming full recall of past interactions rather than the selective memory processes of users [74]. Memory models informed by datasets like the MeMo corpus [36] can help agents align with users' true shared memories, improving their social presence and ability to adapt over time.

Summarisation and personalisation. Conversational memory models have potential in aiding meeting summarisation by emphasising the information most likely to be retained by participants, rather than

relying solely on generic measures of importance (e.g. such widely used measures as involvement [50, 75]). This approach allows for the creation of personalised and concise summaries, tailored to the specific recall patterns of individuals or groups. By predicting memory likelihood, systems can also intervene during conversations, reinforcing key points to ensure they are remembered. Such advancements enhance the relevance and effectiveness of meeting support tools, improving communication and decision-making efficiency in team settings.

Memory augmentation. Finally, our research on conversational memory modelling could have potential for memory augmentation applications. With the ever-growing volume of digital content (from photos and videos to messages and meeting recordings), memory augmentation tools are being developed to extract and highlight relevant moments when needed [76]. Current approaches for extracting key moments from meeting footage typically rely on generic measures, such as text characteristics or frequency of topic repetition [77], which may not capture what is personally significant for each individual. By predicting which conversational moments individuals are likely to remember, our approach not only identifies content with high personal relevance for later review but also offers the potential to supplement human memory with details that may otherwise be overlooked. Moreover, as memory augmentation is crucial for supporting individuals with memory disorders (particularly given the increasing prevalence of conditions like Alzheimer's, dementia, and Parkinson's disease [78]), integrating data from healthy populations (e.g. the MeMo corpus) may enhance the performance of life-logging systems by prioritising particularly memorable events.

1.4. DISSERTATION OUTLINE

This thesis is written based on 3 academic papers [36, 60, 79]. *Chapter 2*, *Chapter 3* and *Chapter 4* present these papers with minor modifications. *Chapter 2* aims to answer RQ1 (based on Tsfasman et al. [36]). It presents MeMo, the first dataset with conversational memory annotations, and defines the problems that can be investigated using the dataset. *Chapter 3* investigates the relevance of memory investigation, by researching to what extent group affect, a much more explored internal state, can be used to uniquely derive memory labels (RQ2, based on [79]). *Chapter 4* presents a conversational memory prediction model based on eye-gaze patterns and speaker activity and an empirical investigation into the relationship between those signals and group memorability (answering RQ3 and RQ4). The titles of the chapters are modified, and the references are unified for better readability. *Chapter 5* concludes the thesis, discussing the main results and the potential future directions for conversational memorability modelling and research.

2

MEMO CORPUS

[submitted to Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies] **M Tsfasman**, B Dudzik, K Fenech, A Lorincz, CM Jonker, C Oertel, "Introducing MeMo: A Multimodal Dataset for Memory Modelling in Multiparty Conversations," *arXiv preprint arXiv:2409.13715*, 2024.

ABSTRACT

Conversational memory is the process by which humans encode, retain and retrieve verbal, non-verbal and contextual information from a conversation. Since human memory is selective, differing recollections of the same events can lead to misunderstandings and misalignments within a group. Yet, conversational facilitation systems, aimed at advancing the quality of group interactions, usually focus on tracking users' states within an individual session, ignoring what remains in each participant's memory after the interaction. Understanding conversational memory can be used as a source of information on the long-term development of social connections within a group. This paper introduces the MeMo corpus, the first conversational dataset annotated with participants' memory retention reports, aimed at facilitating computational modelling of human conversational memory. The MeMo corpus includes 31 hours of small-group discussions on Covid-19, repeated 3 times over the course of 2 weeks. It integrates validated behavioural and perceptual measures, audio, video, and multimodal annotations, offering a valuable resource for studying and modelling conversational memory and group dynamics. By introducing the MeMo corpus, analysing its validity, and demonstrating its usefulness for future research, this paper aims to pave the way for future research in conversational memory modelling for intelligent system development.

2.1. INTRODUCTION

Human memory for conversations plays a crucial role in shaping social bonds and fostering relationship building, as well as decision-making in future interactions [11]. Understanding human conversational memory is, thus, essential for explaining and predicting human behaviour in conversations. Conversational memory can be defined as a subtype of autobiographical memory, which manages the encoding, storage, and retrieval of personally experienced events [80–82], particularly within conversational settings (as operated in e.g. [33, 83, 84]). Previous research on conversational memory shows that numerous factors can affect what is encoded and retained from a conversation: the relationship between participants, their characteristics, linguistic features of produced speech and many more [12–17]. However, due to the multitude of variables involved, it remains unclear how these factors interact to determine which memories are more likely to be retained and which are more likely to be forgotten over time.

One way of investigating the intricate, potentially non-linear relationship between contributing factors to a socio-cognitive process, such as memory, is by creating a computational model. Although never applied to conversational memory, to our knowledge, other socio-cognitive phenomena, such as affect, engagement or cohesion, have been previously investigated using computational models [18, 29, 85]. Researchers model these socio-cognitive processes by training machine learning algorithms to predict users' internal states from verbal and non-verbal data. A major issue to take into account when building such a model is the data used for its creation, since the model can only be as accurate as the data it is trained on [31]. Similar to the reproducibility crisis in the field of social sciences [86], there is more and more understanding of how a biased or misconstrued dataset can be detrimental to the reproducibility, generalisability and validity of the resulting models [32, 87]. Many scholars are calling for more careful creation of datasets aimed at computational modelling [31, 32, 87]. Therefore, constructing a computational model of conversational memory requires a dataset that is representative of the modelled constructs and is collected in an ecologically valid setting.

Since there is no such data available for conversational memory research, in this paper, we introduce the **MeMo** (**Me**mory **Mo**delling) corpus - the first conversational corpus with annotations for conversational memory. The *MeMo* corpus is aimed to be used for computational modelling of conversational memory developed with multidisciplinary research in mind. It, therefore, combines validated behavioural and perceptual measures as well as video, audio, and multimodal annotations, individual and group eye gaze behaviour, head pose, low-level hand gestures, and text. The variety of measures, multimodality and ecological validity of the corpus make it a useful resource for

computational as well as behavioural studies on conversational memory and group dynamics. Because of the data complexity, the dataset will be released in batches with the process described in *Section 2.12* once the pseudo-anonymisation process is complete. In this paper, we aim to achieve the following goals:

- **Introducing the *MeMo* corpus.** We describe the *MeMo* corpus, the data collection, and its challenges. The *MeMo* corpus pioneers a way of collecting first-party memory annotations directly usable in computational research - by combining a moment free-recall task with a subsequent first-party annotation of the recorded memorable moments to a video time frame (see *Section 2.5*).
- **Demonstrating its usefulness.** We demonstrate how the corpus can be used to build conversational memory models and summarise empirical results on the corpus (see *Section 2.8.2*).
- **Suggesting potential topics of future research using the corpus.** We describe potential modelling tasks that can be explored with the use of the *MeMo* corpus (see *Section 2.9*).

2.2. MOTIVATION FOR CREATING MEMO

Humans are inherently social creatures, and the quality of one's social connections significantly impacts their psychological and physiological well-being [88]. Feeling listened to, understood, and appreciated within a social relationship is essential for fostering these quality connections [70]. In group settings, such as family gatherings or work meetings, it could be challenging to achieve this quality because of differences in personality, dominance and other factors. Conversation facilitation has emerged as a promising approach to enhancing the quality of these group interactions [89, 90]. Over repeated sessions with a trained facilitator, groups of people can resolve conflicts, deepen conversations, and foster mutual understanding across various social settings [3, 91, 92]. Conversation facilitation can be challenging, requiring undivided attention towards multiple team members, conversation structure and content [93]. Computational systems can support human facilitators in this task, as well as serve as an alternative solution in the absence of a human facilitator [94].

Existing conversation support systems have been shown to improve social interaction satisfaction, encourage equal participation and decrease social inhibitions [4, 6, 95]. A common method of achieving these results involves continuous tracking of users' non-verbal and verbal signals [96]. These low-level signals are then used to infer a real-time measure of users' participation (e.g. [6]) or more complex internal states, such as attention (e.g. [97]), dominance (e.g. [98]) or

social presence (e.g. [99]). Based on these predicted measures, a system then produces suggestions on how to enforce meeting structure and promote equal participation (e.g. [96], [6]). These predicted measures are real-time indicators of a participant's immediate reactions or internal states, captured at specific timestamps or as cumulative measures from the start of the session. Although these measures can be used to represent users' current state or trends within a session, they do not always represent the way the user will feel about the subject in subsequent interactions, since feelings triggered by an event within an interaction could be forgotten or completely changed over time, in retrospect to the event [100]. To sum up, user experience itself may not matter as much as the user's memory of that experience in the context of long-term interaction and future decisions [101]. Consequently, for long-term interaction, a facilitation system needs to not only track the current state of the user but also understand the user's memory of conversational experiences.

Decades of cognitive research show that, due to the selective nature of human memory, only a fraction of perceived experiences are encoded and retained [22]. The retention or forgetting of experiences and the subsequent accessibility of memories are affected by an intricate combination of inter- and intra-personal factors, such as conversational context, linguistic parameters of speech, one's role in the conversation, and conversational skills [15–17, 102]. Apart from these over-arching factors, according to previous research, humans tend to retain experiences that (1) enrich or confirm their self-image [15], (2) connect them to other individuals [16, 62, 63] or (3) guide their future actions, thoughts and responses [11]. While previous research has identified these factors and functions of human memory, how they operate together in spontaneous settings, determining whether an event will be retained or forgotten remains unclear. Corresponding to the mentioned memory functions, understanding what remains in a user's memory after an interaction could help a computational system in (1) understanding a user's personal preferences and identity, (2) keeping track of relational development between conversational partners and (3) understanding the origin of perspectives and decisions in future interactions.

While some studies have explored ways of supporting memory in conversational settings, these have primarily focused on memory augmentation rather than memory modelling. For instance, Niforatos et al. [103] presented a system that improves recall of previous meetings using auto-generated visual memory aids, based on semantic summaries of previous meetings. Bahrainian and Crestani [104] proposed creating models of topics that were most likely to be brought up again throughout the conversation for memory augmentation systems. Though valuable, such studies focus on enhancing memory

retrieval without directly modelling how memories are encoded and retained during the interaction itself by the conversation participants.

Unlocking the potential of memory encoding modelling for memory augmentation, user-modelling and personalisation purposes in conversational settings requires predictive models based on real-world conversational data. In a context of media consumption, researchers have built such models to predict which images or videos are more likely to be encoded and retained by a human viewer based on media features and user characteristics [21, 35, 53, 105]. However, the conversational context is different from media consumption: unlike media consumption, conversational context involves continuous production and comprehension of verbal and non-verbal signals, involving different cognitive mechanisms [33] and producing qualitatively and quantitatively different memories [13]. Moreover, conversational memory has various context-specific factors at play that do not apply to media consumption tasks: conversation-specific verbal and non-verbal signals, the relational dynamics between conversational partners, and many more [12, 14–17].

In the field of ubiquitous computing, memory has also emerged as a target for augmentation, scaffolding, and understanding human cognition in-situ. Research on memory aids and digital prosthetics, such as first-person reconstruction tools [106], wearable reminder systems [107, 108], and memory training interventions [109], showcases a growing interest in building systems that understand and extend human memory. Recent datasets, such as LAUREATE [110], have further contributed to this space by enabling predictive models of memory formation from affective and physiological signals. However, this prior work has largely focused on daily tasks or individual experiences, not on the specificity of memory for social interactions. Therefore, the task of computational modelling of how humans encode, retain and retrieve conversations remains unsolved.

Unlike computational research, cognitive scientists have previously explored memory for free-flowing conversations. These studies, often conducted in dyadic settings, have demonstrated that conversational memory is shaped by expectations of recall, egocentric biases, and the collaborative development of common ground [12, 15, 84]. Linguistic features, such as lexical repetition, discourse markers, and syntactic structure, have been found to influence recall probability [17], while recent work also explores memory in group settings, highlighting the distributed and co-constructed nature of recall across participants [28]. To the best of our knowledge, existing studies have not investigated multimodal predictors of memory in free-flowing conversations, such as eye gaze, hand gestures, or body posture, despite the recognised importance of these cues in social interaction [111]. While many of these investigations recorded audio during conversational tasks, these

recordings are, to our understanding, not publicly available for broader research use. Moreover, although the conversations were naturalistic in nature, they typically took place in controlled, face-to-face laboratory settings. Collectively, these studies offer valuable insights into the cognitive mechanisms underpinning conversational memory. However, the data collected were not explicitly intended to support computational modelling or the development of intelligent systems. For example, most works relied on transcripts of the conversations (e.g. [112]), only several works recorded audio of the conversations (e.g. [17, 28]), and no works, to our knowledge, recorded videos of the interactions. This means that the number of features that can be used for real-time prediction of memorability would be limited to speech, with no possibility of using such prominent non-verbal features as facial expressions, eye-gaze or hand-gestures [111]. As such, there remains a clear need for a publicly available, multimodal dataset with recording, aligned memory annotations, designed specifically to advance research in memory augmentation and technologies for meeting facilitation.

Since there has not been any computational research on the topic and cannot be directly used for computational modelling, an essential step towards conversational memory prediction is constructing a dataset of spontaneous conversations annotated with memory reports. Therefore, we constructed the ***MeMo* corpus to provide a resource for the two primary goals**: **(G1)** for research and computational prediction of participants' memory in spontaneous conversations via verbal and non-verbal signals and **(G2)** for the creation of conversational memory models supporting meeting facilitation in the context of repeated interactions.

2.3. GUIDING PRINCIPLES FOR COLLECTING MEMO

When constructing a dataset suitable for studying and modelling human conversational memory, we argue that several major principles need to be considered.

2.3.1. P1: MAXIMISING ECOLOGICAL VALIDITY

While scraping the internet for datasets has become very common in computer science, researchers increasingly advocate for carefully curated datasets to better predict human behaviour in natural settings [31, 32, 113]. This is particularly important when the computational models are aimed to predict and explain human behaviour in a setting with minimal structural constraints, such as free-flowing conversations [113]. In the context of conversational memory modelling, this is particularly important since an unnatural, scripted setting can change the structure and the content of memories [114, 115].

Therefore, we strove to ensure that the *MeMo* corpus accurately reflects the conditions and variables present in real-life conversations.

P1.1 Preserving Natural Interaction Environment. First, the dataset would need to be recorded in a natural environment rather than in a laboratory setting. The laboratory environment can introduce artificial constraints and biases that may not exist in real-life conversational settings, including intrusive sensors and unnatural conversation settings (e.g. a lab with visible sensors and no natural light). It has been shown that people perform differently in memory tasks in a laboratory setting in comparison to a real-world setting [116]. The dataset, therefore, needs to preserve a natural conversational setting typical of real-world environments.

P1.2 Preserving Spontaneity of Conversation Interactions. Second, for conversational memory reports to represent the processes engaged in an in-the-wild conversation or meeting, the conversation must be as spontaneous as possible. The main reason is that the processes involved in comprehension and production of spontaneous speech are different from reading out text or following a script (as in scripted corpora, e.g. [117]). For example, memory for self-produced statements can differ from reading or hearing a statement [15, 33, 118]. Letting the conversation flow emerge by itself rather than imposing lab-created tasks or structure as much as feasible is important for the ecological validity of such conversational data.

P1.3 Ensuring Representativeness of Participants. Lastly, it is important for a dataset to recruit a representative sample of participants from diverse demographics and backgrounds. It is, unfortunately, a common practice in datasets and experimental studies to mainly recruit university students and staff, biasing the data towards a demographic of highly educated English-speaking white young women [119, 120]. An alternative to university students is a more diverse demographic of participants recruited through specialised websites, such as Amazon Mechanical Turk. While this method provides a more diverse demographic, participants recruited this way might be 'professional participants' who go through a multitude of studies daily and are therefore biased in how they respond to the experiment questions [121]. In either case, study results can greatly depend on the demographics of its participants [122], and it is, therefore, important to report the demographics along with the dataset to understand the limitations of the data. So far, unfortunately, it is not a common practice and many datasets do not report the demographics of their participants (see section *Section 2.4.5* for examples).

The *MeMo* corpus aims to support ecological validity through three key design choices. First, conversations were recorded in participants' own homes within the context of online meetings, reflecting a format increasingly common in contemporary professional settings. Recordings

were conducted using Zoom, a widely adopted video-conferencing platform which, at the time of data collection, was estimated to have around 300 million daily active users worldwide [123]. Each conversation had an assigned leader, a moderator, which is also typical in work settings. Second, conversational spontaneity was encouraged by maintaining minimal structural constraints. Although discussions were guided by a moderator and focused on the topic of Covid-19, participants were invited to speak freely, engage with one another, and ask questions, while moderators intervened primarily to maintain the flow and depth of the discussion. Third, efforts were made to include a diverse sample of participants across different age groups and demographic backgrounds (see *Section 2.5.1*). Third, we strive to maximise the diversity of the recruited participants by recruiting participants of different ages and demographics (see *Section 2.5.1*).

2.3.2. P2: MAXIMISING THE CONSTRUCT VALIDITY OF CONVERSATIONAL MEMORY MEASURE

A principal goal of developing MeMo is to collect a corpus that facilitates the identification of moments in a conversation likely to be retained by its participants (→*G2*). Collecting such data is challenging due to the fundamentally different organisation of human experience as context-delineated episodes and the typically timestamp-delineated segments used to annotate moments in multimodal data (see Dudzik et al. [124] for a discussion). Aiming for construct validity involves ensuring that our chosen measures accurately reflect conversational memory while also recognising the limitations of the selected metric.

The existing approach to estimating which events are encoded from a conversation is a free-recall task - asking participants to report what they remember from the conversation in their own words, usually in writing [24]. To then access the events that these reports refer to for analysis, external annotators review the reports and identify the events mentioned within the conversation [17, 112]. Outsourcing this task to external observers may impact the construct validity of the resulting memory measure, as multiple moments might match a description, making it difficult to determine the specific event without asking the participant directly.

Given that the *MeMo* corpus is designed to model conversational memory, we propose an alternative method. After a conversation, participants complete a moment free-recall task, describing what they remember. They then watch a recording and pinpoint the exact moments that match their memories (see *Section 2.5.2* for details). This approach ensures that the identified moments accurately reflect the participants' memories, preserving the validity of the memory annotations. These annotations provide a reliable temporal link between

memory reports and conversation segments, serving as ground-truth labels for computational modelling.

2.3.3. P3: CONSIDERING CONTEXT-SENSITIVITY OF MEMORY PROCESSES

Maximising ecological validity (P1) might imply that fewer variables are controlled (e.g. more diverse demographic, an in-the-wild experimental setting, etc.). Therefore, while maximising the internal validity of the data, it might reduce the external validity [125]. To avoid this, it is particularly important to track as many potential confounding variables as possible using validated questionnaires accepted by the scientific community.

Specifically, conversational memory can be affected by communication skills [16], mood [126, 127], personality [128], values [129] and the relationship dynamics between participants [14, 130]. In addition, factors concerning group perception should be measured, such as group entitativity, cohesion and rapport. While they have not been investigated in relation to conversational memory, the research shows that they can influence learning [131, 132], which is inherently related to memory.

2.4. RELATED WORK

While there are some cognitive studies of memory for free-flowing (mainly dyadic) conversations (including the work by Diachek et al. (2024) that examines linguistic features predicting recall [17]), to our knowledge, there are no multimodal datasets aimed for computational modelling of memory in multiparty conversations. To contextualise *MeMo*, in this section, we describe existing datasets that are aimed to support computational modelling research on memory in one way or another. In addition, we describe most related behavioural studies on the topic of conversational memory. We compare the data designs using the criteria most relevant for the context of the *MeMo* corpus as shown in Table 2.1. We describe each criterion in the following subsections.

2.4.1. FACILITATED MODELLING PERSPECTIVE

When it comes to datasets for modelling memory processes, we distinguish between two different modelling perspectives that they facilitate: Situation-centred and Individual-centred (defined below).

Corpora supporting *Situation-centred* perspectives facilitate modelling how specific properties of a defined situation (e.g., exposure to a video) are expected to give rise to memory responses in members of some population (e.g., how specific video content is likely to be remembered

Table 2.1.: The comparison between MEMO and related corpora [21, 23, 54, 133] and the most related study [17] *(Part. count - number of participants)*)

Dataset/ Study id	Memory sub-process	Memory task	Recorded behaviour	Perceptual measures	Task context	Part. count	Time (h)	Longi-tudinal
Video Mem	encoding & retention	recognition (short & long-term)	-	-	media consumption	3246	19.4	51
EEGMem	encoding	recognition (long-term)	EEG recording	-	media consumption	12	8.3	-
Mementos	retrieval	auto-biographic retrieval	video	personality, mood, affect	media consumption	300	33	-
WoNoWa	retrieval & use in collabora-tion	transactive memory perception	video, audio, transcripts	perceived leadership & group performance	collaboration task (group)	45	17	-
Home Birth helpline	retrieval	spontaneous retrieval in dialogue	transcripts	-	spontaneous conversation (dyad)	56	NA (80 calls)	51
Diachek et al. 2024	encoding & retention	free recall reports	disfluencies from transcripts	-	spontaneous conversation (dyad)	118	14.8	-
MEMO	encoding & retention	free recall reports + timing annotation	video, audio, transcripts	individual, task, group & others' perception	spontaneous conversation (group)	53	31	51

by people in general). Datasets focusing on a situation-centred modelling perspective often attempt to capture a large range of distinct situations but typically have a small number of distinct individuals responding to them (e.g. [21]).

In contrast, datasets with an *Individual-centred* perspective typically focus on more fine-grained modelling of variation in memory processes across specific individuals, possibly considering interactions with the situation (e.g., ways in which individuals behaviourally express when specific video content triggers a memory in them [54]). Datasets focusing on supporting an Individual-centred modelling perspective often contain only a relatively small number of distinct situations but a relatively large number of individuals responding to them. Note that datasets facilitating an Individual-centred perspective can often also support a Situation-centred one, but not the other way around (because responses are typically aggregated from individual responses to the situation level).

In *MeMo*, we aim to combine the two perspectives as much as the conversational context permits. From an individual-centred perspective, *MeMo* includes various perceptual and audio-visual measures from individual participants, providing resources to study how humans behave during memorable moments. From a situation-centred perspective, the data design allows for the investigation of the entire situation using audio-visual data from all group members, along with memory data aggregated across the group. This approach helps identify what makes a moment more memorable for a set of participants, abstracting from individual differences (for an example, see *Section 2.9*).

2.4.2. MEMORY PROCESS

Memory-related datasets vary in their primary goal, specifically the memory sub-process they aim to investigate. Human memory can be divided into three sub-processes: *memory encoding* (processing the experience), *memory retention* (preserving the experience), and *memory retrieval* (extracting the retained experience) [22].

These processes are closely intertwined with each other. For example, any study of memory involves some measure of memory retention - whether or not a memory was preserved and for how long. Whether a moment has been retained cannot be completely measured, since some memories might have been retained but are not accessible at the moment of the measure [134]. Therefore, when it comes to retention, researchers usually focus on investigating memories available for retrieval at the moment of a memory test. Forgetting then refers to the moments that are not available at the moment of collecting the memory measure. Studies that focus on the **retention** process usually investigate the forgetting curve - collecting memory reports at several points in time and seeing how much information will be retained across the time [34]. Two datasets have collected such data in the context of media retention and forgetting [20, 21]. In a conversational context, Diachek et al. (2024) [17] have investigated recall rates in dyadic interactions. Brown-Schmidt et al. (2024) analysed how different conversational roles impact memory retention in group settings, finding that active participants recall more content and source information than passive overhearers [28]. Other behavioural researchers, such as Stafford et al. (1987), have conducted studies on conversational memory retention [84]. While some of these studies recorded audio of the interactions ([17, 28]), to our knowledge, no studies on conversational memory have recorded video of the conversations. Therefore, the number of features that could be usable for memorability modelling would be limited to speech-related features.

Most measures of memory involve memory retrieval - for example, free-recall tasks, which ask participants to freely report what they recall from the given stimulus [24]. This said, most of these studies use retrieval as a memory-measuring tool to study encoding and retention. In contrast, memory **retrieval** dataset papers investigate the moments when memories are (spontaneously) triggered and extracted - whether it is a memory from childhood prompted by music videos [54] or memories relevant to a collaboration task at hand [23]. Several studies have explored how individuals retrieve shared conversational experiences in subsequent interactions. Horton and Gerrig (2005) investigated memory retrieval in conversations by examining how speakers access partner-specific information during language production tasks [135]. Similarly, Clark and Wilkes-Gibbs (1986) examined collaborative processes in establishing mutual understanding during conversations [136]. Shaw et

al. (2007) [133] investigated how callers and call takers indicate prior interactions.

While studies of retrieval focus on the moment a memory is being extracted, memory **encoding** investigation focuses on the specific stimulus or human response to the stimulus at the very moment the memory is being encoded. This means that, while the memory task itself might involve retention and retrieval (e.g. free-recall or recognition [24]), the focus of the study is the specific stimulus viewed or behaviour displayed during the event mentioned in the reported memory. For example, [21, 35, 137] that investigate the features of memorable media or [20] investigating the brain signals at the moment of viewing the media that is to be retained. In the educational context, [110] have collected a dataset for memory augmentation with regular performance tests on study-related recall and understanding. In conversational context, Diachek et al. (2024) [17] have previously investigated how linguistic features predict whether or not a conversational event will be encoded.

The *MeMo* dataset has been designed to study and model two sub-processes of human episodic memory. First, the corpus allows for studying when information has been **encoded** through the investigation of first-party temporal labels of events registered in participants' moment free-recall reports (see *Section 2.5.2*). Second, the corpus is designed to study conversational **retention**, with memory reports collected immediately after interactions for short-term memory and after 3-4 days for longer-term memory (see more details in *Section 2.5.2*; also notice the limitations for this task in *Section 2.11*). In principle, *MeMo* could also be used for modelling memory retrieval processing during conversations (e.g., in terms of how people use memories for social bonding purposes during interactions [11]). However, the annotations provided with the current release do explicitly support this task (see *Section 2.11* for a discussion). This said, the three subprocesses cannot be completely separated from each other and focusing on one subprocess (e.g. encoding) does not mean that there are no elements of other subprocesses involved in the study. For example, while focusing on memory encoding, we still use annotations based on the memory recall self-reports, which have been retained in participants' memory and retrieved during the recall task.

2.4.3. RECORDED BEHAVIOUR & MEASURES

Some datasets for memory research involve recording **participants' behaviour** and/or **perceptual measures** (i.e. self-reported individual traits or questionnaires on participants' perception of the task and other participants). These measures can vary from EEG brain signal recordings during the task [20] or wearable physiological monitoring

data [110] to video of participants' non-verbal signals throughout the task (in [54]). In datasets involving conversation or interaction between participants, typically, there is also a recording of the speech, in the form of audio or transcripts [23, 133]. In addition to recorded behaviour, some datasets measure various self-reported perceptual measures [23, 54, 110]. Since the *MeMo* dataset is focused on human behaviour and perception in the conversational context, it includes video and audio recordings for behavioural measures and various self-reported measures of participants' individual characteristics, their perception of the interaction, group, and other participants.

2.4.4. TASK CONTEXT & MEMORY TASK

An important parameter informing corpus design is the context of the task participants perform, such as solitary media consumption or human-human interaction. This context is closely linked with the memory reporting measure. For media consumption, memory is often measured with recognition tasks where participants identify previously seen videos, indicating memory retention or forgetting [20, 21, 35]. Alternatively, some studies use free recall for autobiographical memory triggered by media [54] or study-related performance tests [110].

In spontaneous conversation contexts, recognition tasks are impractical due to the variable content of free-flowing conversations. Instead, memory encoding and retention studies find free-recall self-reports to be more suitable as they allow participants to report memories without additional constraints [13, 17]. For example, these have been used in a recent behavioural study predicting memory encoding using linguistic features [17]. For conversational memory retrieval datasets, other measures have been employed: a task-related memories survey in [23] and a third-party observer annotation of memory retrieval moments in [133]).

2.4.5. SAMPLE REPRESENTATION

Lastly, participant samples vary in size and diversity. The sample size depends on the task context and length, ranging from 12 to 3246 subjects (*Table 2.1*). Considering demographics and representativeness of datasets' samples, three datasets do not report participants' demographics [20, 21, 35], one dataset had students and university staff as participants [23] and one had female-only participants [133]. Only two out of seven memory-related datasets had a more balanced sample (except for the bias towards US residents) [17, 54].

2.4.6. DATA QUANTITY

The dataset comprises over 30 hours of annotated multiparty dialogue, which is comparable in scale to other multimodal corpora used for affective computing and cognitive state modelling. For example, WoNoWa [23] includes around 17 hours of recordings of collaborative interactions. For memory modelling outside social context (i.e. life-logging), the "Naturalistic Free Recall" dataset includes approximately 38 hours of data [138]. As summarised in Table 2.1, MeMo falls within the typical range for datasets designed for multimodal behavioural analysis and modelling.

2.5. METHOD: DATA COLLECTION PROCEDURE

We describe the *MeMo* experimental procedure shown in *Figure 2.1* in this section. We highlight how these relate to our stated primary goals (*G1* and *G2* described in *Section 2.2*) and guiding principles (*P1* to *P3* described in *Section 2.3*).

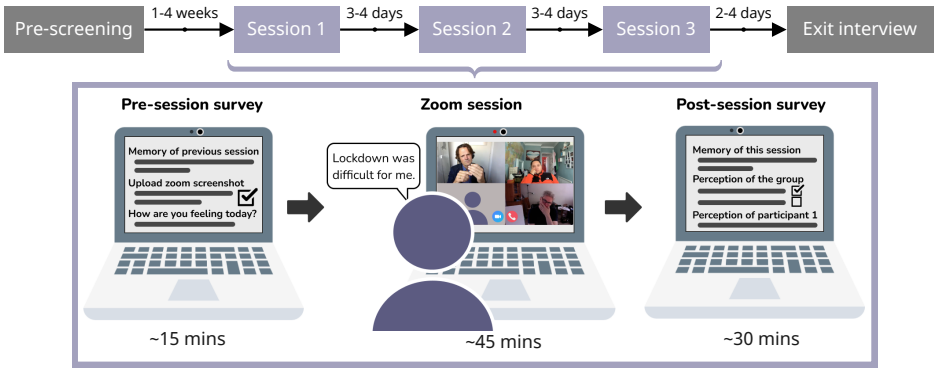


Figure 2.1.: Experimental set-up. Upper flowchart - overall set-up. In the lower part - illustration of the procedure for every group session. On the second screen, there is a screenshot of a discussion from the MeMo corpus, except for the 4th participant, shown with a person icon. The phrase the icon person produces is made-up but matches the conversations included in MeMo.

2.5.1. OVERALL PROCEDURE

The overall procedure of acquiring the *MeMo* corpus is shown in *Figure 2.1*. In this subsection, we will guide the reader through the procedure step-by-step.

ETHICAL APPROVAL.

The Human Research Ethics Committee of TU Delft approved the *MeMo* corpus data collection. Before the experiment, participants filled out an informed consent form permitting us to collect their personally identifiable data, such as audio and video recordings, to be later accessible to the research community under CC-by-NC license. Before the recordings, participants were asked to come up with a pseudonym for themselves for the entirety of the recording. Participants were allowed to avoid answering questions if they did not want to and invent information about themselves throughout the recording.

CONVERSATION SUBJECTS.

Participants. Participants were recruited using Prolific Academic recruiting service [139]. All participants were required to reside in the UK, fluently speak English and be ready for a video-call study (i.e. possessing a laptop with a working camera and a headset with a microphone). The UK residency requirement served two purposes: first, the Prolific platform provided a large number of UK participants; second, research shows memory can be influenced by shared experiences [12]. With Covid-19 selected as a discussion topic, focusing on UK residents helped control for differences in pandemic experiences, thus enhancing the validity of memory comparisons between groups (→ *G1*).

To simulate a situation where facilitation is needed, such as in the case of differing opinions and perspectives in the group, we tried to maximise the diversity in opinions on our target topic - Covid-19 pandemic (*G2*). We, thus, targeted specific demographics differently affected by the pandemic. For example, parents of young children had to suddenly homeschool their children or business owners faced work-related challenges, having to adjust their business to changing regulations. The targeted groups were the following: parents with young children, older adults (50+), students, business owners. Five prescreening surveys were conducted on Prolific, each tailored to one of these groups (see *Appendix A.3*). Participants could only select one prescreening survey to avoid duplicate inclusion. Additionally, we ensured gender balance in our sample (→ *P1.3*).

Group composition. The *MeMo* corpus is designed around small-group discussions, typical of both work and informal settings, to ensure ecological validity (→*P1.1*) and simulate potential facilitation scenarios (→*G2*). "Small groups" here refers to 3 to 8 participants, an optimal size for allowing everyone to share their thoughts [140]. Participants completed a prescreening survey specifying their availability, and 5 to 8 participants were recruited per group, ensuring representation from each demographic. The minimum of 5 was set to maintain at least 3 participants per group, accounting for no-shows and dropouts.

All participants were **zero acquaintance**, meaning they had never met before, allowing us to study the development of within-group relationships with no prior interactions [14]. The group composition stayed the same throughout the experiment (except for having fewer members in later sessions if a participant dropped out after the first or the second sessions). Each participant took part in only one group, to avoid participants confusing memories from interactions with different groups.

Moderators. For facilitation purposes (→*G2*), discussions were guided by professional moderators who ensured a safe and inclusive environment, encouraging free-flowing conversation with spontaneous turn-taking (→*P1.2*). Moderators were confederates aware of data collection goals and were allowed to use any methods of their liking for the facilitation of a free-flowing conversation. Moderators were not familiar with any participants before the experiment. Each group was assigned one moderator for the entirety of the experiment (3 sessions and an exit interview). Along with guiding the sessions, moderators had to fill out the same surveys as participants before and after each session, including memory reports and all other measures.

PRE-SCREENING SURVEY.

The prescreening questionnaire included the consent form, participants' demographics (participants' age, gender, employment status, English fluency, country of residence), personality [141], values [142], technical requirements and online meeting experience (see *Appendix A.1*). Personality was included as it influences recall, with extroverts recalling more positive memories than those higher in neuroticism [128]. Values were also included, as they can enhance recall accuracy for items related to personal values [129].

PRE-SESSION SURVEY.

Questionnaire data was collected using Qualtrics X platform [143] (for the full content of the questionnaires see *Appendix A.1*). Participants started each session by completing a pre-session questionnaire that took ∼15 minutes to complete. The pre-session survey included the participant's mood assessment [144] before all sessions, as mood can affect memory encoding [126, 127]. It also included long-term memory retention task (see *Section 2.5.2*) before all the sessions except for the first one. Before the exit interview, there were some extra questions added to the pre-session survey - participants had to report a moment they found most important in all the past interactions and provide feedback on the moderator facilitation skills (see *Section 2.5.1* for more details). At the end of the survey, participants joined a scheduled Zoom link for the discussion session. They completed this survey by uploading

a screenshot of their Zoom layout when all the participants were present in the Zoom session to facilitate eye-gaze target extraction (see *Section 2.6.4*).

CONVERSATION SESSION.

All discussion sessions happened online, through a Zoom video-call platform, a typical software used for video-calls. This ensured that participants are in the comfort of their own homes, rather than in the lab (→*P1.1*). They used their own computer and headsets, having their natural lighting, which also added to how comfortable they felt, as well as the naturalistic setting of video-call discussions. To ensure that we can track participants' eye-gaze, the moderator asked participants to keep Zoom in 'gallery' mode so that all the participants are on the screen at the same time and their location on the screen stays the same throughout each session. In addition, for a secure recording, there was a technical assistant involved in the call (with no camera or microphone on and no interaction with participants), who recorded the session and resolved any arising technical issues.

After making sure that all participants had completed the pre-session survey, the moderator (or the technical assistant) would start the recording. The session began with head pose and gaze calibration for automatic post-experiment gaze direction annotation (see *Section 2.6.4*). Participants, guided by the moderator, first rotated their heads and then looked at each named participant on their screen.

After that, guided by their moderator, participants started the discussion. To ensure a natural yet directed conversation (→*P1.2*), the Covid-19 pandemic was chosen as a relatable topic, relevant to participants worldwide at the time of data collection (the year 2021), with diverse experiences and opinions. At the start of the first session, participants were informed that they would discuss the past, present, and future of the pandemic over three sessions, aiming to design a better future in case of a recurrence (the memory study focus was disclosed only at the experiment's end to prevent priming participants' memory). Each session lasted 45 minutes, providing sufficient time for conversation to emerge. This duration aligns with typical meeting and facilitation session lengths (→*G2*, →*P1.1*) [93].

To reflect real-world conversations that repeat over time (e.g. work meetings), the corpus included 3 sessions spread out over 3-4 days, reflecting the frequency of real-world facilitation sessions, occurring once or twice a week [93]. This longitudinal approach aimed to capture the evolution of participant relationships and conversational memory trends over time. (→*G1*,→*G2*). This setup also provided repeated measures of memory at different points, capturing both short-term and long-term memory reports (→*G2*).

POST-SESSION SURVEY.

At the end of the Zoom session, the moderator reminded participants to open up the post-session survey link and start the questionnaire. After that, the recording stopped and the Zoom session was closed. See the summary of all measures used in the post-session survey *Appendix A.1*. The post-session survey started with moment free-recall self-reports (described in *Section 2.5.2*) and a qualitative question for facilitation application (*Section 2.5.2*). Since interpersonal skills can be of effect [16], participants' communication skills were evaluated by having participants rate each other's listening and conversational abilities with 2 one-item questions (see *Appendix A.2* for question formulations). Several scales were collected to measure relational growth and mutual understanding for the ultimate facilitation application of the *MeMo* corpus (→*G2*). Relational development was measured using the IOS scale [145] to assess perceived closeness and a single-item scale was used to assess personal attitude (see *Appendix A.2*). Mutual understanding was assessed by comparing participants' pre-reported values with others' post-session evaluations using the Short Schwartz's Value Survey [142]. In addition, group perception was tracked with measures of cohesion [146], entitativity [147], perceived interdependence [148], situational characteristics [149], syncness, and rapport, as these factors can influence learning and group performance [131, 132].The post-session survey finished with encoded event annotation (*Section 2.5.2*), and reasons for remembering (*Section 2.5.2*). After submitting the post-questionnaire, the participants had to wait 3-4 days for the next scheduled session (or exit interview in case of the 3rd session), and then repeat the procedure (pre-session survey → conversation session → post-session survey).

EXIT INTERVIEW.

3-4 days after the final conversation session, there was a ∼15-minute exit interview with each participant. Similar to the conversation sessions, there was a pre-session survey before this Zoom session. The only difference was that participants were asked what was the most important moment for them in all the previous discussion sessions. They then discussed that moment with the moderator one-on-one and answered a list of questions on the topic of the required capabilities of a social robot supporting public discussions, especially about what such a robot should remember.

2.5.2. MEMORY MEASURES

A common practice in user internal state modelling is to rely on third-party annotations of the investigated internal state. This can oversimplify these states and skew model accuracy, neglecting the

first-party perspective and potentially introducing bias. This challenge applies to memory encoding studies, in which, traditionally, the free-recall reports are traced back to the encoded event by third-party annotators [17,112]. Here, we describe our method that mitigates these issues by leveraging participants' self-reports and the first-party task of aligning the reports with specific segments of recorded interactions. This method is aimed to preserve validity and minimise bias, particularly crucial for datasets like *MeMo* corpus aiming to accurately represent the memory content as well as the event that that memory might be based on within the recorded data (→*P2*).

MOMENT FREE-RECALL SELF-REPORTS.

To minimise the bias towards external stimuli and the type of memory events, we have used a task inspired by the traditional free-recall task for memory self-reports [24]. The traditional free-recall paradigm asks participants to remember and recount as much as they can from what was said or occurred during the conversation. In contrast, we have asked participants to recall as many conversational segments as possible, so that each part of the free-recall self-report can be related to a specific event that occurred within the conversation, we call those events 'moments'. This modification was done to ensure that we can then ask participants to complete a first-party annotation of the encoded events (described in Section 2.5.2).

The moment free-recall task was the first task in the post-session questionnaire, completed immediately after the end of the session, to avoid interventions of any additional bias that could modify the memory. The task formulation was open-ended to account for any conversational events recalled (spoken information, participants' feelings, context details etc., see the exact question formulation in *Appendix A.2*). Participants were meant to report a memorised 'moment' in each field in their own words without a word limit. They could report from 3 to 10 moments. Participants could move to the next survey questions only if they did not remember more moments or if they had already reported 10 moments. The maximum of 10 moments was set to avoid fatigue and leave time for answering the next survey questions. This way, we tried to capture all the retained and currently accessible [134] memories (unless there were more than 10 moments to report). The idea of having participants report memory 'moments' aimed to capture the content of the memories as well as the way participants conceptualise the continuous stream of perceived and self-produced social signals into specific memorable events [150].

ENCODED EVENT ANNOTATION.

Unlike in previous research [17,112], the participants themselves did the assignment of their self-reports to specific events that happened within the conversation. This way, we wanted to ensure that the self-reported memories were correctly assigned to the encoded event they referred to, ensuring the validity and the accuracy of the resulting memory measure (→*P2*). Within this encoded event annotation task, participants were given a link to the interaction recording and had to write down the start and end time (minute and second) of each moment they reported in the survey before (the memory self-reports were quoted back to them). Participants had the freedom of scrolling through the video and did not have to re-watch the entire recording to avoid additional fatigue. They also had an option of leaving the timing blank in case they could not find it, the moment was related to an overall feeling of discussion or other kind of memorable moments that cannot be connected to a particular interval in the interaction. We ensured that the free recall reports cannot be modified at this stage (→*P2*). The encoded event annotation task was presented to participants at the very end of the post-session questionnaire so that the other survey questions would not be affected by seeing the video of the interaction either.

REASONS FOR REMEMBERING.

In addition to moment free-recall reports, we asked participants about the perceived reason or motivation behind their memory of each reported moment (see question formulation in *Appendix A.2*). The perceived reasons for memory were aimed to capture information that might not have been described in the moment free-recall reports - about the underlying personal significance of the specific moment and the underlying thought process as opposed to details of the event itself from the moment free-recall reports (e.g. if the memory report is "I remember participant 3 said that they suffered from the lockdown", the reason could be "I remembered this moment because I also found it very difficult"). This information helps uncover intrinsic motivations for memory, useful for qualitative analysis and understanding in meeting facilitation (→*G2*). Participants could provide as much detail as they wanted, and the question was placed at the end of the questionnaire to avoid biasing their recall.

LONG-TERM MEMORY RETENTION.

Apart from the moment free-recall task immediately after the interaction (see above), the *MeMo* corpus also included the measure of long-term retention. This measure was collected to investigate what kind of memories stay after the interaction and which memories are more likely to be forgotten (or less accessible for retrieval) (→*G1*). To

assess long-term memory retention, participants returned 3-4 days after the interaction, just before the next session, to answer the moment free-recall questions about the previous session. This interval was chosen because most forgetting occurs within this timeframe, after which memory stabilises, as shown by Ebbinghaus (1880) and later confirmed by Murre et al. (2015) [34]. Therefore, the task meant to capture a stable representation of what participants would remember in the long-term. The long-term memory question was exactly the same as the post-session moment free-recall task (see above). Similar to the main moment free-recall task, participants could report from 3 to 10 moments in text description fields with no word limit. Unlike the short-term annotations, this time the participants did not have to re-watch the video and map the timing to each moment, to avoid excessive fatigue before the conversation session.

QUALITATIVE DATA FOR FACILITATION APPLICATION.

Since the models trained on the *MeMo* corpus are aimed to be applied to automatic meeting facilitation (→*G2*), there was one qualitative question about what participants would want such a system to recall in the next sessions (for the task formulation see *Appendix A.2*).

2.6. METHOD: DATASET PROCESSING AND CURATION

In the following, we describe distinct processing and filtering steps that we have applied to the raw dataset, resulting eventually in a curated version for the purpose of analysis and eventual sharing with the research community (see *Section 2.12* for more details).

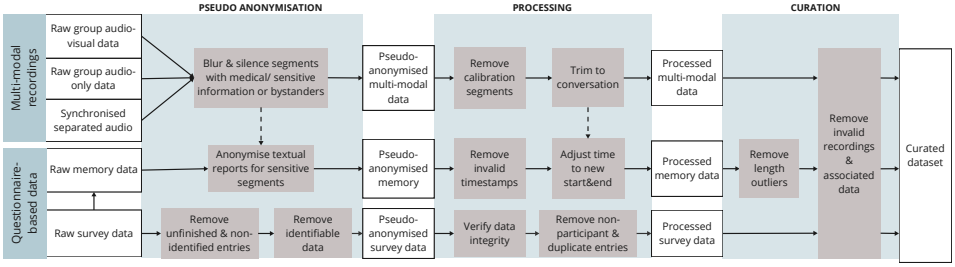


Figure 2.2.: MeMo corpus processing and curation steps

2.6.1. PSEUDO-ANONYMISATION

To maintain the privacy of participants and compliance with ethical guidelines, the dataset is being reviewed and processed to remove

potentially problematic segments, resulting in a pseudo-anonymised intermediate version, see *Figure 2.2*.

Multi-modal data pseudo-anonymisation. The raw multi-modal data recorded throughout the data collection consisted of group audio-visual recording, a group audio-only file in better quality, and separated audio channels per participant, automatically recorded through the Zoom software [151]. Because of a bug in the Zoom software, the separated audio tracks were not aligned with the video recording. They, therefore, were first synchronised to align with the group audio-visual and audio-only recordings using the procedure described in *Appendix A.4*.

Two types of data needed to be removed from the multi-modal recordings for privacy concerns. First, since the corpus contained discussions on the topic of Covid-19, the participants sometimes mentioned medical information about themselves or people close to them. Second, because the recording was in a natural environment of participants' homes, sometimes there were bystanders passing by or talking in the background. Since they did not consent to be recorded, visible bystander faces and audible decipherable bystander speech needed to be removed. While we removed these types of data, the pseudo-anonymisation process is yet to go another stage of sensitive data removal before it will be shared (see the data release statement in *Section 2.12*).

Questionnaire data pseudo-anonymisation. In regard to questionnaire data, the unfinished and non-identified entries were removed to only contain entries from identified paid participants. Identifiable data associated with the data collected through the Qualtrics survey platform was then removed. This included such data as IP address, location and signatures. The Prolific IDs were replaced with a non-identifiable hash number, since they otherwise could be tracked to a specific account on the Prolific recruitment platform. After extracting participants' on-screen location via screenshots for further eye-tracking, the screenshots were also removed from the data since they were taken before the recording and, therefore, sometimes contained sensitive information, such as participants' real names.

Memory data pseudo-anonymisation. Memory data extracted from the questionnaire data was pseudo-anonymised in accordance with the segments removed in multi-modal recording pseudo-anonymisation. For these sensitive segments, the textual reports of the corresponding memorable events were manually edited to avoid direct references and descriptions of the sensitive information. No free recall reports were removed, but the references to the sensitive information were replaced with "[anonymised]".

2.6.2. DATA PROCESSING

As shown in *Figure 2.2*, we have performed additional processing steps to improve the usability of the data.

Multi-modal data processing. We processed the multi-modal data to keep only the conversation content, removing any technical or organisational parts from the recordings. Specifically, calibration segments for eye-gaze and head-pose, included at the beginning of each recording, were removed. We also trimmed the start and end of each recording to ensure they captured only the actual conversation time, excluding moments where moderators discussed technical or scheduling issues with participants or waited for participants to complete questionnaires.

Questionnaire data processing. The questionnaire data underwent the verification of the integrity process. This constituted manually verifying the correspondence of session, group and alias information to the questionnaire data, since sometimes participants made mistakes in those fields. These were manually verified using the available information, such as the date and time of questionnaire completion, the missing participants and the fields that were filled in correctly. In this processing step, only data for conversation participants was maintained, excluding participants who did not show up to any conversation sessions. The duplicate entries were also removed at this stage. Since sometimes participants forgot to fill in the surveys, there are some gaps in the data. To identify the consistency of the available data across groups, we computed a ratio of participants that completed all surveys (both pre- and post- in all sessions). We called that measure "questionnaire completeness" (see full table for all the groups in *Appendix A.5*). Overall, 13 out of 15 groups had questionnaire completeness at 50% and above. We leave the decision of whether to discard some of the groups using this measure up to the future users of the corpus, since this might not be important for researchers not focusing on the longitudinal component of the corpus.

Memory data processing. We excluded the memory moments with invalid timestamps: if the reported start was further than the reported end of a memory event and if the reported timestamps were outside the duration of the recording. Since the multi-modal data processing included shifting the start and the end of the recordings, the encoded event annotations needed to be adjusted to the new start and end of the recordings, to maintain the references to the originally tagged events.

2.6.3. DATA CURATION

In addition to the necessary data processing, we have curated the processed data to provide the cleanest version of the dataset, which we recommend for further use (see *Section 2.12* on how this data will be

released).

Audiovisual data curation. Within the curation, we have removed a video that was recorded in 'speaker view', with one active speaker in the camera view at a time. All other videos were recorded in 'gallery mode', meaning that all participants were visible at all times. Given these criteria, 1 full session was removed.

Questionnaire data curation. The questionnaire data remained unchanged except for removing the data associated with the removed session within audiovisual curation.

Memory data curation. Apart from removing memory data associated with the removed videos, within the curation, we have removed some memory event outliers. Specifically, we have removed memory moments outliers by duration: remembered events that lasted more than 1 standard deviation away from the mean duration over all the data (longer than 690 seconds). These moments were removed since they did not have enough detail or contained an overall feeling over the discussion rather than a specific event, in case of duration outliers (e.g. a moment with memory report "The agreement on the uncertainty of following rules and what rules were correct to follow etc." lasting 34 minutes). We also considered removing events shorter than 5 words in the moment free-recall memory description, but removing the length outliers also removed reports shorter than 5 words, since the duration of the associated event was always above the outlier threshold.

2.6.4. EXTRACTING MULTIMODAL FEATURES

In addition to dataset processing and curation, we have extracted various features from the multi-modal data that can be useful for machine learning tasks. Where feasible, we have attempted to quantify the quality of the extracted features through the analysis of their temporal stability. In particular, we focus on the features in which one would expect to see smooth and moderately paced motions, such as the hand and body motion and facial action unit activations.

Transcription. We diarised and transcribed the recorded audio of the discussions using the Kaldi Speech Recognition Toolkit [152]. We subsequently conducted manual reviews and corrections to the resulting transcripts where necessary. The timestamps for these transcripts are available at both the utterance and word levels, and we provide word-level transcriptions for each recording.

Eye gaze. In our first corpus-based study, we used GazeSense software [153] to estimate participants' gaze direction throughout the session. We created a customised grid for each participant, matching their Zoom gallery layout based on a screenshot they provided at the session's start. Each session began with a calibration where participants focused on specific screen segments. We used the coordinates of these

segments as calibration points and then obtained gaze estimates for all frames beyond the final calibration. Due to recording imperfections and challenges with some participants' uploaded screenshots, gaze tracking for 13 participants was not possible. Therefore, our dataset includes eye gaze data from 40 participants from 14 groups, which collectively spanned approximately 23 hours of video recordings. This annotation includes gaze targets derived from screenshots capturing participants' screen views.

Prosody. We extracted the eGeMAPS feature set from the default eGeMAPS configuration in the OpenSmile software for prosody analysis [154].

Body Pose. Body and hand keypoints were extracted using MediaPipe [155]. Keypoint prediction was evaluated on cropped segments of the original video to ensure only a single person was visible to eliminate the need for keypoint tracking. The largest available model was used and the confidence threshold for retaining the predicted keypoints was set to 0.5. For each keypoint set, we calculated the frame-by-frame motion measuring the Euclidean distance between corresponding keypoints in consecutive frames. Using these, we compute the average displacement. For the hand keypoints, this was 0.04 and 0.05 for the body keypoints. This suggests that large-scale noise in the prediction is not present in the extracted keypoints.

Facial Action Units. Facial action units and face keypoints for participants were estimated using the OpenFace Software [156]. As with the body pose estimation, these features were extracted using the cropped segments to ensure only a single face was present in the video. To evaluate the action unit (AU) detection quality, we analysed the temporal consistency of the AU presence by calculating the change rate for each detected action unit (see *Table 2.2*). Overall, we find that the average change rate across all AUs is 0.0098, indicating that the presence prediction persists on average for approximately 100 frames. Additionally, we observe that AU9 was the most stable (0.0014) AU15 exhibited the largest change rate (0.17). AUs related to eye movements and more subtle mouth movements typically demonstrated higher change rates. AU45 (blink) had one of the larger change rates (0.015), although we note that this may not be an issue with the software detection, but rather a natural consequence of blinking behaviour. However, the larger change rates observed for AU15 (lip corner depressor), AU25 (lips part) (0.013) and AU28 (lip suck) (0.012) are likely due to challenges in detecting more subtle or rapid movements in the mouth region.

Table 2.2.: Corpus wide average change rates for action unit presence extracted using OpenFace

Action Unit	Description	Change Rate
AU09	Nose Wrinkler	0.00145
AU06	Cheek Raiser	0.00302
AU01	Inner Brow Raiser	0.00422
AU12	Lip Corner Puller	0.00467
AU02	Outer Brow Raiser	0.00571
AU20	Lip Stretcher	0.00762
AU10	Upper Lip Raiser	0.00680
AU14	Dimpler	0.00865
AU26	Jaw Drop	0.01063
AU04	Brow Lowerer	0.01080
AU23	Lip Tightener	0.01181
AU28	Lip Suck	0.01254
AU07	Lid Tightener	0.01270
AU17	Chin Raiser	0.01234
AU25	Lips Part	0.01320
AU45	Blink	0.01542
AU05	Upper Lid Raiser	0.01685
AU15	Lip Corner Depressor	0.01729

2.7. DATASET CONTENTS

2.7.1. DATASET SUBJECTS

Conversation participants. There were 53 participants in the experiment. The demographics of the resulting sample are shown in *Table 2.3*. The sample was balanced across genders (28 F, 25 M), included participants of various age groups (from 18 to 76 y.o, see Figure 2.3), and employment statuses (see *Table 2.3*). The participants all spoke fluent English and were UK residents. Maximising the diversity of opinions, we have recruited demographic groups that were differently affected by the pandemic (see specific criteria and the resulting selection in *Appendix A.3*).

Conversation groups. The curated dataset sample contained 15 groups. The groups contained 4 participants on average, with a minimum of 2 (in two sessions where the third participant dropped out) and a maximum of 5 (see *Table 2.4*). Overall, 49 participants took part in all three sessions, and 4 participated in two sessions, skipping one session for unforeseen circumstances.

Conversation moderators Four moderators (3 M, 1 F; 24-45 y.o.) were recruited to facilitate the conversation sessions. All moderators had 2 or more years of professional facilitation experience. Three

Table 2.3.: Participants' demographics

		Mean +- SD / Male	Min / Female	Max
Age (y.o)		38.8 +- 15.1	18	76
		Male	**Female**	
Demographics	**Full-time employed**	11	11	
	Unemployed	10	3	
	Part-time employed	7	5	
	Business owners	2	8	
	Students	6	7	
	Parents of young children	3	6	
First Language	**English**	23	23	
	Other	2	5	
Total		25	28	

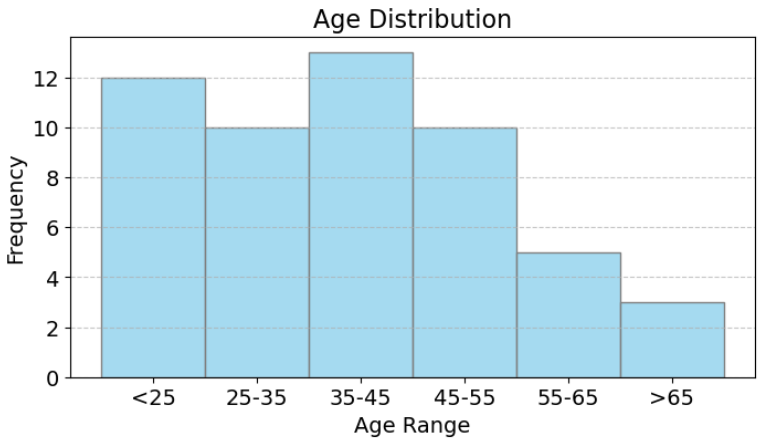


Figure 2.3.: Participants' age distribution

moderators facilitated 3 groups each, and one facilitated 6 groups.

Technical support. It is important to note that apart from group participants and the moderator, within each Zoom session, there was a technical assistant who kept the camera and the microphone off throughout the entire experiment, they were recording the session, and the moderator could communicate with them in case of any technical issues in a private chat. There was no communication between the technical assistant and the participants.

2.7.2. MULTI-MODAL RECORDINGS

The curated dataset contains 31 hours (111674 seconds) of group audio-visual recordings. *Table 2.4* summarises the curated distribution of conversation sessions (the calibration timing is excluded from all the shown durations). Since there were 15 groups with 3 sessions each, a total of 45 sessions were recorded. The average duration of a conversational session was 42 minutes (2538 sec). This said, the duration of the conversation differed across sessions, with the first session being the shortest (35 minutes or on average) and the third being the longest (46 minutes on average, see *Table 2.4* for details). This difference is connected to the fact that, by the 3rd session, participants needed less explanation and had fewer technical issues than in the first and the second sessions.

Video. The videos were recorded with Zoom local recording. The video files have a sample rate of 32000 Hz with 32 bits per sample. The resolution is 1280x720, with a frame rate of 25. The duration of the video recording corresponds to the overall duration of the conversation session: 31. There are 45 videos in total - a video per session. 43 recordings were recorded in the 'gallery view' of Zoom with all participants on the screen, 2 recordings had some technical issues and were recorded in the 'speaker view' and therefore might need to be excluded from video analysis.

Audio: Full and separated. Each video is accompanied by a separated full audio in .m4a format. This audio includes the full audio from the above video automatically recorded through Zoom. This audio is available for all 45 sessions. In addition to the full audio, the dataset includes automatically recorded audio channels per participant available for 42 out of 45 recorded sessions.

Qualitative variations in multi-modal recordings. Since the dataset was recorded in the natural environment of Zoom conversations, the video data has some qualitative variation. This applies to the position the participants were in throughout the recording: while most participants were seated behind their desks, some participants were seated with a laptop on their lap and, in rare cases, a participant was lying down throughout the recording. Another variation was the participants' location of taking the video call - while most participants were in the comfort of a home, one participant was seated outside, and two were taking a call from their car. In addition, although we asked the participants to keep their background as it is, some participants had a blurred background setting and, in rare cases, they had a virtual background. Another variation was the device used for the call - although a laptop was required, some participants joined from their tablet or a phone because of technical issues with their laptops (to our knowledge, this happened in 3 sessions). The final variation to possibly consider is the fact that participants used their own technical setup, with

Table 2.4.: Descriptive statistics of participants' age, group size, conversation duration and memory reports

		M	SD	Min	Max
Group size		3.7	0.8	2	5
Conversation duration (sec)	**Session 1**	2156	377	1290	2775
	Session 2	2671	314	2160	3315
	Session 3	2761	349	1950	3260
	All sessions	2538	431	1290	3315
Memory	**Moment duration (sec)**	141	183	1	1260
	Moment count per person	3.9	1.4	1	10
	Word count per moment	32	21	5	117

different quality microphones, cameras and internet connection. For the same reason, the lighting conditions might vary across participants.

2.7.3. QUESTIONNAIRE-BASED DATA

LONGITUDINAL COMPLETENESS OF QUESTIONNAIRE DATA

Although participants were asked to complete a questionnaire before and after each conversation session, there are some gaps in the data where participants could not complete questionnaires due to unforeseen circumstances. To facilitate the selection of the data for longitudinal analysis, we have computed the continuity of the questionnaire data in each group. Apart from the two groups, every group had 50% or more participants with complete pre- and post-questionnaires in all the sessions. We share the questionnaire completeness scores for each session for an easier subset selection process in future research (see table in *Appendix A.5* for the full table and measure description).

MEMORY DATA

The descriptive statistics of the curated memory annotation are shown in *Table 2.4*. Overall, the curated memory data included 602 moments reported in participants' free-recall tasks. In the context of the MeMo corpus, a moment refers to a subjectively meaningful episode that a participant recalls from a conversation. These moments are not predefined or uniformly segmented, but are self-identified by participants during the moment free-recall task, reflecting their personal memory of the interaction (see Sections 2.5.2 and 2.5.2 for details on how these labels were collected). As such, they

are grounded in participants' subjective interpretations of what stood out to them, rather than objective information conveyed during the interaction. This means that moments may vary in descriptive detail and perceived duration, and may overlap, either because different participants remembered similar content in different ways, or because a single participant described multiple, partially overlapping segments as distinct memorable experiences. As such, 44 instances out of 602 instances involved multiple participants recalling the same event (start and end times aligned with at least 1 other moment start and end times ± standard deviation over the length of all the moments).

The free recall description varied in length, with an average of 32 words per reported moment. The mean duration of the self-reported memorable moments was 141 seconds (2.35 min), with a minimum of 1 second and a maximum of 1260 seconds. Participants remembered ∼4 moments on average (SD=1.4), with 1 as a minimum (in cases where participants didn't remember more or reported non-temporally attached moments) and 10 as a maximum.

Figure 2.4 shows examples of memorable moments variable in report length and moment duration. For instance, a participant recalled the moment in which participant 2 shared their opinions on the ideal future: "P2 wanted a more equal world in the future, more gender equality, racial equality and less of a financial divide between the rich and the poor". There was a variability in memory description word count as well as the reported moment duration. The memory description word count does not correlate with the duration of the moment annotated in the video (Spearman R=0.04 and p>0.05). This can be illustrated with the moment associated with the green point (top report) in *Figure 2.4*, with a long description of a memory associated with a moment that lasted for 4 seconds in the video recording.

During the data processing and curation 235 moments were removed (from the initial 853 total reported moments in the raw dataset), these mainly include invalid entries, but probably also fake memories that participants reported but could not find in the recording during the encoded event annotation task (see Section 2.5.2 for task description and Section 2.6). This said, we did not collect any measures to track which memories were fake, which ones were real but not found by the participants in the recording and which ones were simply invalid entries. The memorable moments data was not meant to be used for calculating recall rates (e.g. as computed in psychological research, such as [17,84]) or forgotten memories, but rather the encoded memories accessible at the time of the free-recall task. This is due to the primary goals of the corpus (see *G2*) being for memorability prediction for meeting support applications. In these technologies, continuous memory prediction is most beneficial, since it can help participants or moderator of the meeting while it is still happening. In contrast to memory encoding

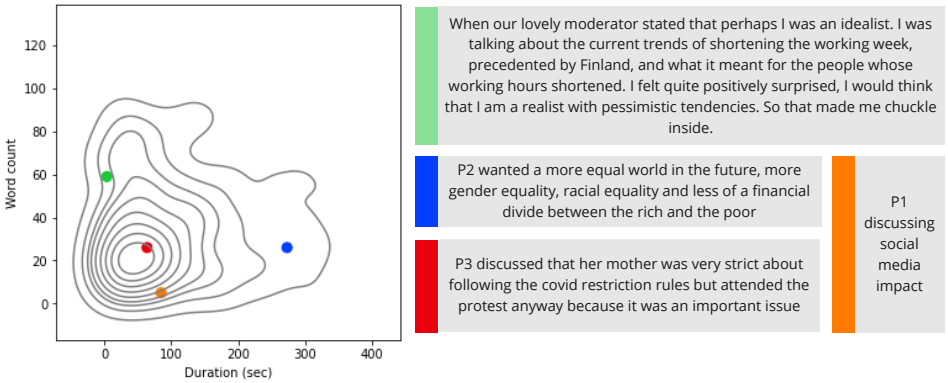


Figure 2.4.: Duration and word-count distribution of memorable moments with examples of memorable moments shown with points and full reports on the right.

annotation, as used in this corpus, recall rates usually provide one measure of recall quality for the entire session and therefore have not been previously operationalised as continuous.

OTHER QUESTIONNAIRE MEASURES

Regarding other questionnaire measures, the collected self-assessment measures show high variability. For instance, participants' mood was measured before each session with AffectButton [144], assessing participants' pleasure, arousal and dominance from -1 to 1. The mood dimensions had a mean of 0.35 ± 0.38 for Pleasure, -0.14 ± 0.77 for Arousal and 0.21 ± 0.57 for Dominance across all participants and sessions. In other words, participants' moods were generally mildly positive, less aroused than average and slightly more dominant than average. The low arousal might be connected to the general setting of at-home video calls, illustrating the informal in-the-wild setting of the dataset (in accordance with P1.1, *Section 2.3*). This said, the variation in participants' mood was quite high, as shown by high standard deviations across all scales. This might indicate that the dataset might be usable for analysis of conversational memory in connection to the mood at the start of the conversation.

Other self-reported measures, such as situation, group and relationship perception, also show high variability. See *Section 2.8.1* for more details on these measures.

2.8. VALIDATION: USING MEMO CORPUS FOR COMPUTATIONAL MODELLING

2.8.1. DEPENDENCY ANALYSES

TEMPORAL DEPENDENCY

Conversational data intrinsically contains temporal dependencies. These and time-related biases related to human cognition have to be accounted for in the design of the computational models.

In relation to memory labels, there might be evidence for a particular time-dependent bias. Specifically, humans tend to recall the first and the last events from a sequence, a phenomenon referred to as recency/ primacy bias [157]. This, however, has not been investigated in the context of long interactions (rather than word lists and media-watching recall). If this hypothesis applies to the long discussions, most reported moments would occur in the beginning or the end of the session, and fewer in the middle of the session.

To test the hypothesis, we compared the memorability index (percentage of participants that included the segment in their memory reports, see labels used in [60]) of moments that occurred in the first 1/3 of the session, in the middle and in the end 1/3 of the session. Judging by an ANOVA followed by a paired t-test, the memorability index of the moments at the start and the end of the conversations is significantly higher than the ones in the middle ($p<0.005$). This therefore seems to confirm the recency and primacy bias in relation to the occurrence of memorable moments. Therefore, the memory labels should not be treated as independent of the temporal context within a session.

SESSION DEPENDENCY

Another factor to consider in the *MeMo* dataset relates to its' longitudinal quality. Since each group participated in 3 consecutive sessions, starting as strangers and gradually getting to know each other, there might have been some evolution in their relationships and group perception. This matters for two reasons. First, a session-dependent variation would indicate the success of moderated sessions in creating a connection within the group ($\rightarrow$ *G2*) and show whether the *MeMo* corpus is representative of general societal trends. Second, this variability is important for computational modelling to determine whether the sessions can be treated as independent of each other. To investigate whether there is a consistent change in participants' perceptions of each other and the group, we analysed self-reported ratings collected from participants after each session during the experiment.

On the **group level**, to investigate the development of interaction and group perception, we compared the questionnaire ratings task across sessions. In case the facilitation sessions were successful, the

hypothesis is that there would be higher ratings of group cohesion [146], entitativity [147], rapport and syncness in the consequent sessions in comparison to the first one. The significance of the differences was evaluated with the Friedman chi-square as a non-parametric analogy of repeated measures ANOVA (since the assumption of normality was not met by the data). The comparison of group perception scores between different sessions showed that the hypothesis is confirmed in relation to syncness (Friedman χ^2=13.7, p<0.005), rapport (Friedman χ^2=10.1, p<0.005) and entitativity (Friedman χ^2=31.7, p<0.005). In other words, participants considered that the group was significantly more harmonious, in sync and united in the 3rd session in comparison to the 1st session. This, however, did not hold for the group cohesion measure, with insignificant differences between the 1st and the 3rd sessions (Friedman χ^2=5.4, p>0.05).

On the level of **individual relational development**, to show whether there was a similar development in how close participants felt to each other throughout the three sessions, we investigated the change in IOS scores [145]. IOS (Inclusion of Other in Self) scale is a validated and comprehensible measure used to evaluate perceived relationship closeness between two participants [158]. Using this scale, each participant in *MeMo* evaluated how close they felt to every participant in their group after every session. The scale implies that the more familiar the participants felt with each other, the higher they would rate their subjective proximity on IOS scale. The hypothesis was that, with every next group session, participants would feel closer to each other, indicating growing closeness between participants. After comparing participants' IOS scores between sessions, the Friedman chi-square test showed that, indeed, participants felt closer to each other with every subsequent session (Friedman chi-square = 331.3, p<0.005). *Figure 2.5* illustrates how the distribution of IOS assessment gradually moves from the mode of 2 (little overlap) in the 1st session to 4 (equal overlap) in the 3rd session. This is consistent with previous findings showing that less acquainted participants score an average of 2 on the IOS scale, with friends scoring about 4 [158], therefore showing that the majority of participants developed a friendly relationship after the 3 discussion sessions facilitated by professional moderators in the *MeMo* dataset.

These results confirm that the *MeMo* corpus setup oversees the development of interpersonal relationships throughout the 3 interactions and therefore can serve as a useful tool for the evaluation of research questions related to longitudinal change in group dynamics and interpersonal relationships throughout repeated deliberation sessions guided by a professional moderator. This also shows that the sessions cannot be treated as independent from each other, with systematic relational and group perception differences between consecutive sessions.

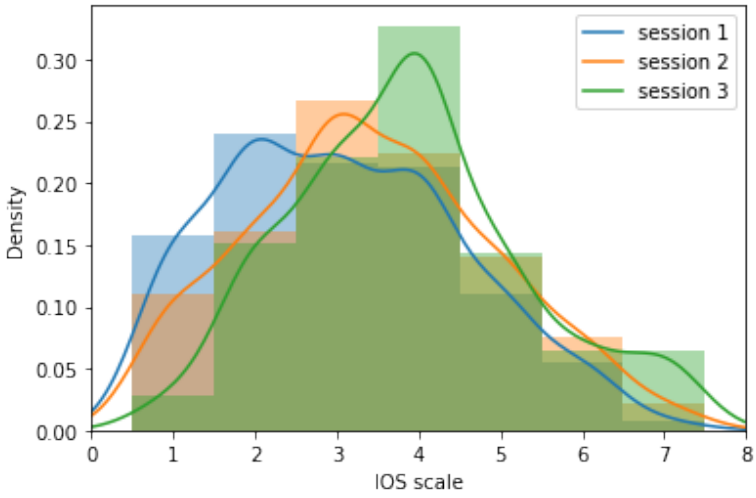


Figure 2.5.: The change in perceived social distance between participants throughout the 3 sessions of the interactions, reported through IOS scale, with 1 = no overlap, 3 = some overlap and 7 = most overlap [145].

2.8.2. EXAMPLE: GROUP-LEVEL CONVERSATIONAL MEMORY PREDICTION

The first baseline for memorability prediction based on the *MeMo* corpus has been explored and shown to be promising in [60]. Tsfasman et al. [60] used group gaze features to automatically predict group memorability levels.

Group memorability levels were computed from the first-party memory annotations described in *Section 2.5.2*, which provided the opportunity to calculate the percentage of participants who remembered each 5-second time-window of recorded discussions. Tsfasman et al. [60] used these proportions to create four labels of memorability level: 'zero' if no one from the group included a time window in their reports, 'low' if less than 30% remembered that time slice, 'middle' - 30 to 70% of participants considered a moment memorable, 'high' - if more than 70% reported an interval in their annotations.

Eye gaze features. To analyse group eye gaze behaviour and train the classifier, Tsfasman et al. [60] engineered multiple features computed from raw eye gaze direction and speaker activity annotation (see *Section 2.5.2*). Specifically, such group features as eye gaze presence, MaxGaze and Entropy. They also introduced two additional measures based on speech - the proportion of participants looking at the speaker at each given time slice and the proportion of participants

speaking in the given time slice.

Modelling. Tsfasman et al. (2022) [60] trained a Random Forest classifier and a Multi-layer Perceptron on the *MeMo* data with input being eye gaze features and output being four group memorability levels (zero, low, middle and high group memorability). Both models have been shown to predict the memorability labels above chance (chance being 0.25) with balanced accuracy scores of 0.42 and 0.43 respectively. These first results show how the *MeMo* corpus can be used for both computational (prediction of group memorability levels) and qualitative (analysis of different categories of memorable moments) studies. Training a rudimentary classifier on non-verbal features to predict the memorability level of a conversational interval resulted in above-chance performance. For future work, there is a need to create more complex models that take into account the temporal and session dependencies of the data points.

2.9. DISCUSSION: POTENTIAL FUTURE TASKS

This section introduces a series of potential multimodal modelling tasks for which MeMo might be a useful research resource. For each of these tasks, we provide a brief motivation and task description, and discuss the specific MeMo resources useful for the task.

We believe that MeMo provides a foundation of sufficient scope and size to facilitate preliminary investigations into these tasks. On the one hand, MeMo is comparable in size to existing datasets covering related phenomena (see Section 2.4.6 and our comparison in Table 2.1). On the other hand, substantial progress in technical research on (multimodal) foundation models and related procedures, such as fine-tuning pre-trained models ([159]), meta-learning ([160]), and few-shot learning ([161]) has led to an increase in data efficiency across diverse task settings. Together, these conditions make exploring these novel task settings using our corpus at least a plausible target for future research. We believe that MeMo provides a foundation of sufficient scope and size to facilitate preliminary investigations into these tasks. On the one hand, MeMo is comparable in size to existing datasets covering related phenomena (see Section 2.4.6 and our comparison in Table 2.1). On the other hand, substantial progress in technical research on (multimodal) foundation models and related procedures, such as fine-tuning pre-trained models (e.g. [159]), meta-learning (e.g. [160]), and few-shot learning (e.g. [161]) has led to an increase in data efficiency across diverse task settings. Together, these conditions make exploring these novel task settings using our corpus at least a plausible target for future research.

As mentioned before, memory can be broken down into three interdependent subprocesses - encoding, retention and retrieval. Since

memory representations are not directly observable, modelling all three processes relies on memory retrieval tasks (e.g. moment free-recall reports in our case) and the three subprocesses are never completely separable from each other. However, the main focus can be on one of the three. As such, we divide the possible memory modelling tasks by the primarily modelled memory subprocess (see *Section 2.1*).

2.9.1. CONVERSATIONAL MEMORY ENCODING MODELLING

Task Description. Modelling conversational memory encoding is the primary task of the*MeMo* corpus. Investigating the moment encoding, or preserving an event in memory, means focusing on the features of the timeframe of the actual event that a person's memory refers to. The task involves predicting the likelihood of a conversational segment being encoded by the conversation participants. In other words, a computational model would be trained on (non-)verbal behaviour of the participants at the moment of either encoding or not encoding a specific segment.

A predictive model could infer the likelihood of encoding an event either by individual participants (human-centred approach) or by the group as a whole (situation-centred approach). In a human-centred approach, the model could focus on individual behaviours, such as eye gaze, gestures, and speech, in relation to their own memory rating, investigating what behaviours indicate that a moment is important enough for a person to encode. Other participants' behaviours can also be used to understand what conversational context makes a moment memorable for an individual. In contrast, a situation-centred approach could focus on what qualities make a moment "universally" memorable for the group as a whole. Here, the model could use the behaviours of all participants, with the output being a cumulative metric of the percentage of participants who successfully encoded that specific moment (see *Section 2.8.2* or [60] for an example).

Task Relevance. Modelling how conversational moments are encoded through participants' (non-)verbal behaviour could be useful for various applications. In meeting facilitation systems, identifying which segments are likely encoded by participants can help focus on moments that enhance mutual understanding and shared history [3,162]. Sharing this information might highlight points one person found important but others missed, prompting further discussion. In automatic meeting summarisation, recognising which segments are likely encoded can produce summaries that emphasise key information or enhance memory by focusing on less memorable points [50,75].

MeMo Resources for the Task. The **output** labels for this task could be the memory features described in *Section 2.5.2* - free recall reports combined with participants' encoded events annotation (see descriptive

statistics of the memory measure in *Section 2.7.3*). These memory data could then be divided into binary labels per chosen time slice (retained/ non-retained for each specific participant) or a cumulative group-level aggregated measure as described in *Section 2.8.2* or [60]. One more possible option for an output variable could be a textual description from the participants' memory report(s). The raw video or audio (full as well as separated channels) data that can be used for the **input** measures is described in *Section 2.7.2*. Along with that, a variety of automatically extracted features is available for this purpose (see *Section 2.6.4*).

2.9.2. CONVERSATIONAL MEMORY RETENTION MODELLING

Task Description. To model **how humans retain conversations on a long-term**, it is possible to use the *MeMo* corpus with the long-term memory reports. Modelling retention implies predicting whether a moment will be remembered long-term using free-recall reports from two points in time: shorter term - straight after the interaction, and longer-term - 3-4 days after the interaction, when most forgetting would have occurred, and only the most persistent memories would have stayed [34]. The task, in this case, is similar to the encoding modelling, except for more incremental output labels - a time slice could be forgotten (not mentioned in any memory reports), retained in the short-term (only mentioned in the short-term memory reports), and retained long-term (mentioned in both short-term and long-term reports). For this purpose, however, more annotation is required to connect the long-term memory reports to the events in the recording. The input features could be the same as those in the encoding task - various verbal and non-verbal behaviour during the conversation.

Task Relevance. This task is relevant for long-term facilitation systems and conversational agents. Since long-term user engagement remains a challenge [163, 164], efforts have been made to enhance agents with shared memory models [73, 165–167]. However, there is no way to identify which memories are shared or forgotten by the user. Memory retention modelling could improve agents' long-term understanding of users and adapt dialogue strategies based on the likelihood of events being retained. Facilitation systems could also benefit from knowing which events are retained long-term. These models could be personalised for a better user experience or modified to support dementia patients.

MeMo Resources for the Task. The output labels for this task could be the memory reports described in *Section 2.5.2* - free recall reports, participants' encoded events annotation and long-term memory reports. The input features could be similar to the encoding task above (audiovisual data described in *Section 2.7.2* and features from *Section 2.6.4*).

In our initial modelling experiments (Section 2.8.2), we show that memory encoding can be predicted above chance using relatively simple classifiers and a limited set of non-verbal features. While further work is needed to evaluate the generalisability of these results to more complex modelling tasks, these findings suggest that the dataset may be sufficient in quantity for the other described computational tasks.

2.9.3. PERCEIVED REASON FOR RETENTION MODELLING

Task Description. To model the reasons why conversational segments were encoded and retained, the *MeMo* corpus contains self-reported reasons why participants recalled each moment. The reasons were then categorised by two annotators, with the label frequencies reported in [60]. In future research, it would be important to further investigate the types of reasons that participants report for considering a moment memorable. This could be a separate modelling task of inferring a perceived reason for remembering a moment from verbal or non-verbal data as well as memory reports themselves.

Task Relevance. This task offers a qualitative perspective on conversational memory modelling, providing an opportunity for deeper user understanding in applications, such as meeting facilitation and conversational agents. Inferring and understanding the perceived reasons for remembering can help extract information about the underlying relevance of a specific memorable event for a user. This reasoning could be crucial for further dialogue strategies in human-agent interaction. What is more, perceived reasons can serve to improve the accuracy of the encoding prediction and to refine the encoding memory prediction labels into more specific memorable event categories, potentially reducing noise in the data.

MeMo Resources for the Task. The output labels could be the description of the perceived reasons mentioned in *Section 2.8.2*, as well as a categorisation of these descriptions described in [60] which are available by request. The input labels could be the moment free-recall memory reports as well as any conversational behaviour, similar to the encoding and retention tasks above.

2.10. CONCLUSIONS

In the present paper, we introduce *MeMo* - the first multimodal corpus with first-party memory annotations. The multi-party interactions were conducted in an ecologically valid, spontaneous setting in a typical online meeting environment and participants in the comfort of their own homes. For each group, there were three 45-minute group meetings spread 3-4 days apart, which provided opportunities for investigating group dynamics and relationships in newly formed groups.

The rich collection of perceptual measures from pre- and post-session questionnaires can serve as a source for studies on how individual participant characteristics and perception of interaction, group, and other participants affect group dynamics and conversational memory. With the first investigation on how non-verbal signals can indicate conversational memorability, we show that group eye-gaze behaviour can discriminate over conversational memorability levels [60]. We also show that throughout 3 sessions there was an interpersonal relationship development: participants assessed their groups as having more rapport, syncness and entitativity. Participants assessed their social distance from other participants as closer in the 3rd session in comparison to the first session.

With this research, we hope to pioneer multi-modal corpus research of conversational memory and create opportunities for studying conversational memory and group dynamics, along with other topics related to longitudinal group discussions.

2.11. LIMITATIONS

As with any dataset, the *MeMo* corpus has its limitations. In the following, we will highlight several key aspects relevant to the intended use of the corpus.

Validity of Memory Annotations *MeMo* contains two types of memory annotations: Moment Free-Recall Self-reports (MFRS), and Encoded Event Annotations (EEAs). While we argue that jointly these provide valid resources for modelling both the encoding and retention processes of participants, there are some caveats we want to highlight.

First, our corpus assumes that each MFRS corresponds to a single EEA, i.e., a remembered moment in the conversation corresponds to a single segment in the multimodal signal. However, this might not always be true, for example, when multiple similar events are reported as a single memory (with more than one EEA included in one MFRS).

Secondly, it is likely that there are events that participants remembered but failed to either report or annotate. Our protocol required participants to identify multiple segments in the recording at a fine granularity, which is a cognitively taxing task. This might impact the comprehensiveness and accuracy of the resulting annotations.

Finally, our protocol asked participants to report MFRS from previous sessions (see *Section 2.5.2*). We have argued that this data forms a viable source for modelling retention. However, we cannot rule out that participants could access memories from previous sessions but failed to report them (i.e., we do not know if they forgot about a moment or simply failed to produce a matching MFRS).

In addition, the dataset therefore does not provide any metrics of 'recall rates' used in previous literature [17], or any measures of recall quality (e.g. percentage of the information that was retained vs forgotten). The memory annotation only provides the encoded conversational events, accessible to participants at the time of the moment free-recall task. This said, it could be possible to compute recall quality using third-party annotators, for example, by annotating free-recall reports as well as conversation transcripts with idea units and computing the percentage of the units that have been reported within memorable moments.

Comprehensiveness of Memory Annotations While we believe that MFRS and EEAs capture properties relevant for modelling retention and encoding, the current annotations do not facilitate work on the functional use of memories in conversations (i.e., how people use memories to convince others or for bonding purposes [11]). However, it is likely that *MeMo* captures a substantial amount of such instances, and we intend to investigate their presence in future work (i.e., to provide additional annotations). Note that even though our protocol for obtaining MFRS explicitly asks participants to "recall" memory, this is merely a methodological necessity (i.e., it is not possible to study memory content without prompting for it); it does not facilitate modelling retrieval processes as such (i.e., the conditions under which content is dynamically accessed or accessible).

Choice of Conversational Setting First, the *MeMo* corpus was recorded online. Therefore, it is specific to the memory of online video-call conversations, which could be different from the memory of face-to-face interaction, because of different conversational dynamics [168]. This said, the research on how humans remember information from online and face-to-face lectures shows no differences in recall quality between these settings [169].

Second, the topic of the conversations in the *MeMo* corpus is limited to the Covid-19 pandemic. At the time of the recording, this topic was naturally engaging since it noticeably affected most people's lives. Humans tend to remember information that is personally relevant to them [170] and, therefore, with a different topic, the trends in memorable moments might have been different.

Additionally, having a trained moderator may create a sense of hierarchy in the group as well as introduce different moderation styles, potentially affecting the discussion structure and group dynamics. A different setting might lead to different results, depending on the environment, goals, and roles in the conversation.

Notes on ecological validity Although we strove for maximising ecological validity in the MeMo corpus (see Section 2.3), it still has some differences from the real-world interactions. Although the online setting of the corpus is highly representative of a typical online meeting setup (i.e. in the comfort of people's homes and with their own technological set-up, using Zoom software that at the time of the data recording had est. 300 million daily active users worldwide [123]), it was still an experiment and therefore does not represent complete real-world interactions. First, participants had to complete surveys before and after the recording, which does not usually happen in real-world settings. Second, participants were recruited for the experiment and were not an already acquainted team, as they would be in a real-world work meeting. For the same reason, although paid for the experiment, they did not have the social and professional incentives in mind that a real-world work team would have. Third, the conversations were guided by moderators, and although it is typical to have a professionally assigned or emerging leader in team meetings, the moderation style or the fact that they are assigned rather than emerging meeting leaders could affect the representativeness of the real-world team meeting dynamics [171]. In addition, although representing a wide range of demographics (e.g. age groups), our sample includes people who fluently speak English and reside in the UK, and therefore only represents a specific subsection of (mostly white) people in a developed English-speaking country and therefore might not generalise to the rest of the world.

2.12. DATASET AND CODE AVAILABILITY

The MeMo dataset will be made publicly available following a comprehensive review process. This process aims at the removal of sensitive and potentially harmful information to the best of our knowledge. Once this step is complete, the wider release will occur gradually through multiple subsets of data, each focusing on different aspects of the dataset. The first release, associated with this paper, will include audio-visual recordings and temporal memory segment annotations. Unfortunately, specific timelines for the dataset release cannot be provided at this time.

ACKNOWLEDGEMENTS

This material is based upon work supported by Delft Institute for Values; Hybrid Intelligence Center, a 10-year program funded by the Dutch Ministry of Education, Culture, and Science through the Netherlands Organisation for Scientific Research; European Commission funded project "Humane AI: Toward AI Systems That Augment and Empower Humans by Understanding Us, our Society and the World Around Us"

(grant # 820437); the National Science Foundation (NWO) under Grant No. (1136993) and Grant No. (024.004.022); TAILOR, a project funded by EU Horizon 2020 research and innovation programme under GA No 952215; European Union project RRF-2.3.1-21-2022-00004 within the framework of the Artificial Intelligence National Laboratory. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the supporting institutions. The support is gratefully acknowledged.

3

THE RELATIONSHIP BETWEEN MEMORY AND AFFECT

[submitted to IEEE Transactions on Affective Computing] **Maria Tsfasman**, Ramin Ghorbani, Catholijn M. Jonker, Bernd Dudzik, "The Emotion-Memory Link: Do Memorability Annotations Matter for Intelligent Systems?" *arXiv preprint arXiv:2507.14084*, 2025. https://arxiv.org/abs/2507.14084

ABSTRACT

Humans have a selective memory, remembering relevant episodes and forgetting less relevant information. Possessing awareness of event memorability for a user could help intelligent systems in more accurate user modelling, especially for such applications as meeting support systems, memory augmentation, and meeting summarisation. Emotion recognition has been widely studied since emotions are thought to signal moments of high personal relevance to users. The emotional experience of situations and their memorability have traditionally been considered to be closely tied to one another: moments that are experienced as highly emotional are also considered to be highly memorable. This relationship suggests that emotional annotations could serve as proxies for memorability. However, existing emotion recognition systems rely heavily on third-party annotations, which may not accurately represent the first-person experience of emotional relevance and memorability. This is why, in this study, we empirically examine the relationship between perceived group emotions (Pleasure-Arousal) and group memorability in the context of conversational interactions. Our investigation involves continuous time-based annotations of both emotions and memorability in dynamic, unstructured group settings, approximating conditions of real-world conversational AI applications such as online meeting support systems. Our results show that the observed relationship between affect and memorability annotations cannot be reliably distinguished from what might be expected under random chance. We discuss the implications of this surprising finding for the development and applications of Affective Computing technology. In addition, we contextualise our findings in broader discourses in Affective Computing and point out important targets for future research efforts.

3.1. INTRODUCTION

Memory for conversations and other social interactions plays a crucial role in shaping social bonds and fostering relationship building [11]. Considering human conversational memory in intelligent systems is, thus, essential for explaining and predicting human behaviour in conversations, including their affective responses. Conversational memory can be defined as a subtype of episodic memory, which manages the encoding, storage, and retrieval of personally experienced events [80], particularly within conversational settings.

Affective Computing (AC) has long focused on recognising and interpreting human emotions to enhance interactions between users and intelligent systems [18]. Emotions are considered to be central to human experience, shaping decision-making, social interactions, and memory. Their automatic detection is valuable for intelligent systems because emotional responses often signal moments of high personal relevance to users. To capture these signals, Multimodal Emotion Recognition (MER) commonly uses human behavioural cues, such as facial expressions, speech patterns, and physiological signals, to infer emotional states. While MER has made significant strides in detecting momentary affective states, its potential to model longer-term cognitive processes, such as memory, remains underexplored.

Both theoretical and empirical research suggests that the way we emotionally experience events is strongly linked to how well we remember them [37]. Emotional arousal enhances memory encoding and retrieval, with emotionally charged events being remembered better than neutral ones [38]. The effect is linked to hormone release during arousing experiences, which strengthens memory formation in the brain [39]. Both valence [172, 173] and arousal [39, 40], as well as their combination [174], have been shown to enhance memory processes. Additionally, affect and memory are closely tied to the personal relevance of a stimulus, as relevance influences both emotional experience [175] and the likelihood of remembering an event [45]. Given these well-documented relationships, it is reasonable to hypothesise that perceived affect could serve as a proxy for memory.

These and similar findings have been used to motivate various intelligent systems to integrate emotional components into computational memory models, e.g., to drive interactions between users and virtual agents [41–43], social robots [176, 177], or between agents in multi-agent systems [178].

However, despite this prevalent conceptual connection between emotional responses and memorability, Affective Computing research has not yet explored this connection in the context of MER technology. Unfortunately, without targeted empirical exploration, it remains unclear to what extent theory and findings from the behavioural sciences connecting the two phenomena actually translate to many of the

settings in which MER technology is developed or expected to operate. Notably, the following aspects span crucial practices of the development and deployment of MER technology but are not sufficiently covered in existing findings connecting the two phenomena:

1. **Choice of Annotation Perspectives:** Although research robustly links experienced emotions (measured through self-reports and physiological signals) with memory encoding, it remains uncertain whether third-party observed affect annotations, which are widely used in affective computing, can reliably serve as proxies for personal memorability (more detail in *Section 3.2.1*).
2. **Continuous Conceptualisation**: Previous research on the link between affect and memory operationalises those as static states, while in MER systems it is more common and desirable to view those as continuous. The link between continuous annotation of affect and memory has not been studied, to our knowledge (see *Section 4.2.3*).
3. **Group-based Analysis**: Group-based MER systems are crucial for real-world applications like meetings and collaborative tasks, yet existing research on the emotion-memory link primarily focuses on individuals, overlooking the social dynamics that shape both affect and memory in group interactions (see *Section 3.2.3*).

Given these gaps, the extent to which third-party affect annotations capture memory-relevant information in real-world settings remains an open question. To address this question, in this paper, we present an empirical investigation evaluating the association between annotations of Perceived Group Emotions (Pleasure-Arousal) and Group Memorability. Our study leverages time-continuous annotations of emotions and memorability in dynamic, unstructured group interactions, mirroring real-world conditions relevant to MER applications like online meeting support. We discuss the implications for Affective Computing, situate our findings within broader Affective Science discussions, and highlight key directions for future research to bridge the gap between computational modelling of affect and memory.

3.2. BACKGROUND AND MOTIVATION

In this section, we briefly expand on some of the core properties of Affective Computing development practices that we believe could limit the insights that existing empirical findings on the emotion-memorability link can provide for the development and applicability of *Multimodal Emotion Recognition (MER)* technology. To support this discussion, we provide a graphical overview of the components and relationships involved in the development practices we discuss in *Figure 3.1*.

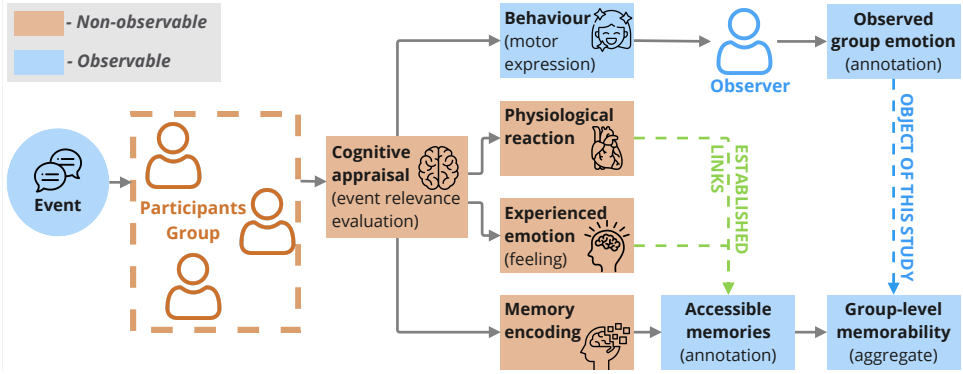


Figure 3.1.: An illustration of perceived vs experienced emotion and the object of this study (shown with a blue dashed arrow). The states with observable information are shown in blue, and the states we have no access to are shown in orange. The green dashed arrow represents the relationships that have been studied in previous literature. (Based on the component model of emotions according to Scherer [179])

3.2.1. CHOICE OF ANNOTATION PERSPECTIVE

As mentioned in *Section 3.1*, there is substantial empirical evidence indicating a link between emotional episodes and the memorability of events. However, the findings supporting this link are largely grounded in physiological signals of emotions and self-reports of first-person experience (e.g. [38–40, 172, 173]), i.e., they capture what we could call "Experienced Emotions" and "Self-reported Memorability" of events. When a person encounters an event, the event is thought to be experienced through a process of cognitive appraisal - evaluation of event relevance, implication, coping potential and normative significance [180]. Such appraisal is thought to result in motor expression (behaviour), physiological reaction and a subjective feeling (experienced emotion) according to Scherer's widely accepted component model of emotion [179]. Along with that, a process that is triggered is memory encoding (if the event is considered relevant). The well-established link between emotion and memory described above applies to the relationship between memory encoding and measures of experienced emotion (subjective feeling) [45, 172–174] or the physiological responses related to emotions [38, 39] (shown with a green dashed arrow in *Figure 3.1*) [1].

In the context of affective computing, however, it is more common to focus on observed emotion - a hired annotator (observer in *Figure 3.1*)

[1] Note that these studies have focused on individual memory and emotion, not memory in a group context.

watches videos and labels arousal and valence of participants' emotional behaviour [44]. Relying on observed annotations is done on the one side for pragmatic reasons - it is easier to employ external annotators than to collect self-reports, and it is possible to hire many annotators for each temporal segment to reduce annotation subjectivity. On the other side, it is more practical for modelling reasons, since annotators rely on external behaviour to make their judgement, similar to what an emotion recognition model would do.

This distinction between experienced and observed emotion is important since the observed emotion is an external evaluation of a participant's behaviour by a third-party observer. Since such third-party annotation is solely based on the observed behaviour, it may contain some inaccuracies. For example, not all emotions might be expressed through behaviour or sometimes the expressed behaviour might not reflect the experienced emotion because of social norms (e.g. covering anger with a polite smile to avoid confrontation). In addition, in the case of group affect [181], the third-party annotation reflects the observed emotions of the group rather than the individual participants one-by-one and might not be equal to specific emotions of each group member (e.g. if one participant displays a negative valence while three display highly positive valence, the valence of the group would probably be labelled as high).

In conclusion, while the literature points towards a clear physiological link between experienced affect and memory encoding, the question of whether observed affect annotations are representative of memory labels remains open. It remains uncertain whether previous insights from the behavioural sciences can inform research about the ability of Multimodal Emotion Recognition (MER) to provide insights into the memorability of situations.

3.2.2. TIME-CONTINUOUS OPERATIONALISATION OF CONCEPTS

A common practice in developing MER systems is to collect data that operationalises emotional responses through *time-continuous measurements* (e.g., annotations collected for every frame in a video stream [46]). Some of the proposed benefits [47] of this practice are considered to be its high temporal granularity (i.e., being able to capturing nuanced changes in emotional qualities over time), but also its capacity to capture emotional variability (i.e., being able to describe changes in emotional qualities within some specified unit of analysis, such as a video clip). Pragmatically, it seems also plausible that time-continuous estimates of emotion are highly desirable from the point of view of many applications, since they might enable systems to respond to changes observed in users dynamically.

While common in data collection for MER development, studies inves-

tigating the emotion-memorability link in the behavioural sciences do not operationalise either concept with such time-continuous measures. Instead, these concepts are typically operationalised with self-reports describing events as reconstructed from memory, without access to information at encoding time or fine-grained breakdowns of parts of the event (e.g., see Talarico et al. [182] for an example).

Given this misalignment in operationalisation, it seems unclear to what extent existing findings about the emotion-memorability link can generalise to the outcome achieved with practices used for developing MER. For example, memory biases might distort how emotion is attributed to remembered events or emotional connections to memorability might be due to increased rehearsal over time [37] (see also Dudzik and Broekens [183] for a more extensive discussion of potential influences manifested by the choice of when to provide emotional self-reports). Overall, it does not seem self-evident that estimates based on datasets that have a time-continuous affective ground truth reliably approximate memorability when similarly operationalised.

For this reason, our study operationalises both concepts in a time-continuous way, leveraging the MeMo dataset [36] that contains relevant annotations for memorability, which were recently complemented with associated time-continuous Affect Annotations by Prabhu et al. [181].

3.2.3. GROUP-BASED ANALYSIS

Many applications for which Multimodal Emotion Recognition is beneficial are often expected to operate in group settings, such as work meetings [5], educational settings [184], and collaborative tasks [185]. In these contexts, MER can not only be used to analyse individual emotions but also to assess group-level affective dynamics, where the target of predictions extends beyond individuals to an entire team and its emergent characteristics.

Group affect, the collective emotional state of a group, has gained attention in affective computing and computational modelling. It encompasses shared moods and emotions among group members during interactions [25, 48]. Research has explored integrating group affect into decision-making processes, developing computational models that consider individual, group, and emerging processes [186]. Studies have investigated the dynamics of group affect, including convergence and divergence of affective expressions, using multimodal approaches to extract synchrony-based features from audio and visual cues [181].

Existing research on the relationship between emotion and memory in the behavioural sciences has primarily focused on individual experiences rather than group-based settings. Most studies investigating this link have been conducted in controlled environments where participants

engage with stimuli in isolation, rather than within dynamic social interactions (e.g., [38–40, 172, 173]). Social interactions such as conversations differ from individual contexts because of the continuous exchange of (non-)verbal signals and relational dynamics, changing the resulting quality and quantity of remembered information [12–17, 33]. Although there is some research on emotion and memory in social contexts, the literature linking the two concepts to each other typically conceptualises them as personal cognitive-affective processes, without considering how these phenomena may emerge differently in conversations and other group contexts. This individual-level focus limits the applicability of prior findings to real-world scenarios where memory and emotions are often shaped by collective interactions.

This gap is particularly relevant for Affective Computing applications, as the cognitive-affective processes at play in group settings may differ significantly from those observed in isolated individuals. Social dynamics such as emotional contagion and regulation not only influence individual affective states but also the overall emotional climate of a group, potentially affecting how shared experiences are remembered, while also significantly impacting group interaction outcomes, including creativity, analytical performance, sense of belonging, and information sharing [132, 187, 188]. These factors suggest that findings from traditional emotion-memorability studies may have limited capacity to inform the development of MER systems intended for group-based applications, underscoring the need for research that explicitly addresses affect and memory at the group level.

For these reasons, we believe that our study's setup provides a meaningful addition to the existing body of research since it explicitly focuses on group-based analysis: it 1) takes place in group-conversational settings and 2) conceptualises both emotion and memorability as group-level constructs.

3.2.4. RELATED WORK: MEMORABILITY PREDICTION

The few studies that do investigate memorability from a computational perspective focus on memorability of media stimuli [20, 21, 35]. These studies have generally been successful in identifying features (such as semantic richness, emotional valence, and visual distinctiveness) that contribute to enhanced memorability of media segments across individuals [20, 21, 35]. Such memorability modelling has the potential to advance user modelling to understand what is actually relevant for the user and what needs to be repeated or reframed for a greater impact on the user in the long term. Nevertheless, in an arguably more common and socially important context of conversations, memory modelling remains underexplored. Unlike media memorability, where the stimulus is a fixed and repeatable entity, conversational memory

emerges from dynamic, interactive, and multi-speaker contexts, making it more complex to model. Moreover, many conversations happen in a group of people, offering an additional complexity as well as a source of insights into what is considered memorable in a group (e.g. team meetings, friends and family gatherings).

3.2.5. CONVERSATIONAL MEMORY MODELLING

Although understanding the way humans remember conversational stimuli has been the subject of decades of cognitive research (see [12–17] for illustrative examples), it has only recently been approached from a computer science perspective. To our knowledge, only one study and dataset have addressed the task of predicting conversational memory: Tsfasman et al. [60] has introduced a baseline model using the MeMo conversational memory corpus [36].

3.3. METHODS: DATASET

3.3.1. DATA SOURCE

In this paper, we use data contained in the recently created MeMo dataset [36]. It was collected in an online video-conferencing setting of 45-minute longitudinal group conversations. Each group included 3 to 5 participants and one professional moderator. The moderator was tasked with keeping the conversation going and trying to keep the atmosphere in the group comfortable for participants to be ready to openly express their opinions and emotions. The topic of the conversations was the Covid-19 pandemic, focusing on people's experiences and opinions about the pre- and post-pandemic world (a relatable topic for many at the time of the recording of the corpus in 2021). Each group participated in three 45-minute long conversations scheduled 3-4 days apart. While the MeMo corpus originally does not contain any affect annotations, these were provided in a follow-up by Prabhu et al. [181].

3.3.2. DATA PREPARATION

In this paper, we used a subset of the original MeMo corpus: the sessions and specific timestamps for which Prabhu et al. [181] have collected affect annotations. This selection resulted in 3 groups from the original MeMo corpus being excluded and the timestamps at the start and end of the recordings being trimmed to match the timestamps of the annotations (for details on the exclusion criteria, see the relevant publication [181]). In addition, we excluded 4 sessions with gaps in memory annotation (i.e., instances where at least one participant did not fill out the post-session survey). The subset of MeMo used in this paper consisted of 30 conversational sessions totalling 1457 minutes of

recording (mean of 41.6 min. per video +- 7.5 min.), 12 groups with 42 participants in total.

3.3.3. ANNOTATION COLLECTION

Perceived Group Affect. Group affect annotations are present in 15-second intervals across two affect dimensions: arousal and valence, based on Russell's circumplex model [189]. The annotations are on an ordinal scale (1-9), allowing annotators to express varying intensities of affective states in a continuous manner. The study employed 8 annotators with backgrounds in organisational psychology and prior experience in annotating social behaviours. Each annotator underwent training designed to ensure consistency and reliability in their evaluations. This training emphasised the importance of focusing explicitly on observable emotional behaviours, ensuring that annotations were grounded in visible cues rather than inferred internal states. Each group interaction video was assessed by at least 6 annotators. Inter-annotator reliability was evaluated, revealing moderate agreement across both affect dimensions. For additional details, see Prabhu et al. [181].

REMEMBERED MOMENTS

Memory annotation procedure. The memory annotation procedure consisted of two stages. First, immediately after each conversational session, participants completed an open-ended free-recall task, reporting up to 10 moments they remember in their own words with as much detail as possible in a free form. This step ensured that memories reflected participants' accessible recollections at the time (for more detail, see Tsfasman et al. [36]). Second, participants reviewed the session recordings to match their reported moments to specific events by providing start and end times or indicating if the memory lacked a precise interval. This self-assignment ensured memory-event alignment based on participants' perspectives rather than third-party interpretations. The process was designed to prioritise construct validity while managing participant fatigue and minimising bias (see more details on the MeMo setup in Tsfasman et al. [36]).

In total, there were 419 self-reported memorable moments in the data used in this paper, with a mean duration of 110 seconds (+- 116, from 1 to 580 seconds)[1].

[1]The minimum duration of a memorable moment is equal to 1 second because the videos were cropped based on the valid segments used in the curated dataset and affect annotations. A one-second moment is likely part of a longer event, with its start or end removed for varying reasons (see Tsfasman et al. [36] and Prabhu et al. [181] for details on video cropping).

3.3.4. PROCESSING AND DERIVED MEASURES

In this section, we briefly outline how we processed the annotations described above to arrive at operationalisations of *Perceived Group Affect* and *Group Memorability*.

GROUP MEMORABILITY MEASURES

Continuous Group Memorability Index. To compute the combined measure of what segments were memorable to the group, for each second of the recording, we calculate a Group Memorability Index. It is defined as the ratio of participants from a group that included those timestamps in their annotations for Remembered Moments.

Binary Group Memorability Index. From the continuous memory measure, we compute a Boolean metric of whether each individual second was recalled by at least one member from within the group. The memory Boolean value was 0 if the memory index was 0. If the memory index was greater than 0 the value was 1. In essence, this variable encodes the most generous operationalisation of the group memorability concept.

PERCEIVED GROUP AFFECT MEASURES

Continuous Affect: Valence and Arousal. We aggregate the arousal and valence annotations across raters using a median of all the reported scores for each second. In addition to aggregation, we shift the scales to be able to compute an additional measure of affect intensity (see below).

Specifically, we shifted the values for both dimensions to Likert scale from 0 to 8. For arousal we used the following formula:

$$A'_t = A_t - 1 \tag{3.1}$$

where t represents the timestamp (in seconds).

To better reflect the bipolar nature of valence, we shifted and rescaled the original values according to the following formula:

$$V'_t = 2 \cdot (V_t - 5) \tag{3.2}$$

This changes the original 1-to-9 scale to a symmetric range from -8 to 8, ensuring that the intensity of both negative and positive valence is represented with equal magnitude.

Continuous Affect: Intensity. Some versions of the emotion-memorability link conceptualise the emotional component not in terms of specific emotional qualities (e.g., pleasure), but instead as the intensity of the emotional episode at the time of encoding [190]. According to Reisenzein [191], emotional intensity refers to the strength or magnitude of an emotional experience, which can be understood

in terms of the circumplex model of affect [189]. This model posits that emotions are organised along two primary dimensions: pleasure (valence) and arousal. Emotional intensity represents the degree to which a person experiences these two dimensions, ranging from mild to strong. In this framework, the intensity of an emotion can be computed based on the values of arousal and valence [192]. In this paper, emotional intensity labels **I** for timestamp **t** is calculated as the Euclidean norm of valence score V_t and arousal score A_t:

$$\mathbf{I_t} = \sqrt{V_t^2 + A_t^2} \tag{3.3}$$

The resulting intensity metric ranges from 0 to $\sqrt{8^2 + 8^2}$(∼11.3).

Affect - Binary. To investigate if the relationship between group affect and memory annotations is clearer with binary affect values with a meaningful threshold, we computed binary affect labels with the middle of each annotation scale as a threshold: 0 for valence, 4 for arousal and 4 for intensity. The threshold of 4 for intensity does not correspond to the mathematical midpoint of the intensity range, but is instead derived from a functionally meaningful point: the intensity calculated when arousal is at its midpoint (4) and valence is neutral (0). That is:

$$\mathbf{I_{mid}} = \sqrt{V_{mid}^2 + A_{mid}^2} = \sqrt{0^2 + 4^2} = 4 \tag{3.4}$$

This threshold reflects a moderate level of emotional engagement based on the meaningful input scales, rather than a purely statistical midpoint. It is thus used to distinguish between lower and higher affective intensity in a way that reflects the underlying annotation schema and preserves interpretability in relation to arousal and valence contributions.

3.4. METHODS: ANALYSIS

3.4.1. METRICS

Comparing human internal states to each other to evaluate their alignment brings several challenges. First, human data is typically characterised by noise and is prone to errors, especially when it comes to perceived measurements such as affect (e.g., there might be a delay in annotation [46] or misinterpretations or lapses in annotators' attention). Second, there might be confounding variables at play, such as the introduction of new stimuli or non-measured factors. Third, when dealing with continuous data such as human interactions and changing internal states, there is temporal context that needs to be taken into account. To address these challenges, it is essential to use evaluation metrics that are sensitive to the temporal nature of the data. Traditional measures of time-series comparison, such as Mean Squared

Error (MSE) or Mean Absolute Error (MAE), penalise slight temporal shifts or discrepancies too harshly.

PATE

Therefore, the first metric we rely on is a recently introduced metric designed specifically for time-series predictions that accounts for such temporal shifts - the Proximity-Aware Time series Evaluation (PATE) metric [193]. PATE is conceptually similar to the Area Under the Precision-Recall Curve (AUCPR) but introduces additional considerations for time-series. Specifically, PATE does not treat all mismatches between predictions and ground truth equally; instead, it introduces a tolerance window around each ground truth event, recognising that in real-world human-centred data, small temporal shifts in responses (such as memory recall or affect expression) are expected. The metric assigns partial credit to predictions that fall within this window, reducing the penalty for minor misalignments. The PATE framework introduces the assumption that temporally proximate events are functionally related, making it particularly well-suited for evaluating human behavioural data where slight timing discrepancies should not be treated as errors.

In this paper, we use the standard **PATE** measure, which operates on one binary time-series (memory Boolean in our case) and one continuous (affect annotations), by setting various thresholds on the continuous data and computing the area under the curve as the output PATE value. Additionally, we use **PATE F1**, which compares two binary time-series representations, to assess whether a meaningful manually set threshold on affect data, such as the midpoint of the scale, would indicate a stronger relationship with memory labels.

EUCLIDEAN DISTANCE

The PATE metric assesses relationships based on discrete categories rather than continuous values. This binary nature can lead to limitations when examining more nuanced or gradual changes in affect and memorability traces, as it does not account for the varying degrees of response or activation that may occur outside of strict categorical boundaries. Therefore, to quantify the relationship between emotion and memorability annotations without threshold assumptions, we use **Euclidean distance** as an additional metric. Euclidean distance offers a straightforward estimate of how similar the two time-series are to each other.

DYNAMIC TIME WARPING (DTW)

Euclidean distance does not take into account small shifts or stretches in the data, for example, if the affect signal increases consistently some

time before/after the interval becomes memorable, which is plausible based on the literature [115, 194]. This is why our final metric is **Dynamic Time Warping distance** (DTW) [195]. DTW is a technique used to measure similarity between two temporal sequences that may vary in speed or timing. By warping the time axis, DTW aligns sequences in a flexible manner, allowing for the comparison of patterns that may be out of phase. This adaptability makes DTW particularly suitable for analysing time-series data in the context of memory and affect, as it can capture meaningful correlations even when events occur at slightly different times, thus providing a more accurate representation of the underlying relationships.

Since all these metrics denote different types of relationships, we compute all four measures - PATE F1 on binary memory labels (*Section 3.3.3*) and binarised affect with manually set thresholds (*Section 3.3.3*), PATE on binary memory and continuous affect, Euclidean distance and DTW distance on continuous memory (*Section 3.3.3*) and continuos affect (*Section 3.3.3*).

Since our distance-based metrics, such as Dynamic Time Warping (DTW) and Euclidean distance, compute mathematical differences between values at each time-step, it is important to ensure that all time-series are on comparable numerical scales. The memory signal in our analysis is already bounded between 0 and 1, while the affective dimensions (e.g., valence, arousal, intensity) span broader and heterogeneous ranges (e.g., from –8 to 8 or from 0 to 11.3). Without scaling, these differences in range would bias the distance metrics toward dimensions with larger numerical intervals, thereby distorting the results. To address this disparity, we normalised all affective dimensions to a range of 0 to 1, ensuring consistency and meaningful comparisons across metrics. Note that valence was normalised from 0 to 1, resulting in the loss of its original bipolar structure (i.e., negative to positive). This transformation is justified in the context of our distance-based metrics, which assess the similarity in shape and temporal alignment between signals, rather than their absolute semantic values. Since the goal is to compare how closely the dynamics of affective signals track memory over time, rather than interpret the polarity of affect, preserving comparability in scale takes precedence over retaining the original meaning of zero as neutrality. This said, for all other metrics (PATE, PATE F1) the annotation scales are not normalised and remain in their original shape.

3.4.2. STATISTICAL TESTING PROCEDURE

A crucial challenge testing the relationship between our memory and affect annotations lies in the absence of a baseline to interpret the strength of that relationship.

Simply put, while we can compute correlation and similarity metrics (PATE F1, PATE, Euclidean distance, and DTW) to quantify the association between these variables, interpreting these values without a frame of reference leaves us uncertain about whether the observed relationships are relatively strong, weak, or even meaningful.

To address this, we followed the procedure shown in *Figure 3.2*, generating synthetic affect data with different assumptions in three experiments. These synthesised datasets act as estimates of the sampling distribution under the null hypothesis, enabling us to create a comparative framework for understanding the relationship between memory and group affect annotations. In each experiment, we simulated synthetic data with a specific hypothesis in mind: for example, assuming random affect annotations independent of memory (experiment 1), or annotations shuffled over time (experiment 2).

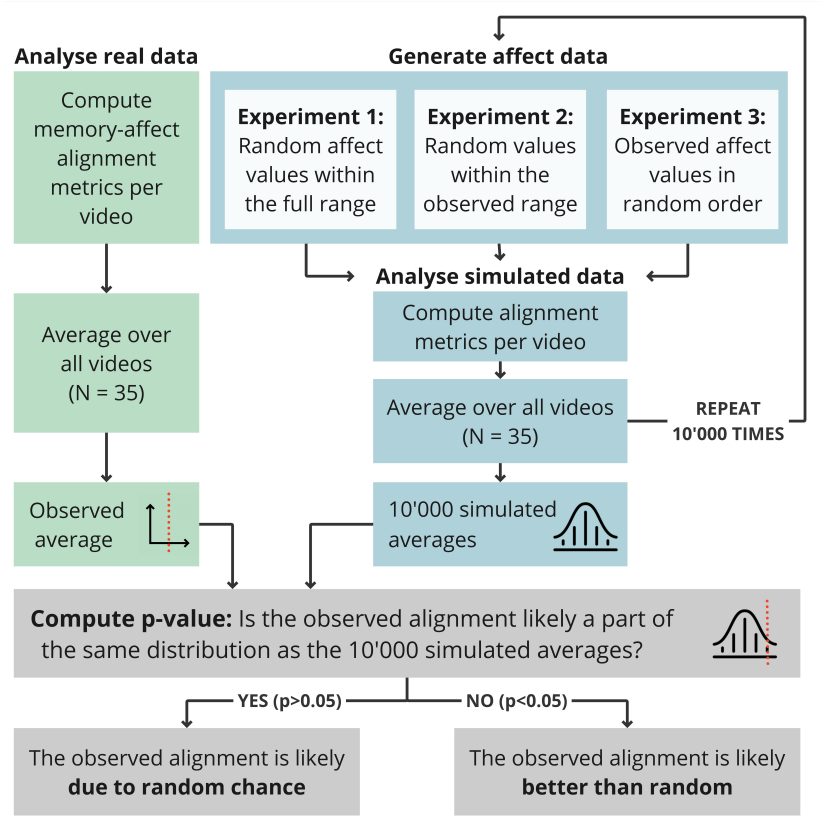


Figure 3.2.: Illustration of the affect-memory annotations comparison procedure

Using the same metrics as for the actual data, we then evaluated the

relationship between each synthetic and the real memory annotations. For each iteration of an experiment, we computed the metric averages over 35 sessions, representing each session's memory data within the dataset and comparing it to a synthesised affect annotation. To ensure the robustness of our results and account for random variations, we repeated this process for 10000 iterations, generating new synthetic data at each iteration. This produced a distribution of 10000 average metric values under the null hypothesis for each experiment.

With this distribution, we could then assess how the actual observed relationship values between real memory and affect annotations compare to the simulated point of reference. This was computed by calculating the probability that the real data metrics belong to the same distribution as the synthetic data. Specifically, by examining how likely the real data metrics are to belong within the range of values produced by the synthetic data, we could assess whether the observed relationships in the real data are likely to have occurred by random chance. If the real data metrics deviate significantly from the synthetic distribution, one would reject the null hypothesis that the association between these two variables is not different from random chance. In other words, this would suggest a meaningful difference in the alignment between group affect annotations and memory in the actual dataset compared to what we would expect under random conditions.

The three computational experiments differed in the way the data under the null hypothesis was generated. In the 1st experiment, we generate the random affect data using the fewest assumptions - drawing random affect values from a random distribution of the full feature range (-8 to 8 for valence, 0 to 8 for arousal and 0 to 11.3 for intensity). This experiment is meant to answer the question of how likely it is that the relationship between memory and affect annotations is completely random. In the 2nd experiment, we generate affect annotations closer to the original data by keeping the observed range of affective labels, i.e. if in one session the arousal annotations range from 1 to 6, this is the range we use for the annotation generation for that session. In the 3rd experiment, we mimic the distribution within the observed affect when generating annotations while destroying the temporal alignment between affect and memory - we take the real affect annotations and shuffle them within the time-series. This experiment is meant to test the strength of temporal alignment between affect and memory. In all the experiments, the comparisons are performed on emotional arousal, valence and intensity annotations separately.

The decision on rejection of the null hypothesis was made using the following rule: if the p-value is significant across all 3 experiments, we could reject the overall hypothesis that the relationship between memory and observed affect annotations could belong to the same distribution as the relationship between memory and random or permuted affect

annotations.

3.5. EMPIRICAL INVESTIGATION

3.5.1. EXPERIMENT 1: RANDOM UNIFORM

SIMULATION UNDER THE NULL HYPOTHESIS.

In this experiment, we compare the actual observed affect to memory alignment metrics (PATE F1, PATE, Euclidean distance, and DTW) to those on random affect data with minimal assumptions (with comparison procedure described in *Figure 3.2* and *Section 3.4.2*). For this, for each of the 35 videos, we simulate affect annotations randomly drawn from a uniform distribution in the range of each affect dimension scale: 0 to 8 when simulating arousal, -8 to 8 when simulating valence and 0 to 11.3 when simulating emotional intensity. To maintain the basic structure of affect annotation, the random affect label is drawn every 15 seconds. Through this simulation, we wanted to test if the relationship that we can observe in our real data is likely random.

RESULTS.

Table 3.1.: Mean metric values for arousal (**A**), valence (**V**) and intensity (**I**). Green cell denotes a significant p-value ($p<0.004$), cell without a colour denotes insignificant p-value ($p\geq0.004$) - 0.004 is the significance level with Bonferroni correction over multiple comparisons. Arrow up (↑) means the higher the measure the more aligned the two annotations, Arrow down (↓) - the lower the more aligned.

Aff.	Mem.	Metric	Mean Observed Value			Experiment 1: H_0 means			Experiment 2: H_0 means			Experiment 3: H_0 means		
			A	V	I	A	V	I	A	V	I	A	V	I
Bool.	Bool.	PATE (F1) ↑	0.69	0.67	0.70	0.63	0.62	0.69	0.66	0.66	0.69	0.69	0.68	0.70
Bool.	Cont.	PATE ↑	0.65	0.66	0.66	0.66	0.65	0.63	0.68	0.68	0.65	0.64	0.64	0.64
Cont.	Cont.	Eucl. dist. ↓	20.7	19.4	15.9	25.1	24.6	24.9	22.9	22.6	18.7	20.8	19.5	15.9
Cont.	Cont.	DTW↓	17.7	16.2	13.1	18.8	18.3	18.4	17.7	17.1	13.9	17.5	15.9	12.9

The comparison between the observed data and the simulated random alignment metrics is shown in *Table 3.1*. For arousal and valence, all the metrics except for PATE showed that the observed relationship between affect and memory is unlikely due to chance ($p < 0.004$, which is a significance level with Bonferroni correction for 12 comparisons - 4 metrics across 3 affect dimensions). For intensity, all metrics except for PATE F1 were significant.

We used PATE F1 to compare binary memory annotations with affect annotations, binarised using a meaningful threshold - the middle of

Likert scales for each dimension of affect. This metric showed significant results for arousal and valence and insignificant ones for intensity. This means that while the relationship between observed arousal/ valence and memory is unlikely due to random chance, it is more likely due to random chance in the case of intensity. This might be due to the fact that for intensity the threshold is not as meaningful as for the measures of arousal and valence, since intensity was not measured on a Likert scale and, therefore, the threshold is not as meaningful for that dimension. Binarising intensity might remove the important trends that contain information about a moment being encoded into group memory. This is consistent with the fact that continuous PATE on intensity showed a significant result.

In contrast, the PATE metric that operated on continuous affect, and the observed relationship showed insignificant results, suggesting that the observed PATE measure is likely to belong to the same distribution as the random simulated data. This was a surprising result, which might be due to the fact that PATE which operates on continuous annotation binarises the data by generating thresholds for every observed value of the data and computes the resulting metric as the area under the curve for all those binary PATE instances. What the insignificant results might point towards is that some thresholds are more meaningful than others (this is why PATE F1 returned significant results for affect dimensions with a Likert scale). This said, PATE for intensity showed a significant result.

Lastly, the metrics that were computed based on continuous metrics of affect and memory, Eucledian distance and DTW distance, returned significant results across all three affect dimensions ($p \leq 0.004$). Judging by these distance metrics, whether or not the metric allows for shifts and stretches in the aligned data (in case of DTW), the observed relationship between group effect and memory is unlikely random. For illustration, *Figure B.1* shows the difference between the observed average DTW distance and the simulated random distribution (the lower the distance the more alignment there is between the two time-series). The top row of the figure illustrates how unlikely it is that the observed relationship would be a part of a random distribution in experiment 1 (for more detail, see figures for other metrics in Appendix B).

3.5.2. EXPERIMENT 2: RANDOM WITH OBSERVED RANGE

SIMULATION UNDER THE NULL HYPOTHESIS.

Experiment 2 is similar to experiment 1, but the data under the null hypothesis is simulated to be closer to the observed distribution of the affect values. This is done to ensure that the differences between the observed metrics and the random data are not due to the differences in the range of affect values represented in the observed data. Specifically,

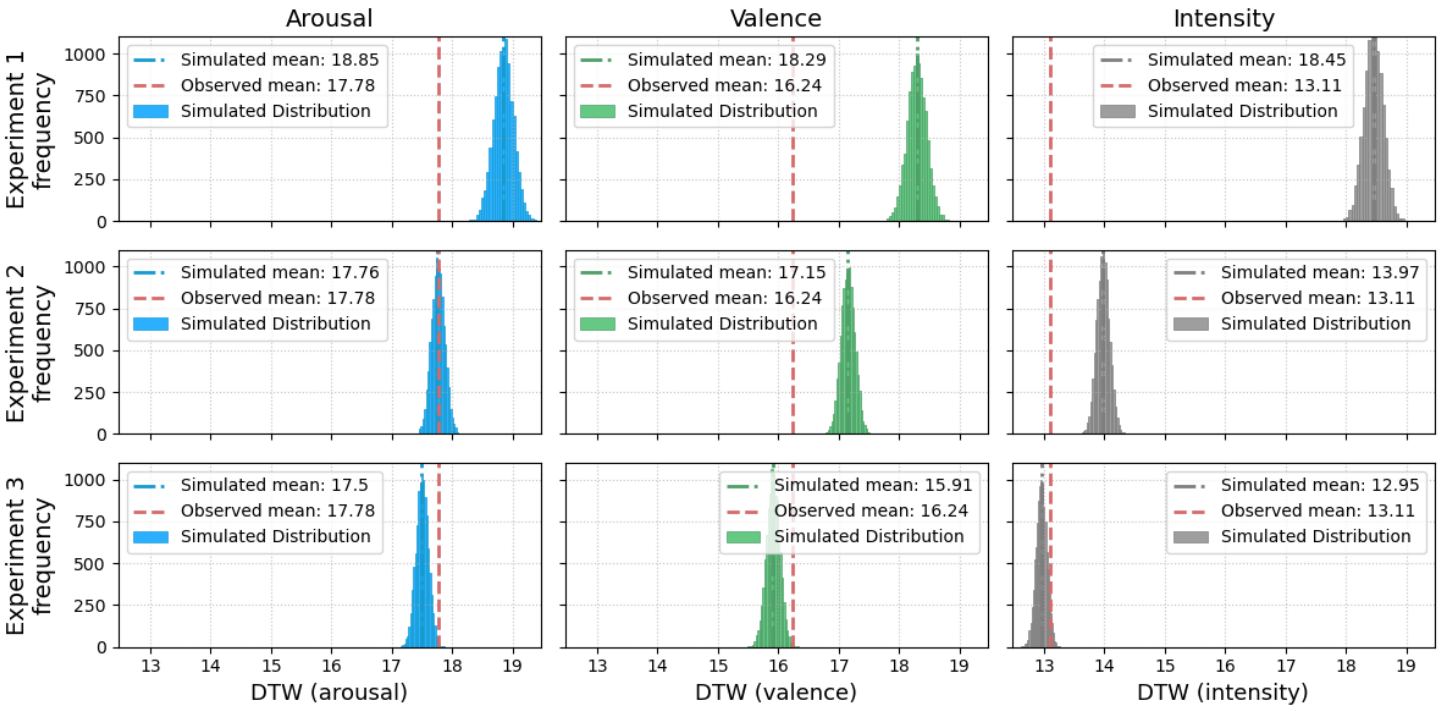


Figure 3.3.: DTW distance results for all the experiments (each row of plots shows data for a different experiment) across the three affect dimensions - Arousal (blue), Valence (green), Intensity (grey). The Colored histogram showed the distribution of averaged DTW distance values under the null hypothesis, the red dashed line shows the averaged DTW distance value for the observed data. *(Metric interpretation: The lower the DTW distance the more alignment is observed in the data.)* For other metrics, see complete figures in Appendix B.

experiment 1 assumes that the affect data ranges across all the possible values of a Likert scale for each affect dimension. This assumption does not hold for the real data. First, the emotion annotations across the dataset are skewed to the positive side: the arousal annotations range from 1 to 8 (with no instances of 0), the valence values range from -6 to 8 (with no instances of -7 and -8). Second, the range of captured emotions differs depending on the session, with the smallest range for arousal being 2 to 5 and the smallest range for valence being -3 to 2 for one session. Consequently, the full emotional range assumption in Experiment 1 does not hold for the data, which might exaggerate the differences between the observed and simulated data. Instead, in experiment 2, we simulate the data based on the affect values represented in each video. Similar to experiment 1, random values are drawn for each 15 seconds of each video, but they are drawn from the set of values present in the affect data for that specific video. For example, if an observed annotation has arousal values [1, 3, 5, 6, 8], the simulated arousal data would draw random values from that set for every 15 seconds of the video. Except for simulation assumptions, the analysis is done according to the same procedure as experiment 1 (see procedure details in *Figure 3.2* and *Section 3.4.2*).

RESULTS.

As shown in *Table 3.1*, the results for Experiment 2 are similar to Experiment 1, indicating that the observed relationship between group affect and memory annotations is unlikely to belong to the same distribution as the relationships between simulated affect and memory. This is true for all the values except for PATE F1 for intensity, PATE for all the dimensions, and DTW for arousal. In comparison to experiment 1, two more values became insignificant (DTW distance for arousal and PATE F1 for valence).

3.5.3. EXPERIMENT 3: TEMPORAL SHUFFLE

SIMULATION UNDER THE NULL HYPOTHESIS.

While experiments 1 and 2 show a comparison of observed affect to memory alignment metrics to completely random data, experiment 3 investigates the importance of temporal alignment between the affect and memory annotations. This means that in comparison to experiments 1 and 2, the simulated data in Experiment 3 represents the exact distribution of affect values in the observed data, but in randomised order within each video. In other words, we simulate data by shuffling the actual affect annotations for each video, destroying the temporal alignment between memory and affect labels. Similar to experiments 1 and 2, we are keeping the integrity of the basic data structure by shuffling windows of 15 seconds within each video. Apart from this,

the analysis procedure is the same as for experiments 1 and 2 (see procedure details in *Figure 3.2* and *Section 3.4.2*).

RESULTS.

Surprisingly, all the metrics across all the affect dimensions showed an insignificant difference from the simulated distribution in Experiment 3 (see *Table 3.1*). This means that under the assumptions of this experiment, we cannot reject the null hypothesis, meaning that the observed relationship between group affect and memory annotation is likely to belong to the same distribution as the memory compared with temporally shuffled affect data.

3.6. DISCUSSION

We conducted three computational experiments to evaluate the relationship between memory and group affect annotations. Despite significant values in the first two experiments, we can conclude that the null hypothesis cannot be rejected, since there are no metrics or affect dimensions for which all 3 experiments produced a significant p-value (see our null hypothesis rejection rule in *Section 3.4.2*).

Annotation Perspectives. One of the core assumptions motivating this study was that emotional experiences influence memory encoding, a well-established link in cognitive science [38, 45]. However, in affective computing, emotional states are often inferred from third-party annotations of observed behaviour rather than first-party reports of experienced emotions [44]. Our findings indicate that this distinction is crucial: while experienced emotions are directly tied to personal relevance and cognitive appraisal [179, 180], third-party affect annotations reflect an external interpretation of group behaviour that may not reliably map onto internal memory processes. Experiment 3 has demonstrated that the significant effects seen in Experiments 1 and 2 were likely due to the differences in distributions of affect labels in the observed data, rather than the fact that affect annotations align with memory annotations better than chance. This discrepancy aligns with potential concerns regarding third-party annotations, which are known to be influenced by external factors such as social norms and individual expressivity [196]. For instance, emotional behaviours may not always correspond to experienced emotions due to social masking, such as hiding frustration with a polite smile.

Continuous Conceptualisation. A second key issue is how affect and memory are conceptualised over time. To the best of our knowledge, traditional memory studies assess emotional experience and memorability as static states, typically using retrospective self-reports [182]. In contrast, MER applications require continuous,

time-aligned affective labels to enable real-time system responses [46]. By testing the relationship between time-continuous affect annotations and memory, we examined whether existing findings on affect and memory translate to a dynamic, multimodal annotation setting. While time-continuous annotations are widely used in affective computing to capture fine-grained emotional fluctuations [18], previous research linking affect and memory has not employed such temporally granular methods. Our results suggest that while emotional intensity, valence, and arousal may contribute to memory encoding, their influence is not reliably captured through continuous third-party annotations. This aligns with previous findings that retrospective memory reports are influenced by post-event reconstruction biases, which are not accounted for in continuous affect or memory annotation frameworks [37, 183]. The lack of a robust relationship between continuously observed affect and memory shown by our results suggests that real-time affective annotations alone may not be sufficient for predicting memory in conversational contexts, necessitating alternative methodologies that consider retrospective appraisal effects and the temporal structure of memory retrieval.

Group-Level Analysis. Finally, this study contributes to the growing need for group-based emotions and memory research. While prior work has examined emotion's role in individual memory encoding [39, 40, 172], our study explicitly considers group dynamics, a key factor in real-world settings like meetings and collaborative tasks [5, 184, 185]. Our results suggest that group emotion annotations, which capture collective emotional states rather than individual experiences, may fail to account for memorability. This could be due to the fact that group memorability annotations might not capture an emergent group-level processes in a way that group affect does [25], since they are aggregated from individual memory reports rather than inferred from group states to begin with. Another possible reason is that people express emotions differently in group settings compared to one-on-one conversations or non-social situations. For example, research suggests that emotions tend to be expressed more strongly in dyads than in larger groups [26]. Additionally, group members often adjust their emotional expressions to match each other, a phenomenon known as emotional convergence [187, 197]. This convergence may dilute individual emotional expressions, driving observable emotion further from the individual experienced emotion that would have been connected to memorability.

3.7. CONCLUSIONS

This study investigated the potential of using group emotion annotations as proxies for conversational memorability in multi-party settings. By analysing the relationship between memorability and affective dimensions using data from the MeMo corpus [36], we conducted a series of computational experiments comparing affect annotations, particularly arousal, valence, and intensity and memorability.

While the relationship between affect and memory showed to be significantly different than random (see experiment 1), when correcting for distribution biases of real affect annotations, experiments showed that such a relationship is insignificantly different from random (experiment 2) or temporally shuffled data (experiment 3). Overall, our analyses revealed that the observed metrics for real data did not deviate meaningfully from the distributions derived from synthetic data generated under null hypotheses. This finding suggests that, within the scope of this dataset and methodology, affect annotations (in terms of arousal, valence, or intensity) do not serve as reliable proxies for conversational memorability. Therefore, despite a common belief that affective states capture inter-personal relevance in alignment to memory, our findings highlight the need for dedicated research on modelling memorability - a distinct indicator of long-term event relevance.

While emotions and memory have been conceptually linked in cognitive science, our findings suggest that this relationship may not translate to the settings typical of Affective Computing applications. These applications traditionally operate with continuos affect annotations, collected from third-party observers, that rely on participants' behaviour to infer their affective states (illustrated in *Figure 3.1*). In contrast, prior research on the emotion-memory link has used static self-reports or physiological measures of emotional experience. Our study shows that observed affect annotations, particularly at the group level, do not meaningfully align with memorability, highlighting the importance of these conceptual differences between the operationalised constructs. Although the memory-emotion link has been treated as a given in some Intelligent Systems applications [41, 177, 178], we urge future research to account for differences in how these constructs are defined and measured (e.g., third-party vs. first-party, group vs. individual level, continuous vs. static). Failing to consider these discrepancies may lead to inaccurate transfers of empirical findings into computational applications.

To better understand the emotion-memorability link, we recommend further research into whether individual-level perceived affect annotations exhibit the same lack of relationship with individual memorability as observed at the group level. We also suggest examining these phenomena in face-to-face interactions rather than online video con-

ferencing to determine whether the setting influences the relationship between memory and affect annotations. Lastly, there may be differences in the types of memorable moments that are linked to affect and those that are not, warranting further investigation into the contextual factors shaping memory and affect.

4

PREDICTING CONVERSATIONAL MEMORY

This thesis chapter is an extended and substantially revised version of a previously published conference paper [1]. The extension includes a significantly extended theoretical framing of conversational memory, with an expanded literature review drawing from cognitive and social sciences, and the introduction of a new visual chart mapping the relationships between memorability, attention, engagement, and involvement. Key constructs were defined more explicitly and supported with visual aids. The reasons annotation procedure was also described in greater detail, with clearer justification and sampling details to improve transparency and reproducibility. Methodologically, the chapter introduces hyperparameter optimisation across all models, corrects the data splitting procedure to prevent session overlap, and increases the number of experimental iterations from 20 to 200, ensuring greater statistical robustness. Permutation importance analyses were replaced with ablation studies to better assess the contributions of different feature sets. These changes yielded a new set of model performance results, which, unlike those in the original version, show no clear differences in predictive power across feature groups.

[1] **M Tsfasman**, K Fenech, M Tarvirdians, A Lorincz, C Jonker, C Oertel, "Towards creating a conversational memory for long-term meeting support: predicting memorable moments in multi-party conversations through eye-gaze," in *Proc. International Conference on Multimodal Interaction (ICMI)*, pp. 94–104, 2022.

ABSTRACT

Humans have selective memory, remembering personally relevant information and filtering out the less relevant details. Understanding what makes certain conversational moments memorable could aid meeting facilitation and conversational AI. While memory modelling (computational prediction of what content people are most likely to remember) has been extensively studied in static media like images and videos, conversational memorability remains under-explored. Conversations are interactive and multimodal, with verbal and non-verbal cues shaping what individuals remember. In this paper, we approach memory encoding in group conversations from a computational perspective, analysing multimodal predictors of conversational memorability using the MeMo corpus - a dataset of recorded group discussions with memory annotations. Our study follows three objectives: (1) to model conversational memorability using verbal and non-verbal features (eye-gaze and speech activity), (2) to empirically investigate the relationship between different non-verbal features and conversational memorability, (3) to investigate the diverse categories of memorable moments for further applications. We show that gaze and speaker activity features can predict group memorability levels significantly above chance. Our empirical investigation shows which group gaze behaviours are particularly indicative of memorability levels. Our analysis of self-reported reasons for remembering a conversational episode confirms the previous research on the importance of social and self-directed memories in a conversational context. The results have practical implications for intelligent meeting support systems, personalised virtual assistants, and collaborative tools designed to enhance retention and knowledge transfer in group settings. By bridging the gap between memory modelling and conversational AI, our work paves the way for developing memory-aware systems for long-term interaction.

4.1. INTRODUCTION

Humans are inherently social beings, and their psychological and physiological well-being is deeply intertwined with the quality of their social interactions [88]. The ability to recall past conversations plays a crucial role in shaping social bonds, guiding future decisions, and reinforcing one's sense of self [11, 15, 16]. In group conversations, such as work meetings or social gatherings, remembering key moments can influence relationship dynamics, collaboration, and decision-making. However, human memory is selective: only a fraction of conversational events are retained over time [22]. While forgetting is a natural and sometimes even adaptive process, understanding which moments are likely to be remembered opens up valuable opportunities for technology. Memory modelling could support intelligent systems that help users retain socially or personally meaningful information, facilitate follow-ups after meetings, or personalise interactions based on past experiences. Despite this potential, the question of what makes certain conversational moments memorable remains largely unexplored in computational research.

Existing research on conversation analysis and meeting support systems has primarily focused on tracking and responding to immediate user states, such as affect [44, 181, 186, 188], attention [97, 198], engagement [6, 95], and social presence [99]. These approaches rely on verbal and non-verbal cues to infer real-time measures of interaction quality, often aiming to enhance participation and group dynamics. However, while these systems capture the present state of a conversation, they do not account for how users remember and interpret these experiences in the long term. Yet, memory plays a fundamental role in human communication, shaping learning, social cohesion, and continuity in interactions [62, 63]. A facilitation system that understands which conversational moments are likely to be remembered could provide more personalised and contextually aware support.

While the study of memory modelling has gained traction in other domains, such as image and video memorability prediction [35, 53, 137], memorability of conversations has not been computationally modelled before, to our knowledge. Unlike media, conversations involve complex multimodal dynamics, requiring continuous verbal and non-verbal processing [13, 33]. Previous psychological studies have identified conversation-specific interpersonal and linguistic factors influencing memory, such as self-relevance, linguistic features, and conversational role [14, 15, 17]. However, there has been no systematic effort to computationally model conversational memorability using real-world multimodal data.

This work strives to address several gaps in the literature. First, we shift the context of analysis from visual media to spontaneous conversation, a setting characterised by interpersonal coordination,

dynamic turn-taking, and embodied social cues. Second, rather than focusing on individual recall alone, we examine memory on a group level, exploring which moments are consistently remembered across participants and why. This enables us to investigate whether there exist properties of conversational events that increase their likelihood of being encoded. Finally, we adopt a time-continuous modelling approach that allows for fine-grained analysis of behavioural signals during interaction.

We focus on eye-gaze and speaker activity as primary input features for several reasons. First, these signals are continuously measurable, accessible through standard audio-visual recording tools. Second, both eye-gaze and speaker activity have been previously used to predict attention and involvement - the internal states that are known as memory modulators [58, 199–201]. While attention and involvement are not directly observable, prior work shows they can be reliably inferred from gaze patterns and speaking behaviour [56, 202]. Given their predictive value and traceability in naturalistic settings, these multimodal signals provide a promising foundation for developing intelligent systems capable of detecting and modelling shared memorable moments in conversation [19, 29, 30].

By linking real-time indicators of attention and involvement (gaze, speaker activity behaviour) to later memory outcomes, this study aims to identify multimodal cues that signal memorability and explore the technical feasibility of conversational memorability modelling in group interactions. We use the MeMo corpus [36], a dataset of recorded group discussions with memory annotations to achieve the following overarching goals:

(G1) Model conversational memorability using verbal and non-verbal features (eye-gaze and speech activity).
(G2) Empirically investigate the relationship between different multimodal features (eye-gaze and speech activity) and conversational memorability.
(G3) Analyse the reasons behind why certain moments are remembered.

By advancing computational models of conversational memory, this work strives to provide insights into human recall processes and contribute to the development of memory-aware conversational AI systems, such as meeting facilitation, automatic summarisation and memory augmentation.

4.2. BACKGROUND AND MOTIVATION

In human memory research, memory is typically conceptualised as comprising three interconnected subprocesses: encoding (processing an

experience to be preserved or forgotten), retention (the preservation of that experience over time), and retrieval (the subsequent extraction of the preserved information) [22]. Although these processes are intrinsically linked and cannot be entirely disentangled, empirical studies often choose one subprocess as the primary focus of investigation. The focus on one primary subprocess matters for the methodological and contextual choices in the study. For instance, retrieval, intentional (voluntary) or spontaneous (involuntary) is commonly studied in the context of how and when memories are accessed, such as during collaborative tasks [23]. In contrast, studies with a focus on memory encoding investigate the specific stimuli or human behaviour at the very moment of an event that would subsequently form into a memory. In other words, although memory tasks (e.g., free recall or recognition [24]) inherently involve all three subprocesses, encoding studies focus on the contextual and behavioural factors present at the time an event is experienced and the likelihood of that event being encoded or forgotten.

While modelling all three subprocesses has potential applications in intelligent systems, we argue that memory encoding is particularly useful. Arguably, encoding occurs at the time of the event or shortly after, making it possible to observe and analyse in real time. By predicting which events are likely to be encoded, we can estimate their intrinsic relevance as they happen. This opens up the possibility of identifying what makes an event memorable and enabling real-time interventions that could influence what participants will ultimately remember from a meeting. Applications could include personalised user interfaces that adapt content based on what users are likely to remember, meeting facilitation tools that emphasise key discussion points to improve retention, and conversational agents that selectively recall relevant past exchanges to enhance rapport. Memory encoding modelling has been ly referred to as ’memorability modelling’ - the likelihood of a stimulus or an event to be encoded in user’s memory [137]; therefore, for the purpose of this study we use these terms interchangeably.

In fact, stimulus memorability has been computationally modelled before, in the context of image and video memorability [20, 21, 105]. These studies typically focus on identifying stimulus-driven features (such as visual composition, semantic content, or motion patterns) that contribute to the likelihood of a media item being remembered across a wide audience. While some findings suggest a degree of universality in what makes visual media memorable [137], these models largely generalise over controlled, short-form stimuli and do not account for social, contextual, or interactive dynamics. In parallel, other work has shown that media memorability can also be predicted from neural responses such as EEG signals during viewing [20], highlighting the role of individual cognitive processing at encoding.

However, these approaches are not directly applicable as baselines in our work, for two key reasons. First, they model memory for externally presented media, rather than for personally experienced conversational moments, which are shaped by autobiographical, social, and interactional factors. Second, the features they rely on (stimulus-centred or neurophysiological) do not capture the interpersonal and temporal dynamics central to group conversation. While our dataset (Zoom recordings) does contain audiovisual data, the nature of the memory task (first-party recall of co-constructed interaction) requires modelling beyond stimulus salience. Therefore, our approach shifts focus from what makes a media stimulus memorable to what makes a social interaction memorable from the perspective of its participants.

This paper aims to analyse and model memory encoding (1) in a conversational setting, (2) on a group level, (3) with continuous annotations. These are the gaps within the existing literature that will be described in this section.

4.2.1. CONVERSATIONAL SETTING

Although the media memorability have been approached from computational perspective before, it has never been modelled in conversations, despite it being a common setting for intelligent systems (e.g. meeting support [5, 92, 97], meeting summarisation [75, 203–205], public deliberations [4, 91] or collaborative educational tools [185, 206]).

Conversations are a systematically different setting than media consumption for several reasons. First, unlike media consumption, which is largely passive, conversations require active engagement in both comprehension and production of verbal and non-verbal signals. This continuous exchange places additional cognitive demands on participants, influencing which moments are encoded into memory [13, 33]. Second, memory formation in conversations is shaped by interpersonal dynamics, such as turn-taking, speaker emphasis, and shared attention, which do not exist in media consumption even if consumed media involves a recorded conversation [12, 13, 15, 16]. Third, social factors such as emotional contagion and group cohesion influence memory retention, as shared emotional experiences can strengthen collective recall [132]. Finally, conversations often involve goal-directed interactions, such as persuasion, problem-solving, or bonding, which impact what is remembered based on its social relevance rather than its intrinsic stimulus features [11, 45].

These differences imply that the media memorability models may not translate to the context of conversations. Existing computational work on memorability has primarily focused on media that differ considerably from natural conversation, such as highly edited advertisements, static images, or curated video clips (e.g. [20, 35, 53]), rather than

on interactive, spontaneous dialogue. Highlighting this distinction is important, as a critical reader might reasonably ask: if existing models are indeed unsuitable, why was this not tested empirically? One way to address this is by emphasising the difference in task formulation. Media memorability studies typically concern semantic memory (recall of facts or content), whereas this work is concerned with autobiographical memory, in which individuals recall personally experienced conversational moments. Just as prior psychological studies have focused specifically on the mechanisms of conversational memory [12–14, 16, 33], there is a clear need for computational approaches that are likewise tailored to the distinctive nature of conversations and the ways in which they are encoded and remembered.

4.2.2. GROUP-LEVEL ANALYSIS

Many intelligent system applications operate in group settings, such as workplace meetings [5], educational discussions [184], and collaborative problem-solving tasks [185]. In these contexts, memory plays a crucial role in shaping group decision-making, knowledge retention, and long-term interaction outcomes [11]. Previous research has highlighted that dyadic interactions are different from group interactions, with distinctive patterns and emergent states [26,140,197,207]. Nevertheless, the focus of conversational memory encoding studies has only been on dyadic and not group interactions, to our knowledge [13, 14, 17, 33]. Understanding which conversational moments are collectively remembered in a group context is essential for improving intelligent systems that support collaboration, facilitate discussion, and enhance knowledge transfer in teams.

Similar to media memorability research, which aims to identify stimulus characteristics that make images or videos universally memorable [53, 105], it is valuable to understand what qualities make conversational segments consistently memorable across different participants. Unlike many media memorability studies, which often assess recall of watched stimuli, conversational memory involves remembering personally experienced interactions. As such, it is typically more autobiographical in nature, shaped by personal relevance, subjective interpretation, and emotional salience [45]. This said, by treating moments within a conversation as stimulus events, we could potentially consider their memorability as partly driven by features intrinsic to those moments. If multiple participants independently recall the same conversational moment, this may suggest that it may possess properties (such as emotional intensity, novelty, or salience) that make it more universally memorable. Thus, by aggregating individual memory reports across participants, we can approximate a measure of shared or conversational memorability, reflecting which moments are more likely to be

encoded and retained by different people. Identifying these inherently memorable event features could inform intelligent systems designed for meeting summarisation, meeting facilitation, and conversational AI, enabling them to highlight key moments that are not just personally memorable but collectively significant.

By investigating group-based memorability, this research aims to bridge the gap between individual and collective memory modelling, providing insights into how intelligent systems can better support long-term information retention in group interactions.

4.2.3. CONTINUOUS OPERATIONALISATION

A common practice in developing predictive models of internal states (e.g. affect, involvement and attention) is to collect data that operationalises user interactions through *time-continuous measurements* (e.g., annotations collected for every frame in a video stream [46]). Some of the proposed benefits [47] of this practice include its high temporal granularity (i.e., the ability to capture nuanced changes in user engagement over time) and its capacity to track dynamic variations in behaviour (i.e., identifying fluctuations in attention, involvement, or interaction patterns within a specified unit of analysis, such as a conversation segment). From an application perspective, time-continuous estimates of user states are valuable, as they enable systems to adapt dynamically to shifts in user behaviour.

While common in user state modelling, this perspective has not been applied to memory, to our knowledge. Media memorability prediction studies focus on short video fragments with one memory annotation per video [20, 21, 35, 105]. Contrary to media consumption setting, conversations in the focus of many intelligent system applications are typically long (from 10 minutes to several hours [4, 5, 75, 91, 92, 97, 185, 203–206]). Since human working memory capacity has been shown to last up to 20-60 seconds [208, 209], it is logical to expect to have more than one (non-)memorable event within one conversation. In psychological research, this has been approached from the perspective of recall rates - the percentage of idea units that were retained over the duration of the entire conversation [17, 28, 83]. Although this approach offers a valid estimation of recall quality, it is still a singular measure for the entirety of a conversation session and, therefore, is not suitable for the development of a real-time conversational memorability prediction.

Empirical studies of conversational memory commonly use self-reports that are not grounded in specific moments of encoding (i.e. recording of the interaction). The few studies that do attempt to ground these in moments of encoding [17], ask third-party annotators to relate self-reports back to the events in the interaction, bringing potential issues concerning the subjective nature of memory reports and therefore a

threat to the validity of such annotation [182]. Since memory reports can provide quite a subjective and condensed summary of the remembered event, it is highly likely that several events would match each memory description, and this could cause mistakes in the annotation. This could be avoided if the participants themselves would provide an encoded event annotation, relating their free-recall reports to specific moments of discussion they are referring to. This approach to increasing the validity of continuous conversational memorability annotation has been implemented in the dataset we base this study on [36].

In conclusion, we highlight three major gaps in previous research on memory modelling: conversational context, group-level analysis and continuous operationalisation of memory encoding labels. In this paper, we aim to fill those gaps with empirical and computational investigation of group-level conversational memorability.

4.3. THE OVERALL APPROACH

4.3.1. MULTIMODAL ANALYSIS

For continuous group memorability modelling in conversations, it is important to choose the input features that could serve as reliable predictors of memory and be usable in an intelligent system setting (e.g. meetings, deliberations, learning settings mentioned above). For this, the input features need to be continuous (see *Section 4.2.3* for why) and traceable via standard meeting software. Previous studies of continuous modelling of internal states (e.g. affect, involvement, attention) in social settings have shown that multimodal social signals, extractable from audio and video recordings, are promising for such a task [19, 29, 30].

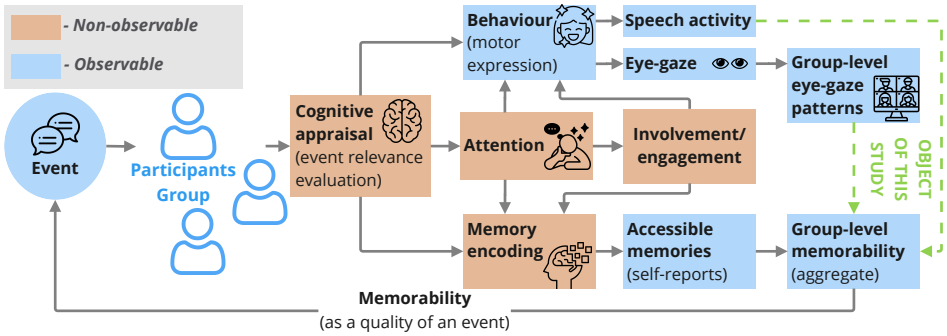


Figure 4.1.: Visualisation of the interaction between the key concepts of the present study and the main object of the study (see green dashed arrow).

Figure 4.1 outlines our overall approach, the major concepts at

operation of this study and how they relate to each other. When an event occurs in a conversation, participants in a group engage in cognitive appraisal, evaluating the relevance of the event, its' implication, coping potential and normative significance [179, 180]. This appraisal is thought to result in some memories being encoded and some forgotten. The encoded memories can then be measured through self-reports. These self-reported memories can then be aggregated to estimate group-level memorability, reflecting the extent to which an event is universally memorable for a group. This is how memory construct transitions into memorability - a quality of an event as memorable for its participants (shown by an arrow at the bottom of *Figure 4.1*). Cognitive appraisal also influences whether the event is considered relevant enough to pay attention, which in turn affects both behavioural expressions (e.g., speech activity and eye-gaze) and whether the participant will be involved/ engaged in the next event.

Cognitive appraisal, attention, and involvement play an important role in memory encoding, shaping what individuals later report as accessible memories [58, 199–201]. Since attention and involvement are not directly observable to an intelligent system, external behaviours could potentially be used as their indicators. Specifically, observable behaviours, such as eye-gaze and speech activity, have been previously shown to be predictive of the level of attention as well as involvement [17, 56, 202]. Since attention and subsequent (physical) involvement have been previously linked to memory, it is logical to assume that similar (non-)verbal signals would also be predictive of conversational memorability.

Given previous research, we hypothesise that it is possible to predict which moments will be more memorable for the group using eye-gaze and speaker activity features previously connected to attention and involvement. Therefore, the first research question (RQ1) that the present paper aims to answer is following:

RQ1: Do humans non-verbally signal which moments they are more likely to encode in their memory? If so:

RQ1(a) What patterns in eye gaze and speaker activity behaviour indicate memorability?

RQ1(b) Can group eye gaze behaviour and speaker activity be used to predict conversational memory?

We approach RQ1 from two perspectives: RQ1(a) - with empirical analysis and RQ1(b) - from a computational modelling perspective.

4.3.2. QUALITATIVE ANALYSIS OF MEMORABLE MOMENTS

Conversations are a complex setting with heterogeneous characteristics. Conversational memory holds three different functions in our social

brains - people are more likely to remember experiences that (1) reinforce or shape their self-image [15], (2) strengthen social connections [16, 62, 63], and (3) inform their future decisions, thoughts, and behaviours [11]. According to those three functions of memory, the encoded moments could be systematically different and therefore hold different characteristics, including non-verbal signals. Although widely accepted as memory functions, they have not been studied in a conversational context and therefore it is unclear which functions are most commonly reported within conversations. For an intelligent system, knowing what kind of moments are more likely to occur could help understand how to reuse those memories in future interactions. For example, a conversational AI that recognises when a memory serves a social bonding function could strategically reference shared past experiences to enhance rapport with users, while a system that detects directive memories could prioritise recalling task-relevant details to provide personalised recommendations or reminders. Nevertheless, previous research in conversational memory has not attempted to separate memorable events into such categories, to our knowledge. We approach categorising memorable moments through analysing participants' self-reported reasons behind remembering a particular segment. Through annotations of these reasons, we divide them into subgroups by underlying appraisal and memory function. Based on this data, we aim to answer the following research question (RQ2):

RQ2: What kind of self-reported reasons for remembering a conversational moment are the most common and how do they align with the previous theories on memory functions?

General contribution. In conclusion, the present study pioneers a largely unexplored topic of conversational memory modelling in group interactions. This study provides the first step on the path of characterising the multimodal features that are predictive of conversational memory. This paper introduces a novel computational framework for modelling conversational memorability, addressing gaps in previous research by focusing on multimodal analysis of memory encoding in group conversations with continuous memory annotations. By leveraging group-based memorability, this study identifies conversational moments that are consistently memorable across participants, providing insights that have the potential to enhance intelligent systems for meeting summarisation, facilitation, and conversational AI.

4.4. DATASET

4.4.1. GENERAL DESCRIPTION

For conversational memorability modelling, there needs to be a dataset of conversations, annotated with memorability labels. For memorability labels to be valid, the annotation needs to be first-party - collected from participants of the conversations themselves. This is particularly important for the validity of the memory report, since previous research has shown that observers remember conversations qualitatively and quantitatively different than participants [13, 33]. In addition, given the highlighted gaps in *Section 4.2* above, the corpus needs to have group conversations, contain continuous annotations of memorability and recordings of multimodal signals. To our knowledge, there is only one dataset that satisfies these criteria - the MeMo corpus [36] .

The MeMo corpus is the first conversational dataset annotated with participants' memory retention reports, aimed at facilitating computational modelling of human conversational memory. The dataset consists of video-call discussions in small groups over three consecutive sessions distanced 3-4 days apart. Throughout ∼45 minutes long sessions, participants discussed COVID-19 and their experiences in the pandemic. To facilitate an active discussion, each group was paired with a moderator with experience in moderating meetings, facilitating creative sessions, and conducting interviews. The discussions were conducted in English, all participants were fluent English speakers and resided in the UK. Participants were divided into groups with 3-5 participants and 1 moderator per group.

Before and after each session, participants and moderators filled in a series of surveys. The surveys included a wide range of perceptual measures, for conciseness, we only mention the ones used in the current analysis. These are described in the following subsections (for further details on the corpus see Tsfasman et al. [36]).

4.4.2. INDIVIDUAL MEMORABILITY ANNOTATION

In order to capture memorable moments from the interaction and collect ground-truth labels of when they occurred, the memory annotation consisted of two stages illustrated by *Figure 4.2* - free-recall self-reports and encoded event annotation.

Free recall self-reports. The free-recall task was the first in the post-session questionnaire, administered immediately after the session to minimise potential biases that could influence memory recall. The task was open-ended, allowing participants to recall any aspect of the conversation (see the exact task formulation in Tsfasman et al. [36]). Participants were asked to describe each remembered "moment" in their own words, without a word limit, and could report between three and ten moments (the upper limit was set to prevent fatigue and ensure

Figure 4.2.: The procedure of memory annotation after the discussion session. Free-recall reports on the right and timing annotation on the left. The moment mentioned on the screen is an example from the data: "I remember participant 2 sharing he had also started exercising during lockdown."

sufficient time for subsequent survey questions). They could proceed to the next section only if they had no additional memories to report or had already reached the maximum of ten moments. This approach aimed to capture all currently accessible memories [134] while also providing insight into how participants segment the continuous flow of social interactions into discrete, memorable events [150].

Encoded event annotation. In contrast to previous studies [17,112], participants were tasked with assigning their self-reports to specific events within the conversation. This approach was designed to ensure that the self-reported memories were linked to the corresponding encoded events, thereby preserving the validity of the memory measures. For this task, participants were provided with a link to the recorded interaction and asked to note the start and end times (down to the minute and second) for each moment they reported within the free-recall task (they were brought back to their free-recall answers without an option of changing them). They could freely scroll through the video, without the need to watch it in its entirety, to minimise fatigue. If they struggled to pinpoint a specific moment, they had the option to leave the time blank, particularly for instances tied to broader feelings or memorable moments not linked to a specific time interval. Additionally, we ensured that the free recall reports could not be altered at this stage.

4.4.3. MEMORY REASON SELF-REPORTS

After each memory encoding annotation question, participants were also asked to self-report the reason why they thought they remembered each particular moment.

4.5. DATASET PRE-PROCESSING

4.5.1. MEMORY PREPROCESSING

GROUP-LEVEL MEMORABILITY INDEX

For group-level analysis, the individual memory annotations were aggregated to a group level (see *Section 4.2.2* for the motivation behind group-level analysis).

Figure 4.3 illustrates the process of creating the group-level memorability labels step-by-step. We employed a 5-second sliding window approach and considered the memorable moment as a binary variable derived from the individual annotations (see *Section 4.4.2*). For **individual memory annotations,** each moment was represented as an array of the time slice t per participant i. It was 1 if at least half of the time slice t was included in the moments remembered by the participant i, and it was 0 otherwise.

We considered individual memorable moments as consecutive in case they overlapped in time, unless one of the moments lasted longer than half of the discussion session. All annotations encompassing more than half a session were discarded as they did not apply to a particular moment but rather "an overall feeling" of discussion.

After that, we computed the **group-level memory index** as the proportion of participants who considered each time slice t memorable.

We then divided the group memory indices into four **memorability level labels**: *zero* - if nobody remembered a slice; *low* - if > 0 and < 30 % remembered a slice; *middle* - if 30-70% considered a slice memorable; *high* - >70% reported a slice as remembered (see the last line in *Figure 4.3*). We used these memory level labels in the classification and for other analyses further on.

MEMORY REASON ANNOTATION

To answer RQ2 (see *Section 4.3.2*), self-reported reasons were manually annotated by third-party observers. *Figure 4.4* shows the developed annotation scheme. These particular labels and sub-labels were created to capture the diverse reasons why people remember conversational moments, distinguishing between different cognitive, emotional, and social factors that contribute to memory encoding. Each label represents a unique memory trigger: "Facts about others" and "Facts about the world" focus on external information, yet they differ in scope: the former pertains to interpersonal knowledge (e.g., opinions, surprising details about people), while the latter encompasses broader factual statements about events and entities. "Self perception" reflects how individuals integrate conversations into their self-concept through emotional responses or personal narratives. "Shared experience" highlights social bonding through mutual experiences, while "Meta-behaviour of other" label focuses on the observed actions or emotions of others. "Cognitive

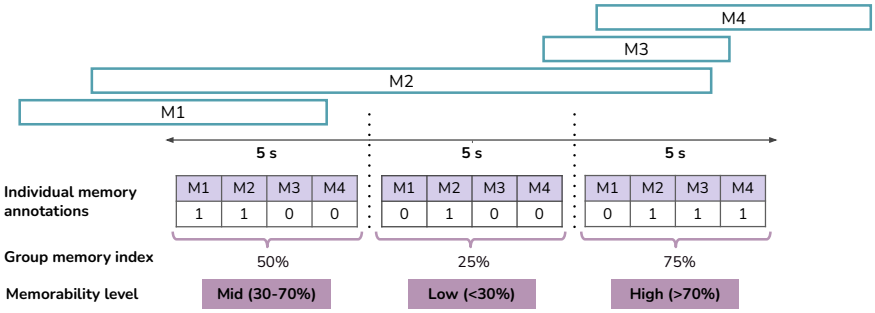


Figure 4.3.: Memorability level annotation. The blue frames on the top illustrate moments remembered by four different participants. There are three consecutive time windows on the x-axis (5 seconds each). The memorability is considered True (=1) if the moment lasts for half or more of the specific window. Therefore, M3 is 0 for the second window.

Labels	Sub-labels	Description
fact_about_others	view	agreement or disagreement with the view of another participant
	social_facts	speaker describes something about his/her activities, family or friends
	unexpected_info	speaker describes something which is unexpected from the annotator's perspective
fact_about_world	entities	view about entities such as lockdown, vaccination, etc.
	people	view about people
self_perception	annotator_feelings	feelings of the annotator such as happy, sad, embarrassed, etc.
	annotator_stories	stories or views of the annotator
shared_experience	shared_story	annotator has a similar experience or feeling with the speaker
meta_behaviour_of_other	emotional_moment	emotional moments of other participants
	behaviour	behaviour of the participants during the session such as laughter, anger, etc.
cognitive	cognitive_empathy	annotator sees the situation from the perspective of the speaker in a logical way
time_label	first	annotator remembers the first moments of the session
	last	annotator remembers the last moments of the session

Figure 4.4.: Multi-layer annotation scheme for memorability reasons with corresponding descriptions. "Annotator" is a participant for whom the segment was memorable. "Speaker" is the main speaker of the segment.

empathy" label captures moments where understanding another's perspective required active cognitive engagement. Finally, the "Time" label captures moments remembered due to primacy or recency effects, a common memory bias where people tend to recall the first and last events in a sequence more easily than those in the middle [157]. This is a multi-layer annotation scheme, since more than one of those labels

can be assigned to the same self-reported reason for remembering a moment.

These labels align with **Bluck's memory functions** [11]: directive memory supports learning from experiences (e.g., remembering factual insights within "Facts about the world" category), self memory category aids in identity construction ("Self perception" category), and social memory strengthens relationships ("Facts about others", "Shared experience", "Meta-behaviour of other", "Cognitive empathy" categories).

These categories can be described with **cognitive appraisal theory**, which explains how emotional responses emerge through primary and secondary appraisals [179, 180]. Primary appraisals (e.g., novelty, goal congruence) determine whether an event is worthy of further processing and, therefore, memory encoding, while secondary appraisals (e.g., accountability, coping potential, and future expectancy) shape how deeply it is processed and whether it affects future decisions (i.e. in how much detail it will be encoded). For instance, "Facts about others" (unexpected_info) engages novelty appraisal in the primary stage, as surprising information naturally draws attention. It can also involve coping potential or future expectancy in secondary appraisal, influencing whether the surprising fact alters future expectations or decision-making. Similarly, "Meta behaviour of others" (emotional_moment) aligns with relevance appraisal, as witnessing an emotional reaction in someone else prompts an assessment of its personal significance. It also engages accountability appraisal, since people naturally evaluate who or what caused the emotional response, influencing their interpretation and memory of the event. Together, these categories provide a structured approach to understanding memory formation in conversations through appraisal mechanisms and functional significance.

The annotation was performed by two annotators with a social signal processing research background. The annotators used reasons self-reports along with free-recall reports and corresponding encoded event video segments when assigning labels. They were instructed to prioritise the information stated in the reasons self-reports, with other information to be used as context rather than a primary source for the labels. The annotation was done using Elan annotation software [210]. To assess the inter-rater reliability, 145 samples were double-annotated. The annotators were not given information about the other annotator's labels. The inter-annotator agreement was measured using Fleiss' kappa statistics, which was found to be 0.60 - moderate agreement according to Landis and Koch [211].

4.5.2. MULTIMODAL FEATURES

SPEAKER ACTIVITY IDENTIFICATION

From the recorded audio of the discussions, we extracted active speaker information using Kaldi Speech Recognition Toolkit [152]. After conducting speaker diarisation, we extracted an active speaker array per time window t containing binary values of each participant i speaking at that time interval ($s_i(t)$). The value was 1 if the participant i spoke for at least half of the slice t and 0 otherwise:

$$s_i(t) = \begin{cases} 1, & \text{if } i \in \text{speaking}(t) \\ 0, & \text{otherwise} \end{cases} \tag{4.1}$$

We then calculated the active speaker index per time window t using the following equation:

$$S(t) = \frac{\sum_{i=1}^{N} max(s_i(t), s_i(t-1), s_i(t-2))}{N} \tag{4.2}$$

Simply put, we calculated the number of individual participants (i) that were speaking in each time slice (t) or in two time slices preceding it ($t-1$ and $t-2$) and divided that sum by the overall number of participants N in the session.

EYE-GAZE ANNOTATION

Eye gaze target extraction. Point of gaze was estimated with GazeSense software [153]. For each participant, a grid matching the gallery layout was defined based on their provided screen capture. At the start of each session, a calibration stage was performed: participants were required to fix their gaze on the screen segment containing the current target participant. We estimate the target calibration point of the i-th segment of the grid $\mathbf{p}_{grid,i}$ to be the coordinates in the centre of the segment. Point of gaze estimates $\mathbf{p}_{gaze}$ were then obtained for all remaining frames beyond the final calibration frame. The final gaze target T_{gaze} for each frame was determined as

$$T_{gaze} = \begin{cases} \alpha \arg\min_i \|\mathbf{p}_{gaze} - \mathbf{p}_{grid,i}\|, & \text{if } \mathbf{p}_{gaze} \text{ detected} \\ -1, & \text{otherwise} \end{cases} \tag{4.3}$$

Group gaze features. As illustrated by *Figure 2.5.1*, a significant body of research indicated that there is a link between attention and involvement and memory encoding [58, 199–201]. As attention and involvement are not directly observable by intelligent systems, external behaviours like eye gaze and speech activity, which are known to correlate with attention and involvement levels [56, 202], may serve as indicators of memorability. Specifically, Oertel and Salvi [56] have

shown a connection between group eye gaze behaviour and participants' conversational involvement. There was a series of group-level eye-gaze features that have been shown to correlate with perceived involvement: presence, maxGaze, entropy, and symmetry [56]. In this paper, we use the first three of these, since the gaps in data made the symmetry feature unreliable.

All the features are calculated from the gaze matrix g with $N \times K$ dimensions: N being the number of participants with valid gaze data and K - the number of targets (number of participants and an additional label for when they look away from other participants or the screen).

Individual gaze matrix g_{ij} consisted of binary measures of gaze for each time slice t. It was 1 if participant i looked at participant j for at least half of the time window t, it was 0 otherwise.

$$g_{ij}(t) = \begin{cases} 1, & \text{if } i \text{ gazes at } j \text{ at time } t \\ 0, & \text{otherwise} \end{cases} \tag{4.4}$$

Unlike Oertel and Salvi [56], a participant can gaze at themself on the screen, so there are no limitations to the value of g_{ii} in this regard. Nevertheless, since each participant could only gaze at one target at a time, the following equation applies:

$$\sum_{i=1}^{N}\sum_{j=1}^{K} g_{ij}(t) = N, \forall t \tag{4.5}$$

The speaker-directed gaze feature $f_s(t)$ was calculated to see how many participants are looking at the active speaker at any time t. It was based on matrix $s_j(t)$, which also consisted of binary measures - it was 1 if j was an active speaker at that time slice t and 0 otherwise. For each participant i and target participant j we then computed a speaker-directed gaze value S_{ij} - it was 1 if participant i was gazing towards j ($g_{ij}(t) = 1$) and j was an active speaker ($s_j(t) = 1$) at that time slice t:

$$S_{ij}(t) = \begin{cases} 1, & \text{if } s_j(t) = 1 \ \ \& \ \ g_{ij}(t) = 1 \\ 0, & \text{otherwise} \end{cases} \tag{4.6}$$

To compute the final speaker-directed gaze feature f_s we then computed a fraction of participants looking at the active speaker for each time slice t:

$$f_s(t) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} S_{ij}(t)}{N} \tag{4.7}$$

The gaze presence feature $f_p(t)$ from Oertel and Salvi [56] is the proportion of participants looking at other participants as opposed to looking away:

$$f_p(t) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} g_{ij}(t)}{N} \tag{4.8}$$

The MaxGaze feature f_m computes the maximal number of participants looking at the same target at a particular time window t:

$$f_m(t) = \frac{max_{j\in[1,K]} \sum_{i=1}^{N} g_{ij}(t)}{N} \tag{4.9}$$

The entropy measure indicates the probability of each target being looked at by all others at each particular time:

$$P(\text{target} = j|t) = \frac{\sum_{i=1}^{N} g_{ij}(t)}{N} \tag{4.10}$$

To compute the final entropy measure $f_e(t)$ the probability is then normalised as follows:

$$f_e(t) = \frac{\sum_{j=1}^{K} P(\text{target} = j|t) log(P(\text{target} = j|t))}{log(K)} \tag{4.11}$$

Therefore, it is the lowest ($f_e(t) = 0$) when all participants are looking at the same target. It is the highest ($f_e(t) = 1$) if all participants are looking at different targets.

4.5.3. MEMO SUBSET USED IN THIS STUDY

Because of recording imperfections and problems with some screenshots participants uploaded, the gaze data had 40 participants (14 groups) and 16248 individual 5-second time slices (∼23 hours). The videos were recorded with a resolution of 1280x720, with a frame rate of 25. In alignment with available gaze data, we used a subset of the data that comprised 34 hours of group discussions (42 sessions). The session duration was 45 minutes long (with a standard deviation of +- 6.6 minutes).

Within this subset, there were 53 participants (28 F, 25 M; 18-76 y. o.) and 4 moderators (3 M, 1 F; 24-45 y.o.).

There were 633 memorable moments (mean duration= 143.5, standard deviation = 183.6 seconds).

4.6. CLASSIFICATION METHODS

4.6.1. MODEL ARCHITECTURES

For classification of memorability levels, we trained standard supervised machine learning models: logistic regression, support vector machine (with RBF kernel), random forest classifier, and a multi-layer perceptron (MLP or "neural network" further). The selected machine learning methods were chosen because they are standard techniques commonly

used in the field. Logistic regression provides a straightforward and interpretable baseline for identifying linear relationships between features and group memorability. The SVM with RBF kernel is well-suited to handle non-linear relationships, making it effective for the complex and noisy multimodal data typical in social settings. Random forests, with their ensemble nature have the potential for capturing intricate feature interactions and can manage high-dimensional data, which is essential for real-world applications like meeting support systems. The MLP, or neural network, enables the exploration of deep, non-linear patterns in data, offering the potential to uncover more complex associations. As standard machine learning methods, they not only provide robust performance for this task but also offer a solid baseline for future research, enabling comparisons with more advanced techniques and ensuring the results' generalisability and reproducibility.

MLP fixed parameters. The neural network consisted of one input layer, one to three hidden layers (with optimised sizes ranging from single layers of 50 or 100 neurons to progressive architectures like (16, 8), (32, 16, 8), and (64, 32, 16)), and one output layer to enable hierarchical learning and capture complex relationships between the input features. For multi-layer architectures, the progressively smaller hidden layers helped prevent overfitting while allowing for sufficient learning capacity. L2 regularisation (alpha) was applied to reduce model complexity and prevent overfitting. We used cross-entropy loss (implicit in scikit-learn's MLPClassifier), as it helps the model optimise the prediction of categorical outcomes, such as the likelihood of a moment being memorable. The Adam optimiser was chosen for its efficiency in handling sparse gradients and adaptively adjusting the learning rate during training. ReLU was selected as the activation function for its ability to introduce non-linearity while being computationally efficient and effective in preventing vanishing gradients. We trained the model for a maximum of 300 iterations with early stopping (patience = 30), ensuring that the network converges effectively without overfitting. The validation fraction was set to 0.2 for early stopping evaluation.

Other ML models fixed parameters. Logistic Regression was configured with balanced class weights to mitigate class imbalance, a maximum of 1000 iterations for convergence, and a tolerance of 1e-4. NuSVC was configured with an RBF kernel for flexibility in capturing complex patterns, a cache size of 1000 MB for computational efficiency, and shrinking enabled for optimisation. Random Forest classifier used 600 estimators, as more estimators typically exhibit higher performance. Other than the optimised hyperparameters described below, we used default parameters defined in scikit-learn 0.24.1 [212] package.

Hyperparameter optimisation. We employed a grid search strategy with 5-fold cross-validation to optimise model hyperparameters, using balanced accuracy as the optimisation metric. We made sure that no

data from the same session was present across training and validation folds, to avoid overfitting. For the MLP, we optimised the learning rate ([0.0001, 0.001, 0.01, 0.05, 0.1]), hidden layer sizes ([(50,), (100,), (16, 8), (32, 16, 8), (64, 32, 16)]), L2 regularisation strength (alpha) ([0.0001, 0.001, 0.01, 0.1, 1.0]), and learning rate schedule ([’constant’, ’adaptive’]) to control overfitting. The progressively smaller hidden layers helped prevent overfitting while allowing for sufficient learning capacity. For Random Forest, we varied the maximum depth ([None, 3, 5, 7, 10, 13, 15]), minimum samples split ([2, 3, 4, 5, 6, 8, 10]), and minimum samples per leaf ([1, 2, 3, 4, 5]) to balance model complexity and performance. Logistic Regression optimisation focused on regularisation strength (C) (100 points from 100 to 0.001, log-spaced), controlling the trade-off between bias and variance. For NuSVC, we optimised the nu parameter ([0.5, 0.3] plus 20 log-spaced points from 0.1 to 0.01), gamma ([’scale’, ’auto’] plus 5 log-spaced points from 0.01 to 10), and class weighting ([None, ’balanced’]) to adjust margin tightness, kernel flexibility, and class imbalance handling.

4.6.2. FEATURES

The models were trained to predict the output label of one of four classes of memorability levels: zero, low, middle and high.

For the input features, we used the 5 gaze and speaker features mentioned above: gaze entropy, gaze presence, maxGaze, speaker-directed gaze, and active speaker index.

4.6.3. TRAINING SAMPLES

The train and test sets were divided into 80% and 20% respectively for all the models. The Machine Learning techniques we employ treat each data point as independent. Yet, this is not the case because of the time-series nature of our data. To ensure robust evaluation of our models, we therefore employed a session-based sampling approach with 100 re-samples. In each re-sample, we randomly split the conversation sessions (videos) into training (80%) and test (20%) sets. We then train all models on this train-test sample. Within validation folds, we also ensure that validation and training sets do not contain data from the same conversation sessions. This session-based splitting ensured that data from the same interaction session did not appear in both training, validation and test sets, preventing potential data leakage and providing a more realistic evaluation of model generalisation. This approach resulted in 100 different train-test splits, allowing us to assess the models’ performance across different random partitions of the data and obtain robust estimates of model performance variability. The performance metrics we present in the results are therefore aggregated over 100 re-samples.

Since the class distribution was severely unbalanced (zero: 8501, low: 2355, middle: 4865, high: 527 instances), in addition to a session-based split, we took an under-sampling approach. Each training set was under-sampled after being re-sampled to have an equal class representation.

4.6.4. RANDOM BASELINE

We employed a stratified Dummy classifier as our baseline, which predicts classes based on their training set distribution, rather than using uniform random prediction. We did not employ undersampling for this model in order to create a realistic baseline. This approach provides a more meaningful baseline for imbalanced datasets, as it accounts for the natural class distribution in the data and represents the performance that could be achieved by simply guessing according to class frequencies (see DummyClassifier function in scikit-learn [212]).

4.6.5. FEATURE ABLATION STUDY

In order to understand which features were most important for the model's predictions, we conducted a **feature ablation study** for all 4 models. This included removing features one at a time and training the models on the remaining features. We trained these models using the same procedure as the original models, using the best-performing hyperparameters from the hyperparameter optimisation described above. The results were aggregated over the same 50 re-samples with the same sampling procedures as with non-ablated models. We then used balanced accuracy scores to compare the ablated models to the models trained on the full feature set.

4.7. RQ1(A): EMPIRICAL RESULTS

To answer RQ1(a) (how do non-verbal behavioural features differ between more and less memorable moments in group conversations?) we investigate whether specific non-verbal cues are associated with moments of varying memorability. We approach this question from two complementary perspectives. First, in the Memory Level Analysis (*Section 4.7.1*), we examine how non-verbal behaviours (such as gaze and speaking patterns) differ across varying levels of group-level memorability. This helps us identify patterns associated with moments remembered by many vs. few or no participants. Second, in the Memory-Level Analysis Across Time (*Section 4.7.2*), we explore how these same features change immediately before, during, and after a memorable moment. This temporal perspective allows us to investigate whether certain non-verbal cues might precede or signal the beginning

of a memorable event, offering potential insights into the dynamics that lead up to memorable moments, not just their characteristics. This extensive empirical analysis aims to provide explainable insights into the dynamics of human behaviour in conversational moments of varying memorability.

4.7.1. MEMORY LEVEL ANALYSIS

To explore the relationship between memorability level and non-verbal features, we need to examine how the features can change depending on group memorability level. *Figure 4.5* illustrates means of different features and their confidence intervals for windows of different levels of memorability.

The gaze presence feature significantly differs depending on how many people reported the moment as memorable. The presence is significantly lower in highly memorable time slices (see 'high' on the y axis of *Figure 4.5*) than in low, middle and non-memorable moments ($\chi^2(3)= 750.96$, $p < 0.001$, with a mean presence score of 0.74 for 'zero', 0.69 for 'low', 0.61 for 'middle' and 0.5 for highly memorable intervals). Dunn's Multiple Comparison post hoc test indicated that presence is significantly different between each pair of memorability labels ($p<0.001$ for each combination of zero, low, middle, and high memorability labels). This means that participants looked away more than they looked at each other in highly memorable moments as opposed to other moments.

Gaze entropy (blue in *Figure 4.5*) follows a similar trend as presence. It is significantly different between the memorability levels judging by the Kruskal-Wallis H test ($\chi^2(3) = 420.65$, $p < 0.0001$), with a mean entropy score of 0.38 for zero memorability, 0.36 for low memorability, 0.32 for middle memorability and 0.23 for moments of high memorability. Dunn's post hoc test reports significant differences for all pairs of memorability levels: zero, low, middle, and high ($p<0.001$ for all combinations). It also has a similar negative trend as the group presence feature: it is significantly lower in moments of high memorability than in middle, low and zero memorability. This means participants were more likely to look at the same target in highly memorable moments and were more likely to look in different directions in less memorable moments.

A Kruskal-Wallis H-test revealed that maxGaze (green in *Figure 4.5*) is significantly different between the different memorability levels ($p<0.001$). A Dunn's post hoc test revealed that all differences between levels are significant except for moments of mid-level memorability to highly memorable ($p=0.062$) and zero to low ($p=0.004$).

Similarly, a Kruskal-Wallis H-test showed that speaker-directed gaze proportion (pink in *Figure 4.5*) is significantly different between the different memorability levels ($p<0.001$). A Dunn's post hoc test revealed

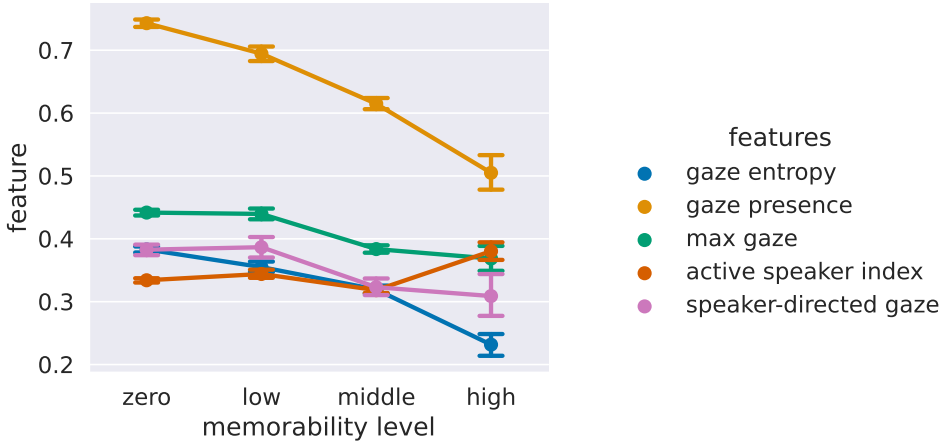


Figure 4.5.: The differences between gaze and speaker features in relation to group memorability levels. On the y-axis: points are means of the feature for specific memorability levels and 95% confidence intervals as bars. (On the x-axis: "zero" is for time slices that no one in the group recalled after the discussion; "low" are moments remembered by less than 30% of participants in the group; "middle" applies to slices remembered by 30-70% of participants; "high" - moments that 70 % or more of participants recalled)

that all different levels except for moments of mid-level memorability to highly memorable (p=0.5) and zero to low (p=0.038). This means, in the moments of high or middle memorability participants were more likely to look away from the speaker and have fewer people looking toward the same person than in moments of low or zero memorability.

Regarding the active speaker index (orange-red in *Figure 4.5*), the proportion of active speakers is significantly higher in highly memorable moments (Kruskal-Wallis H(3) = 124.89 and $p < 0.001$) with a mean active speaker index score of 0.37 for zero memorability, 0.38 for low memorability, 0.35 for mid-level memorability and 0.42 for high memorability. Dunn's post hoc test postulates significant differences ($p<0.001$) for the active speaker index for all pairs of levels, except for low vs zero memorability ($p = 0.011$). The active speaker index is significantly higher in high memorability moments than in middle, low and zero memorability ones. This means that there are more active speakers in highly memorable moments than in moments of lower memorability.

feature	χ2(3)	p-value	μ zero	μ low	μ middle	μ high
entropy	420.65	< 0.0001	0.38	0.36	0.32	0.23
presence	750.96	< 0.0001	0.74	0.69	0.61	0.5
maxGaze	399.94	< 0.0001	0.44	0.44	0.38	0.37
active speaker index	124.89	< 0.0001	0.33	0.34	0.32	0.38
speaker-directed gaze	227.81	< 0.0001	0.38	0.39	0.32	0.31

(a) Memory level

feature	χ2(3)	p-value	μ before	μ within	μ after	μ outside
entropy	46.22	< 0.0001	0.37	0.32	0.35	0.39
presence	579.74	< 0.0001	0.72	0.63	0.69	0.75
maxGaze	275.74	< 0.0001	0.43	0.4	0.42	0.44
active speaker index	64.98	< 0.0001	0.35	0.33	0.40	0.33
speaker-directed gaze	153.55	< 0.0001	0.40	0.34	0.39	0.38

(b) Timing

Figure 4.6.: Kruskal-Wallis H test results when comparing gaze and speaker features for different memorability levels (a) and different timing in relation to memorable moment (b)

4.7.2. MEMORY-LEVEL ANALYSIS ACROSS TIME

We also investigated whether any contextual cues might signal a memorable moment coming up or some changes that occur directly after the moment. For that, we compared how features changed in the two time windows before the memorable moment("BM" in *Figure 4.7*), within memorable moments ("M"), two time windows after each memorable moment ("AM"), and all other windows outside the mentioned groups ("NM"). In this case, a memorable moment is a moment remembered by at least one participant. Therefore, the comparison between the windows that fall into the categories "NM" and "M" was somewhat similar to the results described in *Section 4.7.1*. However, the difference between "NM"/"M" vs "BM"/"AM" is of greater interest, since it sheds some light on whether there might be a cue that indicates the start or the end of a memorable moment.

There were no significant differences in group gaze entropy for different timing as indicated by the post hoc Dunn's test ($p>0.001$). For maxGaze and speaker-directed gaze, there was a significant difference between during vs. before, during vs. after, outside vs. within memorable moments ($p<0.001$ in all three pairs judging by Dunn's test) but there were no significant differences between outside vs. before ($p=0.2$ for maxGaze, $p=0.9$ for speaker-directed gaze) and outside vs.

after (p=0.1 for maxGaze, p=0.8 for speaker-directed gaze). This can mean that while the lower max or speaker-directed gaze features do indicate memorable moments, there are no distinct predictive cues of a beginning or an end of the memorable moment within these features.

For the group presence measure, there was a gradual decrease from outside to right before the memorable moment and an increase from the end of the memorable moment to further outside the memorable segments. Although we can see this trend in *Figure 4.7*, the differences were significant only in the following pairs: during vs after/before/outside, outside vs after ($p<0.001$, Dunn's post hoc). Differences between outside vs. before are insignificant ($p=0.3$, Dunn's post hoc).

The most promising candidate for being a cue in signalling a memorable moment was the active speaker index (fourth subplot in *Figure 4.7*). In the time window directly preceding a memorable moment window, the proportion of active speakers significantly increases ($p>0.001$, post hoc Dunn's test). Although there was also a slight increase in the subsequent time window, this increase was not significant ($p=0.006$, Dunn's test). Interestingly, the proportion of speakers within the memorable moment did not differ from moments further away ("M" vs "NM" $p=0.5$). This finding might serve as an additional indication that, in this case, what matters is how many participants are actively involved in the discussion directly before the moment becomes particularly memorable.

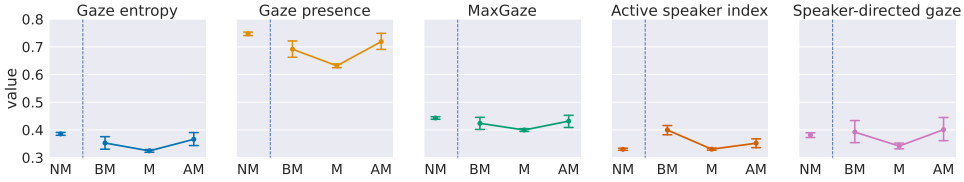


Figure 4.7.: A comparison of gaze and speaker features in different moments in relation to their timing in relation to moments remembered by at least one participant in a group. The windows within such **memorable** moments - "M" on the x-axis, two time slices **before** these moments - "BM", two time slices **after** M intervals - "AM", and all remaining time slices **not included** in the above - "NM".

4.8. RQ1(b): COMPUTATIONAL RESULTS (CLASSIFICATION)

To further address RQ1(b) (can group-level memorability be predicted from non-verbal behavioural features using machine learning?), we evaluate the predictive power of these features through supervised

classification models. While the descriptive analyses in RQ1(a) highlighted feature differences across memorability levels, this section takes a step further to assess whether those patterns are robust and consistent enough to enable computational prediction. We first test whether multiple model types can reliably classify memorable moments above a realistic random baseline. Then, through a feature ablation study, we examine which specific features contribute most to the prediction of memorability.

Algorithm	Balanced Accuracy
Logistic Regression	0.30 (± 0.06)***
Random Forest	0.28 (± 0.03)***
NuSVC	0.27 (± 0.05)***
MLP	0.26 (± 0.02)***
Dummy	0.25 (± 0.01)

Table 4.1.: Model performance metrics (mean ± standard deviation), stars after the performance metric show the significance of the difference between the model performance and the random baseline of a Dummy classifier (*** - $p < 0.001$)

Model performance. *Table 4.1* shows balanced accuracy scores aggregated over models trained on 100 re-samples of train/test split (see *Section 4.6* for more details on the method). For a realistic random baseline, we have trained a Dummy Classifier along with the main models. Paired *t*-test showed that all the models (Logistic Regression, SVM, Random Forest, MLP) performed significantly above the random baseline, with p-value < 0.001 (see a "Dummy Classifier" for comparison in *Table 4.1*).

Feature ablation study. To understand which features are most predictive of group memorability, we have performed feature ablation studies for the same models as in the main task. The ablation results are shown in *Figure 4.8*. For all 4 models, removing 1 feature at a time did not produce significant differences in the model performance, in terms of balanced accuracy (post hoc Dunn's p>0.001).

4.9. RQ2 RESULTS: MEMORY REASON ANALYSIS

To answer RQ2 (What kinds of reasons do participants give for remembering moments from a group conversation?) we qualitatively analysed participants' explanations of their own memorable moments, based on the third-party categorisation, built on theoretical work the

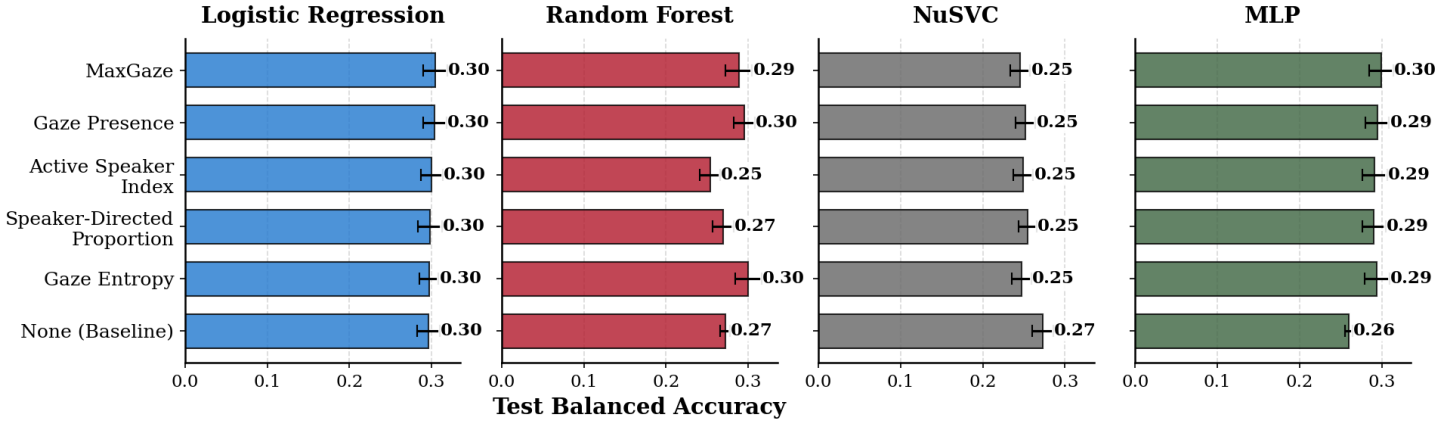
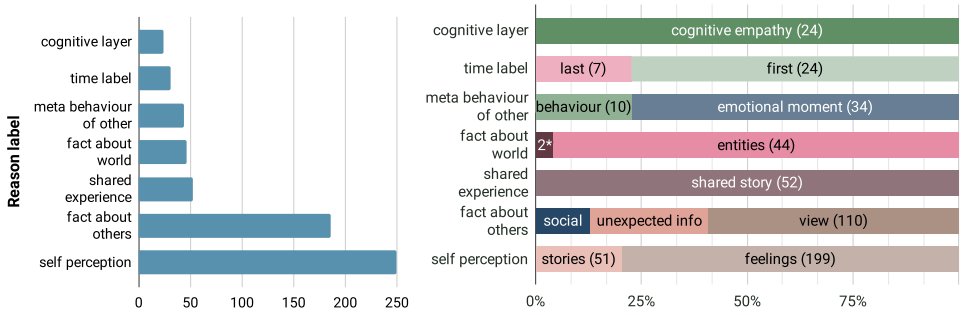


Figure 4.8.: Ablation study results for all the models: Balanced accuracy scores of models trained on different sets of features with 95% confidence intervals over 50 re-samples. "None (Baseline)" at the bottom of the y-axis shows a baseline model with all features included. The other labels correspond to the performance of a model trained on the data with a specific feature being removed.

cognitive and social functions of memory (see *Section 4.5.1* and *Section 4.3.2* for more detail).

The distribution of the memory-reason analysis is shown in *Figure 4.9a*. The most common reason was self-perception (250 of 633 memorable moments). The next frequent reason-label captured facts about other participants in the group (186). Other labels were considerably less frequent: shared experience (52), facts about the world (46), meta-behaviour of other participants in the group (44), time label (31), and cognitive empathy (24).

The sub-level distribution is shown in *Figure 4.9b*. Self-perception labels included more sub-labels related to the participant's feelings (199) than life experiences (51 reasons labelled "stories" in *Figure 4.9b*). The fact-about-other label had the majority of moments with the "view" sub-label (110 out of 186). This means that the reason for remembering the moment was related to the views of other participants in the group (for example, agreeing or disagreeing with their point of view). The second most frequent sub-label in fact-about-other reasons was unexpected information (52 out of 186), and the least frequent was social facts (52 out of 186).



(a) Number of memorable moments per label

(b) Sub-labels distribution within main reason labels

Figure 4.9.: Visualisation of reasons label distribution: main label distribution over the whole data set (plot a), and sub-labels within the main labels (plot b). Important to note that a "memorable moment" in this context is the entire memorable interval, rather than a time slice as in the statistics for the gaze and speaker features.

4.10. DISCUSSION

This study investigated how non-verbal behaviours reflect and predict conversational memory in group interactions. Our overarching research

question (RQ1) was whether humans non-verbally signal which moments are more likely to be encoded into memory. We approached this question from two angles: empirical analysis of behavioural patterns (RQ1a) and predictive modelling (RQ1b). In addition, we explored participants' introspective reasons for remembering certain moments (RQ2).

RQ1(a): Non-verbal behavioural patterns associated with memorability. Our empirical analysis revealed that participants' gaze and speaker activity indeed differ in moments of high versus low memorability. Specifically, gaze entropy, a measure of group gaze divergence, was significantly lower in highly memorable moments than in less or non-memorable ones. This suggests that participants were more visually aligned and attending to similar visual targets during moments they later remembered. This aligns with prior findings linking shared attention to joint engagement and higher information salience [56], and supports cognitive models of memory that highlight the role of attention in encoding [40, 58, 199]. Shared gaze patterns can thus be interpreted as a collective attentional marker of a salient conversational content.

We also found that speaker activity, measured as the proportion of actively speaking participants, was higher in memorable moments. This supports literature on ego-centric memory biases [12, 15], showing that individuals are more likely to remember parts of the interaction where they or others were verbally active. Interestingly, speaker activity peaked just before the most remembered intervals, echoing prior work suggesting that people are especially likely to remember reactions to their own contributions [213]. This temporal alignment suggests that pre-memory speaker dynamics may prime segments for subsequent encoding.

However, not all features conformed to our initial hypotheses. Contrary to previous research linking presence and maxGaze to high involvement and joint attention [56], these features were lower in highly memorable segments in our dataset. A similar pattern held for speaker-directed gaze. One possible explanation is that these divergences reflect a cognitive processing mechanism: gaze aversion has been associated with internal thought and memory encoding [214, 215]. Memorable moments may be cognitively demanding or emotionally salient, prompting participants to disengage from the external environment to process internally. Thus, rather than being contradictory, our findings may illustrate a shift from overt involvement to internal encoding when moments become memorable.

RQ1(b): Predictive modelling of conversational memory. IN order to examine whether gaze and speaker features can be used to predict group-level conversational memory, we trained a series of supervised machine learning models using these behavioural cues. All four tested models (logistic regression, SVM, random forest, and an

MLP) were able to classify memorable moments significantly above a random baseline (with 0.26 - 0.30 Balanced accuracy performance), demonstrating that non-verbal signals carry a meaningful signal for memory prediction. From all the models, Logistic Regression performed the best, suggesting that this relationship is best captured with a linear model.

These findings confirm that the patterns observed in our descriptive analyses are robust and consistent enough to support computational modelling. In particular, they show that even without access to verbal or acoustic information, behavioural traces such as gaze and speaking activity encode sufficient information to distinguish moments of different group-memorability levels in conversation. This supports our broader hypothesis that non-verbal behaviour reflects underlying cognitive and attentional processes involved in encoding conversational experiences.

Overall, a feature ablation study revealed that removing individual features did not significantly impact model performance across our top-performing classifiers. The most resilient to the removal of the features was Logistic Regression, with no changes to model performance when any features were removed. Other models, while affected by feature removal, also did not show significant differences in performance. This suggests that the predictive signal is not dominated by any single behavioural feature, but rather emerges from a distributed pattern across multiple cues. Another most likely explanation is that the selected features are correlated with each other, which gives the models enough information about the removed feature from the remaining ones. Since the features are computed based on the same multimodal data (gaze target data and speaker-activity data), and even when one feature is removed, there are others that could point towards the ablated feature (e.g. in case of speaker activity feature ablation, the information on that feature might be derived from the speaker-directed gaze feature).

To further explain the results, we have reviewed random example videos of correct and incorrect predictions of the best-performing model (logistic regression). Given this preliminary qualitative analysis, it seems that the misclassification might be connected to several reasons. First, technical issues in the videos that result in an incorrect prediction of eye-gaze behaviour, for instance, participants wearing glasses or not enough lighting on a participant's face at a certain time segment. Second, following a thin-slicing approach, we are dividing memorable moments into smaller segments, where each segment has the same memorability rating as the other segments of that memorable moment. This means we are treating memorability as a constant in each moment annotated as memorable. This is not always the case: for example, if there is a pause or a moment of hesitation in a longer memorable segment it would also be classified as memorable, while the non-verbal signals would indicate participants' disengagement. Last,

the neural network might have misclassified some instances because of ambiguities in non-verbal signals. For example, a segment where most participants avert their gaze would be classified as memorable, since, statistically speaking, it is a signal indicative of memorability. However, it could also be a signal of disengagement and, therefore, the lack of attention needed to memorise the moment. In this case, the segment would be incorrectly classified as highly memorable. This highlights the need for a wider context for accurate predictions of conversational memorability. Specifically, introducing additional modalities, such as speech or prosody, along with constructs such as engagement or affect could help to solve these ambiguities (see e.g., [216, 217]).

RQ2: Self-reported reason analysis. To complement the behavioural analyses, we explored why participants remembered particular conversational moments by analysing their self-reported reasons. Drawing on theoretical frameworks of autobiographical memory [11] and cognitive appraisal theory [179, 180], we developed a structured annotation scheme that categorised memory triggers into functional and emotional dimensions. This approach allowed us to investigate not only which types of conversational content tend to be remembered, but also what those memories might mean to participants, whether they support identity construction, social connection, or future decision-making.

Our analysis revealed that the most common reason for remembering a moment was related to self-perception. Participants frequently recalled segments that resonated with their personal feelings or narratives. This goes in line with the previous research on ego-centric bias [15] and means that people remember things that were personally distinct to them because of how it reflects on their personal image. This also confirms the importance of the second memory function described in Bluck et al. [11] - 'self', memories that reinforce or shape one's self-image. The next most common category involved facts about others, particularly moments that revealed another person's opinions or unexpected traits. This aligns with the social function of memory [11], where conversations are used to gather and retain interpersonal knowledge. Less frequent but still meaningful were labels tied to shared experiences, meta-behaviours of others (such as witnessing someone's emotional display), and cognitive empathy, reflecting socially rich but rarer moments of collective or emotional significance. Interestingly, factual memories, aligning with Bluck's directive memory function (e.g. with "facts about the world" label), were less commonly reported. This suggests that in natural group conversations, people are more likely to remember moments that are emotionally salient or socially meaningful, rather than simply informative.

4.11. CONCLUSIONS

In this study, we investigated whether conversational memory could be predicted from non-verbal multimodal cues, focusing on gaze and speech activity as potential indicators of memorability. Our findings showed that these features could successfully distinguish between four levels of memorability on a group level. Specifically, highly memorable moments were marked by distinct patterns: lower participant presence, reduced gaze entropy, and more active speaking engagement before and after the memorable moment. These results suggest that gaze aversion (lower presence) is a key distinguishing feature between memorable and merely highly involved moments. Additionally, the most common reasons participants gave for recalling a specific moment were related to personal feelings and experiences, highlighting the importance of social factors in memory formation. The second most common reason was information about other participants, which underscores the social and relational nature of conversational memory, rather than factual or world knowledge.

With this study we aimed to move towards filling three major gaps:

1. **Conversational setting**: Previous research on memorability has primarily been limited to static media, such as images and videos, and has not explored the dynamic, interactive nature of conversations. This study extends memory modelling into the conversational setting, acknowledging the unique challenges posed by ongoing verbal and non-verbal exchanges. Conversations, unlike media consumption, are inherently social and active, and memory formation is influenced not just by the intrinsic characteristics of the conversation but also by interpersonal dynamics such as turn-taking, emotional contagion, and group cohesion. Our findings suggest that group behaviour in conversations can be used to predict memorability levels of the conversational segment. It also confirms the particular importance of social and self-directed memories in a conversational context.

2. **Group-level analysis**: Much of the existing research on memory encoding has focused on dyadic interactions, where memory is typically studied in one-on-one conversations. However, many real-world applications of memory-aware systems, such as meeting support and collaborative learning, involve group interactions. By investigating memory at the group level, this study bridges a significant gap in the literature. We demonstrate that group-level non-verbal behaviour patterns can be used to predict aggregated memorability levels.

3. **Continuous operationalisation of memory encoding & multimodal analysis**: A key novelty of this work lies in the continuous

operationalisation of memory encoding annotation. This approach is essential in dynamic environments such as meetings or deliberations, where multiple moments within a single conversation may be memorable. Although previous research has analysed conversations continuously [17], they have only focused on the verbal information within the conversation. While verbal information can only provide information about the active speaker, non-verbal signals provide a continuous signal for both speakers and listeners. In this paper, we showed that such non-verbal features as gaze can indeed be successfully used to predict group memorability.

Overall, conversational memory modelling could show promise for interactive systems in various applications. In user modelling, current adaptive systems track engagement and mood [44, 206, 218] but overlook memory, crucial for long-term personalisation [64]. Integrating memory would help differentiate transient from lasting experiences, enabling more tailored interactions. In meeting facilitation, existing tools track verbal and non-verbal cues [4, 6, 95] but lack conversational memory. Detecting shared recollections could strengthen social bonds and common ground [63, 71]. Conversational agents also struggle with long-term rapport, recalling full conversations instead of selectively remembering socially relevant details, which can hinder trust and engagement [61, 65, 163]. Memory-aware systems could foster deeper connections and more meaningful interactions.

5

DISCUSSION & CONCLUSION

5.1. CONTRIBUTIONS AND FINDINGS

While human memory has long been studied in controlled environments, understanding how memory functions in natural conversational settings presents unique challenges. Group conversations, which involve multiple speakers and dynamic exchanges, offer a rich context for examining how different factors (e.g. affect, engagement, and the way content is communicated) can influence what is remembered. This thesis aims to provide insights into how memory can be predicted in context-rich real-world interactions.

Constructing a dataset for conversational memory prediction. While existing approaches to socio-cognitive computational modelling often leverage annotated datasets, the unique requirements of memory research (e.g. self-reported annotations and consistent temporal frames) make it impossible to rely on third-party annotations or adapt existing datasets. Furthermore, no existing dataset captures memory annotations within ecologically valid, multi-modal, multi-party conversational contexts, which are essential for understanding memory processes in real-world interactions. *Chapter 2* addresses this critical gap in the availability of datasets for conversational memory prediction by describing the structure and construction of the MeMo corpus - a multimodal conversational dataset annotated with first-party memory reports. The setup and construction ensure ecological validity. *Chapter 2* thereby answers research question RQ1 in the following way:

RQ1: How can a multimodal conversational dataset be designed to validly capture first-party memory reports to support computational modelling of memory processes in multi-party meetings?

Main takeaway of *Chapter 2*: We address this question by designing and collecting the MeMo corpus using three key principles. First, to ensure ecological validity, we recorded naturalistic, small-group conversations in a typical online meeting settings of Zoom across three sessions over two weeks. Second, to reliably capture first-party memory, we collected individual memory reports after each session and later asked the participants to relate the memory reports to specific events in conversation recordings, ensuring the memory annotations provided first-party continuous ground-truth memory labels directly usable for computational modelling. Third, we ensured construct validity of memory and behaviour measures by using established perceptive measures, integrating them with multimodal signals (audio, video, gaze, and annotations), and aligning them with theoretical frameworks from cognitive psychology. The resulting MeMo corpus consists of 31 hours of multimodal data and offers a rich foundation for modelling conversational memory, with a particular emphasis on real-world meeting setting and group behaviour.

By presenting this novel dataset, we hope to have paved the way for further research on conversational memory modelling, providing a foundation for computational modelling of verbal and non-verbal cues associated with memory and enabling the development of predictive tools for applications, such as meeting support systems.

Identifying if observed group affect annotations can derive memory annotations. Building on the development of the MeMo corpus in *Chapter 2*, *Chapter 3* investigates whether time-continuous, third-party annotations of perceived group affect can serve as an effective proxy for group memorability in naturalistic conversations. Given the substantial effort required to obtain first-party memory labels, which limits scalability, this chapter explores the potential of leveraging affective computing resources as a more practical alternative. While prior research highlights the influence of emotion (in terms of arousal and valence) on memory encoding, it remains unclear how continuously annotated group-level affective signals relate to collective memory in real-world settings. To address this, we examine the extent to which group affect annotations correspond with group memorability labels derived from first-party memory reports. Through this investigation, we aim to clarify whether dedicated memory-specific annotations are indispensable or if observed affective cues, captured through established computational methods, can reliably predict conversational memory in team contexts, thereby addressing RQ2:

RQ2: To what extent do third-party, time-continuous annotations of perceived group emotions (arousal and valence) predict group-level memorability in unstructured, naturalistic conversational interactions?

Main takeaway of *Chapter 3*: This chapter found that although third-party group affect annotations (arousal, valence, intensity) show some relationship to conversational memorability beyond chance, this association largely disappears when controlling for distributional biases and temporal alignment. These results suggest that continuous, observer-rated group affect signals do not reliably serve as proxies for group-level memory in naturalistic conversations. Consequently, dedicated memory-specific annotations remain necessary, and future work should carefully consider the conceptual and methodological differences between affect and memory constructs in affective computing contexts.

The findings imply that relying on third-party continuous affect annotations as a substitute for personal importance or memory labels in conversational AI or affective computing systems may lead to inaccurate predictions of memorability in group discussions. This underscores the need for developing dedicated methods and datasets focused explicitly

on memory modelling, rather than assuming affect can serve as a sufficient proxy. Moreover, it highlights the importance of aligning annotation perspectives and temporal frameworks when integrating cognitive constructs like memory into intelligent systems.

Identifying multimodal predictors of conversational memorability. While prior work has successfully used verbal and non-verbal behaviours to infer socio-cognitive states, such as affect, involvement, and dominance, to our knowledge, there were no existing studies that explore which behavioural signals may be informative for predicting memory encoding and retention in conversational contexts, and whether the conversational memory prediction task is feasible overall. *Chapter 4* addresses this gap by investigating the feasibility of using group eye gaze and speaker activity as behavioural predictors of conversational memorability. Specifically, we use the MeMo corpus (introduced in *Chapter 2*) to explore whether these signals might serve as meaningful features for future computational models of memory in interaction.

The first aim of *Chapter 4* is not to propose a definitive model, but rather to explore the potential of these behavioural signals as predictors, based on their known links to attention and involvement. The main answer that *Chapter 4* gives to RQ3 is the following:

RQ3: Can non-verbal behaviours, such as group eye gaze and speaker activity, serve as indicators of which conversational moments are more likely to be encoded in participants' memory? If so, what specific patterns in these signals predict conversational memory?

Main takeaway: Our research finds that group eye gaze synchrony and speaker activity are feasible predictors of group memorability in conversations, with standard computational models performing with above-chance accuracy. Moments with higher gaze synchrony, where multiple participants focus on the same speaker or event, are more likely to be remembered, supporting the role of joint attention in memory encoding. Additionally, speaker activity patterns (e.g. sustained speaking turns and dynamic turn-taking) are linked to memorability, suggesting that conversational engagement and emphasis influence what is remembered. These findings indicate that non-verbal coordination, particularly through shared attention and speaker dynamics, plays a key role in encoding memorable moments in group interactions.

In addition to exploring behavioural indicators of conversational memorability, we sought to understand the subjective motivations behind what people remember. While earlier parts of the study focus on observable signals of memory encoding, this second component addresses the question of why certain moments are retained. Specifi-

cally, we examined participants' self-reported reasons for remembering specific conversational segments to investigate whether these align with established cognitive and social functions of memory, such as self-relevance, social connection, or future utility [11]. Understanding these motivations offers insight into the underlying goals of memory in conversation, which can, in turn, inform which types of moments might be important to model or support in interactive systems. Following is RQ4 and the main answer to it, based on our research presented in *Chapter 4*:

RQ4: What kind of self-reported reasons for remembering a conversational moment are the most common?

Main takeaway: The findings indicate that the most frequently cited reason category for remembering specific segments of the conversation was "self-perception", coherent with the findings of the main roles of memory [15]. The second most-reported reason category was "capturing facts about other participants", which is also coherent with previous research on the main goals of memory being social [11].

By tackling these questions, this chapter lays the groundwork for understanding how multimodal signals can be used for group-level memory encoding prediction and what drives the personal relevance of memorable moments in conversations.

5.2. POTENTIAL APPLICATIONS AND SOCIETAL IMPLICATIONS

5.2.1. USER MODELLING

When constructing an intelligent system suitable for long-term interaction with a human, the success of a system can depend on how well it can learn and adapt to a specific user [64]. The field of user modelling focuses on building a rich, adaptive user profile suitable for subsequent personalisation of intelligent systems to a specific user [64, 219]. Updating information about the user could be done through explicit questions to the user, but this could interrupt the interaction with the system and, therefore, usually cannot be done throughout the whole interaction without causing disruption in the use of the system. This gives rise to real-time tracking of user internal states, such as engagement [206], affect [44] and mood [218] throughout the interaction via verbal and non-verbal signals coming from the user. The most commonly modelled internal states, such as affect, engagement, attention or mood, can be used for user preferences within a particular interaction with the system [7, 9, 99]. However, in a long-term interaction, it is also important for the system to understand, more globally, what parts of the interaction stay with the user even after the session is complete. In other words, when building an intelligent system suitable for long-term interaction, it is important to model what the user is more likely to remember from previous interactions with the system. This is important because, unlike machines, humans do not retain or weigh all parts of an interaction equally. While a system can access the full interaction history at any time, humans have selective memory, with only some moments encoded, retained, and later recalled. Ultimately, these remembered moments are what shape the user's long-term experience with the system. Therefore, memory modelling is essential for user modelling in the context of systems interacting with a user more than once. In group interaction, it is not only individual users one-by-one that a system would need to adjust to, but also the dynamics of the group as a whole. The system needs to be personalised to the group as a whole, its preferences, needs and behaviours, which is not necessarily equal to the addition of those characteristics for each participant, but might emerge from a unique combination of participants in the group [27]. Group-level memorability modelling presented in *Chapter 4* could be used for this application. The MeMo dataset presented in *Chapter 2* allows for both group-level and individual-level memorability modelling to advance user modelling applications.

5.2.2. FACILITATION

In a world where loneliness has become a pressing societal issue [66, 67], impacting both mental and physical well-being across our society [68, 220], solutions offering to improve the quality of human-human relationships are essential [221–223]. This is particularly important since the feeling of loneliness is not solely determined by the quality of human connections, but rather only by their quantity. Relationship quality hinges on feeling understood and appreciated within social relationships in one's life, with misunderstandings and conflicts diminishing it [70]. Conversational facilitation is a tool that can be useful not only on the level of improving one meeting at a time but also deepening conversations and improving the overall quality of relationships between people.

Existing meeting facilitation systems can enhance social interactions by tracking users' non-verbal and verbal signals in real-time, leading to improved satisfaction, equal participation, and decreased social inhibitions [4, 6, 95]. Although not investigated before, understanding the intricacies of human memory could be essential for facilitating meaningful social interactions. By detecting and analysing encoded conversational segments across users, facilitation systems can foster social bonding, establish common ground, and guide participants towards mutual understanding and appreciation [62, 63, 71]. Thus, integrating memory modelling into facilitation systems holds significant promise for enhancing social interactions and strengthening human relationships over the long term.

5.2.3. CONVERSATIONAL AGENTS

Conversational agents are being continuously developed for various applications: tutor agents designed to train social skills [224], moderator agents that can prevent and resolve conflicts [225], and conversational agents that act as human companions in times of need [226]. However, while these systems have shown promise in the short term, sustaining their impact over time is still a challenge [163]. State-of-the-art longitudinal experiments show that over time, users' engagement with the existing intelligent systems diminishes, resulting in a weakened connection and a decline in the positive effects of the interventions [164, 227–229].

One potential reason for conversational agents struggling in long-term interaction with the user could lie in the role memory plays in conversations. In human-human interactions, memory is closely intertwined with social cognition - through shared memories, humans create and maintain social bonds [62] and identify as part of a certain social group [63]. The ability to remember and recall memories from previous interactions in humans correlates with how well they perform

in a conversation and their perceived social intelligence [16]. Similarly to an interaction with another human, in a repeated interaction with an agent, a user expects the agent to remember and reuse experiences from previous interactions [72]. One could argue that in the absence of memory, a user-agent bond is not possible, which in turn can hinder long-term engagement with an agent [164]. In other words, for a successful long-term interaction with an agent, it should possess the ability to recall previous sessions and leverage shared memories with the user when relevant [162]. In fact, leveraging shared memories of past interactions has already been shown to improve long-term human-agent interaction [73, 165, 166, 230]. Similar to how humans with better social memory are perceived by other humans [16], agents with memory for past interactions are generally perceived as more understanding and socially present than agents without such ability [73, 166].

While these results of memory systems for conversational agents are widely positive, they are not based on how humans recall conversations and misuse the term 'shared memory', assuming the entire past conversation is in the user's memory and is 'shared' with an agent. In reality, if the memories are forgotten or not recognised by the user, bringing them up might not induce any positive effect, as shown in [74, 231, 232]. This discrepancy might come from the fact that while an agent can memorise an entire conversation, a human would only retain some part of that information since, unlike agents, humans have a highly selective memory [61, 65]. Therefore, only certain (if any) episodes from the previous interaction would be in the shared memory between the agent and the user. This misalignment between what an agent believes is a shared memory and what the user actually remembers might prove to be a problem, especially when transferring to a less constrained in-the-wild setting, with a wider range of topics or with frequently repeated interactions over longer time frames [164]. In other words, although these state-of-the-art agents are able to refer to previous interactions, they do not possess the ability to identify memories that are truly shared with the user and are more likely to lose the connection with the user over time, similar to Campos et al. [74] or Croes and Antheunis [164].

To truly leverage shared memory with the user and improve the chances for long-term interaction success, a conversational agent needs to understand which knowledge and experiences the user is more likely to retain from the previous interaction. Therefore, an essential step before designing a memory model for an agent is to first understand which conversational segments are more likely to be retained by the user, why, and when to bring them up. Here is where the MeMo corpus and conversational memory models developed on its basis can be of help. Aside from aligning the 'true' shared memories for conversational agents, a model of the user's conversational memory can help advance

the agent's theory of mind and perception of common ground.

In addition, this topic can be further explored from one-on-one interview data collected after the three group discussions. During the interview, the moderator asked each participant about the preferred capabilities of memory for an agent that is meant to support public discussions. This qualitative data can be used as a basis for a user-memory-informed conversational agent memory design in similar contexts. This data is a part of a conjoint project and, therefore, is not included in the publicly shared MeMo corpus, but is available upon request.

5.2.4. SUMMARISATION AND NOTE-KEEPING

Another topic that conversational memory models hold significant potential for is meeting summarisation applications. The likelihood of information being retained by each team member can be used as a more personalised alternative to other measures of individual or group moment importance (e.g., [50] or [75]). By identifying and prioritising segments with higher probabilities of being retained by participants, summarisation algorithms can focus on extracting and emphasising the most salient information for each participant or group as a whole. This approach enables the creation of more concise and relevant summaries that capture key points and discussions, enhancing the efficiency of post-meeting communication. Additionally, understanding which segments are more likely to be remembered allows for targeted interventions during meetings, such as repetition or elaboration, to reinforce important concepts or decisions. Overall, integrating memory likelihood tracking into summarisation or automatic note-keeping applications can lead to more tailored and impactful meeting summaries that better serve the needs of participants.

5.2.5. MEMORY AUGMENTATION

Finally, our research on conversational memory modelling could aid memory augmentation applications. The amount of content stored in a person's digital footprint is only growing - daily photos, videos, messages, meetings. To help users deal with the amount of the stored data, memory augmentation tools are developed to extract relevant content when needed [76]. When it comes to extracting relevant moments from work meeting footage, the annotation of relevance for the user is currently based on text characteristics or frequency of topic repetitions [77]. While useful, these measures are too generic to measure what each participant would find relevant enough to remember from each interaction. For this purpose, it would be useful for a memory augmentation system to be able to predict what a person would recall from an interaction for two purposes:first, memorability indicates

personal relevance of a moment that a human is likely to want to replay later, after some time has passed. Second, if a system is able to predict what a person would recall after a meeting, it could augment human memory with moments that seem important but were not captured by a participant's memory.

Memory augmentation tools are also useful for humans with memory disorders. With an ageing population, this application has become particularly important (e.g. increasing numbers of people with Alzheimer's disease, Dementia, Parkinson's disease [78]). To help humans with such a disorder, researchers develop life-logging systems that record users' daily activities (e.g. with a wearable camera) and extract the relevant events when needed. Since the amount of data recorded by such a device could be increasingly large, evaluating the relevance of each event is highly important for the usability of such a system. We suggest that memorability modelling has potential for advancing such systems. For example, the system could use data from a healthy individual on which moments are more likely to be memorable (e.g. MeMo corpus described in *Chapter 2*) and prioritise those moments when supporting a person with memory disorders.

5.3. LIMITATIONS

***Chapter 2*:** All the studies in this thesis are based on the MeMo corpus (described in *Chapter 2*). Like any dataset, the MeMo has its limitations. First, the corpus is confined to online video-call settings and a single topical domain (Covid-19), which may not fully represent the range of conversational dynamics present in face-to-face or other contexts. Second, the memory annotations methodology used in the MeMo corpus could be limited by the assumption that each free-recall self-report (FRS) corresponds to a single encoded event annotation (EEA). In practice, participants might merge multiple similar events into one report, or they may simply fail to report or precisely annotate all memorable moments due to the cognitive demands of the task. Furthermore, although the first-party memory annotations are closer to the experienced state, not all encoded memories might be accessible at the time of the survey, and, therefore, the memory annotation only captures a subset of encoded memories that were accessible to participants at the time of the survey. The corpus also does not record measures of participants' fatigue at the time of memory annotation, which might affect the completeness of memory reports and accessibility of memories. Finally, having a trained moderator may create a sense of hierarchy in the group and introduce the confounding variable of different moderation styles, potentially affecting the discussion structure and group dynamics. A different setting might lead to different results, depending on the environment, goals, and roles in the conversation. These limitations extend to this thesis as a whole, since all the studies are based on the MeMo corpus. In future work, it would be great to reproduce the results on another corpus annotated with memory, maybe in the offline setting or a real-world work meeting setting.

***Chapter 3*'s** study design also had its' own limitations in addition to the ones described above. When comparing two types of continuous annotations (memory and affect), we have assumed that if the affect and memory are related, they would have a linear relationship with each other. This might not be the case, and non-linear models might show more of a connection between the two phenomena. Another limitation is connected to the reliance on third-party affect annotations as proxies for experienced emotional states. While these observed annotations provide valuable insights into the group's collective emotional atmosphere, they may not accurately reflect the internal, experienced affect that underlies memory encoding. This discrepancy between observed and experienced affect could dilute the predictive power of the models, underscoring the need for future research to incorporate more direct, physiological, or self-reported measures of affect.

Memorability modelling in ***Chapter 4*** also comes with its limitations. One major limitation is that the analysis focuses exclusively on group-level predictions, leaving individual-level memorability analysis

and prediction unaddressed. As such, the models may overlook the variability in how different individuals encode and retain conversational events. Additionally, the features studied in *Chapter 4* are limited to group eye-gaze behaviour and speaker activity. There could be other multimodal features promising for memorability prediction that have been previously used in other social signal processing applications, such as prosody [216], laughter [217], and participants' individual [233] or group characteristics. Furthermore, the predictive models employed in this work are relatively simple and may not capture the complex, non-linear patterns inherent in human memory. More sophisticated time-series models, as well as online prediction frameworks, would likely improve accuracy and relevance in real-world meeting support applications.

5.4. ETHICAL CONSIDERATIONS

Privacy and consent in memory-related inference systems. Developing systems that predict what individuals are likely to remember based on their verbal and non-verbal behaviour raises significant concerns around privacy, autonomy, and consent. Unlike traditional observable behaviours, memory is an internal cognitive state that individuals may not wish to reveal or have inferred without their explicit awareness. Predictive memory systems, when deployed without appropriate safeguards, may be perceived as invasive or manipulative, particularly in sensitive contexts, such as workplace meetings or educational settings. Therefore, any deployment of such models must ensure clear communication of their purpose and scope, secure informed consent, and provide individuals with the right to opt out of memory-related profiling.

Risks of misinterpretation and misuse. Even when developed with good intent, memorability prediction tools could be misused or overinterpreted in real-world scenarios. For example, organisations might use such systems to assess attention or learning outcomes in job interviews or exams, without understanding the nuanced factors influencing memory. Because memory is influenced by individual traits, emotions, social dynamics, and prior knowledge, treating predicted memorability as a definitive measure could lead to unfair decisions. In addition, predictive models trained on predominantly neurotypical populations may fail to account for the cognitive and behavioural diversity found among neurodivergent individuals, such as those with ADHD or autism spectrum conditions, whose memory patterns, attention, and expressive behaviours may differ systematically. As a result, such models risk embedding normative assumptions and reproducing discriminatory outcomes if applied uncritically. It is important to consider the probabilistic and limited nature of these predictions to avoid their use in high-stakes decisions without robust validation across diverse populations and with appropriate ethical safeguards in place.

Responsible development and limitations. While this thesis presents novel contributions towards understanding and modelling conversational memory, it does not aim to produce deployable systems. The work is intended to support future research under responsible AI development principles, with careful reflection on societal implications. As the field progresses, developers should be guided by fairness, transparency, and inclusivity when designing memory-aware technologies. Future work must consider the broader ethical landscape, including potential unintended consequences, and involve collaboration between ethicists, legal scholars, and human-computer interaction researchers to ensure that these technologies serve users' interests without compromising dignity or autonomy.

5.5. FUTURE RESEARCH

In this section, we summarise several potential areas of research and applications for which the research and data collected within this thesis could be useful.

5.5.1. FURTHER MODELLING CONVERSATIONAL ENCODING AND RETENTION FROM SOCIAL SIGNALS

As described in *Chapter 2*, the first and primary area that the MeMo corpus can be useful for is modelling human conversational memory from (non-)verbal behaviour of the participants during the encoding of a moment. Although in *Chapter 4* we have already presented models that perform above chance in predicting memory encoding in terms of group memorability levels, there is more work to be done until such a model can be used to predict conversational memory in the wild. In our envisioned application, such models would take verbal and/or non-verbal features as input, potentially enriched by interpersonal and intrapersonal traits from questionnaires, and output either (a) a binary label of whether a timestamp was encoded, or (b) a continuous estimate of how many participants remembered a moment, as used in Chapter 4. Other modelling directions, such as predicting retrospective memory or memory decay over time, remain unexplored. Developing predictive models for conversational memory "in the wild" will require technical advances beyond the scope of this thesis. Future work should draw on dedicated technical research in multimodal memory modelling, which addresses the development and evaluation of sophisticated model architectures tailored for social and cognitive prediction tasks. The MeMo corpus could be used for these tasks. The **input features** for these models could include any combination of non-verbal or verbal features, as well as inter- and intra-personal characteristics of participants collected in pre- and post-questionnaires. Based on our annotation of memory (see more detail in *Chapter 2*), the **output** variable in such a model would be either (a) a binary label of whether a timestamp was in the subset of retained moments or (b) a cumulative value of the percentage of participants that had the timestamp within their memorable moments (similar to *Chapter 4*). Another possible option for an output variable could be a textual description from the participants' memory reports.

As mentioned before, memory can be broken down into three subprocesses - encoding, retention and retrieval. Since memory representations are not directly observable, modelling all three processes relies on memory retrieval tasks (e.g. recognition tasks or free-recall reports) and the three subprocesses are never completely separable from each other. However, the main focus can be on one of the three.

As such, the primary focus of the MeMo corpus is on encoding and

retention of conversational segments (see *Chapter 4*), but retrieval can also be studied with some additional annotations. Modelling the spontaneous retrieval of memories from previous sessions is possible since the MeMo corpus contains longitudinal data from 3 conversations spaced by 3-4 days in each group. The research question could be: What kind of memories are more likely to be revisited within the next conversations and how are they used within the conversation? For this research to be possible, one could annotate the moments of the conversation when events from past interactions are verbally retrieved by participants. For computational modelling labels, an annotator could go through these recordings of past interactions and relate the retrieved events to when they were first mentioned in the conversation. This could be a promising area of research for further development of automatic meeting support systems to be able to predict which moments of the previous interaction are particularly relevant within future conversations.

To model how humans retain conversations long-term, it is also possible to use the MeMo corpus with the long-term memory reports. Modelling retention could imply modelling the forgetting curve of a conversational segment with two recall reports: shorter term - straight after the interaction, and longer-term - 3-4 days after the interaction, when most forgetting would have occurred and only the most persistent memories would have stayed [34]. These persistent memories could be considered the most personally relevant for the participant. Predicting those could be used to further specify user profiles or understand the common ground or shared memory within the group.

To model the reasons why conversational segments were encoded and retained, the MeMo corpus contains self-reported reasons why participants recalled each moment. The reasons were then categorised by two annotators with the frequencies reported in Tsfasman et al. [60]. In future research, it would be interesting to further analyse and predict the types of reasons that participants report for considering a moment memorable. This could be a separate modelling task within remembered moments. Possible research questions could be the following: Are memorable moments encoded differently depending on different types of reasons for remembering the moment? Can we automatically predict the type of reason using (non-)verbal features?

5.5.2. FURTHER EMPIRICAL RESEARCH OF CONVERSATIONAL MEMORY AND INTER/INTRA-PERSONAL FACTORS

Previous research on conversational memory mainly focused on factors affecting recall during language comprehension and production tasks. It is common to have controlled experiments focusing on the memory of individual words or sentences [15, 16, 215]. Because of this controlled manner of experiments, these studies might not extend to real-world

situations, such as a spontaneous conversation. This said, some studies did focus on free-flowing conversations [17, 83], proving there are linguistic predictors of a moment that is to be retained by a participant in dyadic conversations. However, individual factors that could be of importance were not investigated in a spontaneous setting. This leads to the question: how do participants' individual characteristics (e.g. personality, values) influence the quantity/type of moments they encode, retain and retrieve in following conversations?

Conversational memory also has not been studied in group interactions. From the studies of group perception measures on learning, it can be inferred that group entitativity, cohesion and rapport that positively affect learning [131, 132], could also have a positive effect on memory. Also, since memory supports social bonding [63] and cohesion is a measure of the quality of social relationships within a group [234, 235], there might be a positive relationship between memory and cohesion. Therefore, another research question that could be studied using the MeMo corpus is the following: How do group cohesion and other group perception parameters influence the remembered moments? Another factor that has only been studied in more contained settings is how the relationship between participants influences memory [14]. This could also be studied based on the MeMo corpus, using the measures of IOS (perceived social distance) recorded after every session [130].

These research questions could be answered through further empirical investigation of the MeMo corpus, with the benefit of ecological validity of its setting, as compared to more traditional in-the-lab cognitive experiments. Answering those research questions could not only bring new insight into cognitive science but also serve as a foundation for further intelligent systems development. For example, to adjust to the context of a particular user or group profile for more accurate memorability prediction.

A

MEMO CORPUS

A

A.1. QUESTIONNAIRES USED IN MEMO CORPUS

Table A.1 shows a full list of measures collected within the surveys in MeMo data collection. See detailed descriptions of each measure in *Section 2.5*.

Table A.1.: A full list of measures collected within the prescreening, pre-session and post-session surveys. Note that the order of the mentioned measures does not correspond to the order in which the measures were presented to participants. '-' in scale column means that the measures were 1-item questions rather than a full validated scale (see *Table A.2* for specific questions).

survey	category	variable	scale	when
prescreening survey	demographics	Age, gender, English fluency, country of residency, COVID-19 affected group	-	before the experiment
	personal charachteristics	Personality	24-item Brief HEXACO Inventory [141]	
		Values	The Short Schwartz's Value Survey [142]	
		Experience with online meetings	-	
	consent	Consent for recording and storing video, audio and survey data	-	
	technical requirements verification	Laptop with working camera and headset with a working microphone	-	
pre-session survey	mood at the start of the session	Mood before the session	Affect Button [144]	right before each conversation session
	screenshot upload	Zoom setup for gaze target extraction	-	
	memory of the previous session	Free recall	-	before 2nd & 3rd sessions & exit interview
	most important moment of previous discussion (start and end times)	Most important moment for grounding the questions in the exit interview	-	before exit interview
post-session survey	memory	Free recall	-	straight after each conversation session
		Timing annotation	-	
		Reason for remembering	-	
	memory for conv. agent	-	-	
	perception of the group and interaction	Task & Group Cohesion	Cohesion in newly formed teams [146]	
		Entitativity	Entitativity Scale [147]	
		Syncness	-	
		Rapport	-	
		Perceived Interdependence	Situational Interdependence [148]	
		Perceived Situation Characteristics	DIAMONDS [149]	
	perception of other participants (one by one)	Relationship (perceived distance)	IOS scale [130]	
		Mutual understanding (perceived values)	The Short Schwartz's Value Survey [142]	
		Quality as a listener	-	
		Personal attitude	-	
		Quality as conversational partner	-	

A

A.2. FORMULATIONS OF ORIGINAL QUESTIONNAIRE

Table A.2 lists the specific formulations of the original questionnaire items that were created specifically for this dataset and were not directly adopted from existing validated scales.

Table A.2.: A list of question formulations for the measures that were specific to this dataset (the ones that were not adopted from an existing validated scale)

Measure	Question formulation	Type of response	Which questionnaire
Free recall reports	"Recall and describe moments of the most recent discussion session in as much detail as you can remember. Any details are great - for example, about the content, other participants, the moderator, you, your feelings, the reaction of others, your words, others' words, timing, or anything that happened throughout the discussion. Recall at least 3 moments. If you remember more, the fields will show up as you go until you leave one of them empty. ... Write the [*first/ second/ third*] moment you remember from the last discussion that you had. The more details the better :) ... Do you remember another moment from the last discussion that you had? In case you remember more moments, another field will show up on the next page. If you don't remember more moments leave the box empty and proceed. "	text	Post-questionnaire (after all sessions) and Pre-questionnaire (before 2nd, 3rd and exit interview)
Timing annotation	"You wrote down several moments you remembered from the previous discussion at the beginning of the survey. Can you now open the video recording and try to find when those moments occurred? It's ok if the timestamp is not too precise. Please, don't rewatch the whole video, only use it to look up the specific time of each memory moment you wrote down. You can move your cursor along the timeline of the video to go to a specific moment. Please close it as soon as you are done with this survey and don't come back to it until you finish the experiment (officially finished the entire experiment on prolific). If you can't find the exact moment, put "0" in the time fields and fill in the option 'Comment' with any details you remember of the timing. Moment[N_{memory}]: At what point in the conversation this moment happened: "[quote from the free recall report [N_{memory}]"."	numbers	Post-questionnaire
Reason for remembering	"Write down why do you think this moment was memorable for you."	text	
Memory for conversational agent question	"We want to build a social robot that really understands you and what is important to you and represent you in the future meetings. It could represent you in the discussions with other people to make sure your perspective is being heard. It could also serve as your personal brain-storming partner with whom you can deliberate important aspects of your life or decisions you need to take. What would be important for such a robot to remember from this meeting? Which specific moments of the conversation would you want it to remember? What details are most important to remember? Write as many details as you can."	text	
Quality as a listener	"To what extent do other participants have the following qualities? - To what extent is [N_{participant}] a good listener? "	7-point likert scale (1=not at all to 7=very much)	
Personal attitude	"- To what extent do you like [N_{participant}]?"		
Quality as conversational partner	"- How would you rate [N_{participant}]'s ability to keep the conversation flowing"		

A

A.3. ADDITIONAL RECRUITMENT CRITERIA

Following are additional recruitment criteria per target group we used to diversify opinions in each group:

- 'Parents with young children': having at least one 2-12 y.o. child at the time of the pandemic.
- 'Students': an active student status.
- 'Older adults': 50 years of age or older (since it was the COVID-19 risk group).
- '(Ex-)business owners': having had an entrepreneur status in the last 4 years before the experiment.

Using these criteria, the resulting sample included 9 parents of young children, 13 students, 10 business owners, 14 older adults (50+), and 7 others, not identifying as parts of the groups above.

A.4. SEPARATED AUDIO SYNCHRONISATION PROCEDURE

Because of a technical issue with the used version of Zoom, these audios were not synchronised to the video. We have synchronised these audios to the original video using Final Cut's synchronise feature [236]. The audio was then manually checked for inconsistencies and corrected to match the original audio. We then plotted the sum of all separated audios against the original audio and calculated the absolute difference between the two audios to find any missed non-synchronised segments. The resulting synchronised separated audio channels are recorded in the original .m4a format, with separate audio for each participant in every session. Since participants used their own computers and headsets, the audio might vary in quality, reflecting an ecologically valid setting of a video-call set-up (see P1.1 in *Section 2.3*).

A.5. LONGITUDINAL QUESTIONNAIRE COMPLETENESS

Table A.3 shows the percentage of participants (including the moderator) that completed pre-and/ or post-questionnaire over all 3 sessions.

Table A.3.: Questionnaire completeness index per group (the percentage of participants that completed pre- and/or post-survey over all the sessions)

Group	**Questionnaire completeness**		
number	**pre-survey**	**post-survey**	**total**
1	0.80	0.80	0.80
2	0.80	0.60	0.60
3	0.50	0.75	0.50
4	0.83	0.67	0.67
5	1.00	1.00	1.00
6	0.60	0.80	0.40
7	0.75	1.00	0.75
8	0.75	0.75	0.50
9	0.50	1.00	0.50
10	0.83	0.83	0.83
11	1.00	1.00	1.00
12	1.00	0.75	0.75
13	0.75	1.00	0.75
14	0.50	0.25	0.25
15	1.00	1.00	1.00

A

A.6. DATASET VERSION STATISTICS

Table A.4 shows the statistics of the data included in each version of the dataset: pseudo-anonymised, processed and curated.

Table A.4.: Statistics of the data included in each publicly available version of the dataset

	Pseudo-anonymized	**Processed**	**Curated**
N_{groups}	15	15	15
$N_{sessions}$	45	45	44
N_p **in conversation**	55	55	54
N_p **in pre-screening survey**	154	55	54
N_p **in post-questionnaire**	54	54	53
N_p **in pre-questionnaire**	54	54	53
recording duration (min)	2050	1942	1892
recording duration (hours)	34	32	31
memorable segments count	853	694	622
memory duration in min (M	3.27	3.27	1.88
+- STD)	+- 6.6	+- 6.6	+- 2.01

A.7. QUESTIONNAIRE DESCRIPTIVE STATISTICS

Table A.5 shows mood and situation perception scores across the curated version of the MeMo dataset.

Table A.5.: Descriptive statistics of mood and DIAMONDS situation perception across MEMO data

	mean	std	min	max
AffectButton: Pleasure	0.35	0.38	-0.61	1.00
AffectButton: Arousal	-0.14	0.77	-1.00	1.00
AffectButton: Dominance	0.22	0.57	-1.00	1.00
DIAMONDS: Duty	3.88	1.62	1.00	7.00
DIAMONDS: Intellect	4.74	1.44	1.00	7.00
DIAMONDS: Adversity	1.42	0.91	1.00	6.00
DIAMONDS: Mating	1.26	0.78	1.00	5.00
DIAMONDS: Positivity	5.75	1.03	3.00	7.00
DIAMONDS: Negativity	2.28	1.41	1.00	7.00
DIAMONDS: Deception	1.61	1.17	1.00	6.00
DIAMONDS: Sociality	5.19	1.51	1.00	7.00

B

THE RELATIONSHIP BETWEEN MEMORY AND AFFECT

B

B.1. VISUALISATION OF METRIC COMPARISON FOR ALL EXPERIMENTS

B.1.1. DTW

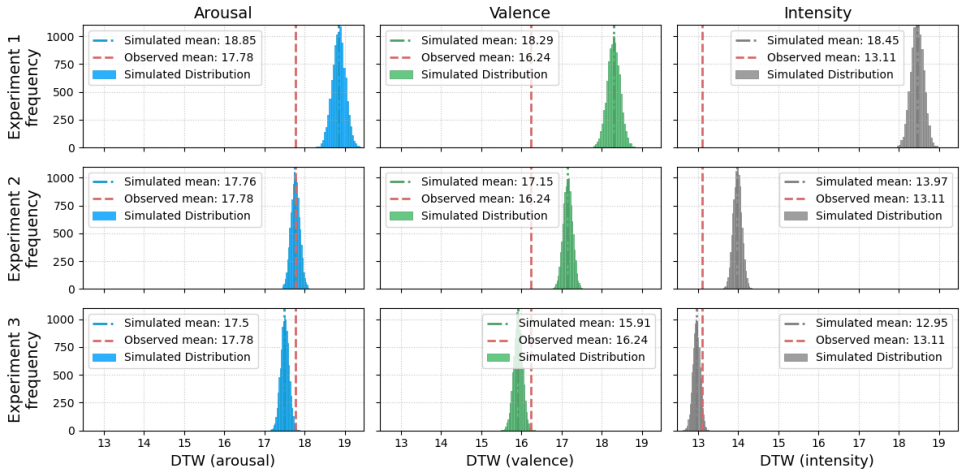


Figure B.1.: DTW distance results for all the experiments (each row of plots shows data for a different experiment) across the three affect dimensions - Arousal (blue), Valence (green), Intensity (gray). The Coloured histogram showed the distribution of averaged DTW distance values under the null hypothesis, the red dashed line shows the averaged DTW distance value for the observed data. *(Metric interpretation: The lower the DTW distance the more alignment is observed in the data.)*

B.1.2. PATE F1

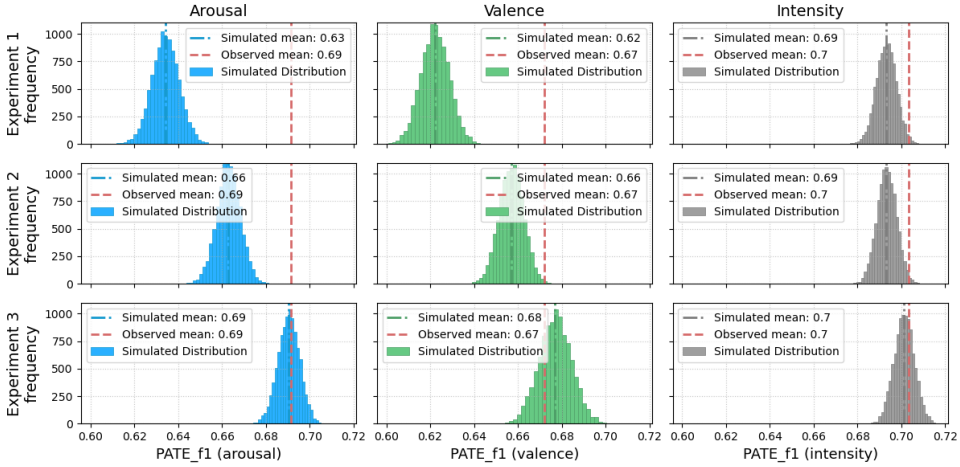


Figure B.2.: **PATE F1** results for all the experiments (each row of plots shows data for a different experiment) across the three affect dimensions - Arousal (blue), Valence (green), and Intensity (grey). The Coloured histogram showed the distribution of averaged PATE F1 values under the null hypothesis, the red dashed line shows the averaged PATE value for the observed data. *(Metric interpretation: The higher PATE F1 the more alignment is observed in the data.)*

B.1.3. PATE

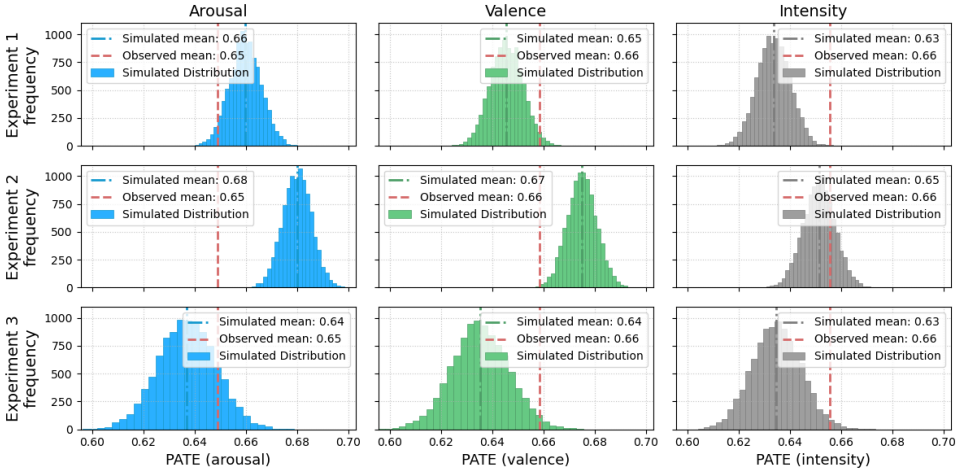


Figure B.3.: **PATE** results for all the experiments (each row of plots shows data for a different experiment) across the three affect dimensions - Arousal (blue), Valence (green), Intensity (grey). The Coloured histogram showed the distribution of averaged PATE values under the null hypothesis, and the red dashed line shows the averaged PATE value for the observed data. *(Metric interpretation: The higher PATE the more alignment is observed in the data.)*

B.1.4. EUCLEDIAN DISTANCE

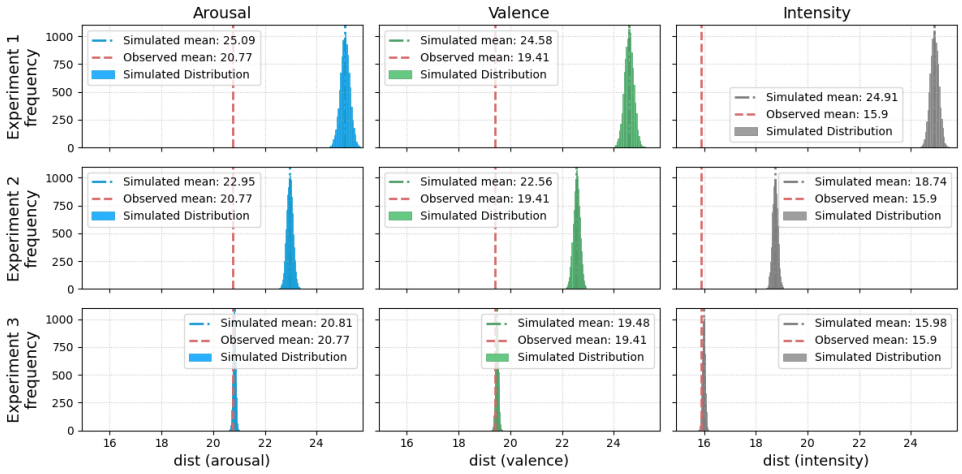


Figure B.4.: **Eucledian distance** results for all the experiments (each row of plots shows data for a different experiment) across the three affect dimensions - Arousal (blue), Valence (green), Intensity (grey). The Colored histogram showed the distribution of averaged Euclidean distance values under the null hypothesis, the red dashed line shows the averaged Euclidean distance value for the observed data. *(Metric interpretation: The lower the Euclidean distance the more alignment is observed in the data.)*

BIBLIOGRAPHY

[1] V. Kaptelinin, K. Danielsson, N. Kaiser, C. Kuenen, and M. Nordin, "Understanding the interpersonal space of online meetings: An exploratory study of "we-ness"," *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*, 2021.

[2] S. Kauffeld and N. Lehmann-Willenbrock, "Meetings matter: Effects of team meetings on team and organizational success," *Small Group Research*, vol. 43, no. 2, pp. 130–158, 2012. [Online]. Available: https://doi.org/10.1177/1046496411429599

[3] J. E. Garon, "Facilitating meetings." *Clinical leadership & management review : the journal of CLMA*, vol. 16 4, pp. 215–23, 2002. [Online]. Available: https://api.semanticscholar.org/CorpusID:36829588

[4] T. Lindblom, M. W. Aiken, and M. Vanjani, "Electronic facilitation of large meetings," *Communications of the IIMA*, 2009. [Online]. Available: https://api.semanticscholar.org/CorpusID:55848946

[5] S. Samrose, D. McDuff, R. Sim, J. Suh, K. Rowan, J. Hernandez, S. Rintel, K. Moynihan, and M. Czerwinski, "Meetingcoach: An intelligent dashboard for supporting effective & inclusive meetings," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: https://doi.org/10.1145/3411764.3445615

[6] A. Shamekhi and T. Bickmore, "A multimodal robot-driven meeting facilitation system for group decision-making sessions," in *ICMI '19*. New York, NY, USA: ACM, 2019.

[7] M. Langner, P. Toreini, and A. Maedche, "Eyemeet: A joint attention support system for remote meetings," *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 2022.

[8] R. Stiefelhagen, "Tracking focus of attention in meetings," *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*, pp. 273–280, 2002.

[9] T. Dacayan, D. Kwak, and X. Zhang, “Computer-vision based attention monitoring for online meetings,” *2022 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI)*, pp. 533–538, 2022.

[10] L. Levine, H. C. Lench, and M. A. Safer, “Functions of remembering and misremembering emotion,” *Applied Cognitive Psychology*, vol. 23, pp. 1059–1075, 2009.

[11] S. Bluck, N. Alea, T. Habermas, and D. C. Rubin, “A tale of three functions: the self–reported uses of autobiographical memory,” *Social Cognition*, vol. 23, pp. 91–117, 2005.

[12] G. L. McKinley, S. Brown-Schmidt, and A. S. Benjamin, “Memory for conversation and the development of common ground,” *Memory & Cognition*, vol. 45, no. 8, pp. 1281–1294, Nov. 2017. [Online]. Available: https://doi.org/10.3758/s13421-017-0730-3

[13] W. L. Benoit, P. J. Benoit, and J. Wilkie, “Participants’ and observers’ memory for conversational behavior,” *Southern Communication Journal*, vol. 61, no. 2, pp. 139–154, Mar. 1996. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/10417949609373007

[14] J. A. Samp and L. R. Humphreys, “"i said what?" partner familiarity, resistance, and the accuracy of conversational recall.” *Communication Monographs*, vol. 74, no. 4, pp. 561–581, 2007. [Online]. Available: https://doi.org/10.1080/03637750701716610

[15] D. Knutsen and L. Le Bigot, “Capturing egocentric biases in reference reuse during collaborative dialogue,” *Psychonomic Bulletin & Review*, vol. 21, no. 6, pp. 1590–1599, Dec 2014. [Online]. Available: https://doi.org/10.3758/s13423-014-0620-7

[16] J. B. Miller and P. A. de Winstanley, “The role of interpersonal competence in memory for conversation,” *Personality and Social Psychology Bulletin*, vol. 28, no. 1, pp. 78–89, 2002. [Online]. Available: https://doi.org/10.1177/0146167202281007

[17] E. Diachek and S. Brown-Schmidt, “Linguistic features of spontaneous speech predict conversational recall,” *Psychonomic Bulletin & Review*, Jan. 2024. [Online]. Available: https://doi.org/10.3758/s13423-023-02440-w

[18] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, “A review of affective computing: From unimodal analysis to multimodal fusion,” *Inf Fus*, vol. 37, 2017.

[19] M. Pantic, R. Cowie, F. D'Errico, D. Heylen, M. Mehu, C. Pelachaud, I. Poggi, M. Schroeder, and A. Vinciarelli, *Social Signal Processing: The Research Agenda*. London: Springer London, 2011, pp. 511–538. [Online]. Available: https://doi.org/10.1007/978-0-85729-997-0_26

[20] A. García Seco de Herrera, M. G. Constantin, C.-H. Demarty, C. Fosco, S. Halder, G. Healy, B. Ionescu, A. Matran-Fernandez, A. F. Smeaton, M. Sultana, and L. Sweeney, "Experiences from the mediaeval predicting media memorability task," in *The NeurIPS MemARI Workshop proceedings*, New Orleans, USA, Dec. 2022. [Online]. Available: https://doras.dcu.ie/27948/

[21] R. Cohendet, C.-H. Demarty, N. Duong, and M. Engilberge, "VideoMem: Constructing, Analyzing, Predicting Short-Term and Long-Term Video Memorability," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, Korea (South): IEEE, Oct. 2019, pp. 2531–2540. [Online]. Available: https://ieeexplore.ieee.org/document/9008778/

[22] N. M. Long, B. A. Kuhl, and M. M. Chun, "Memory and Attention," in *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, J. T. Wixted, Ed. Hoboken, NJ, USA: John Wiley & Sons, Inc., Mar. 2018, pp. 1–37. [Online]. Available: http://doi.wiley.com/10.1002/9781119170174.epcn109

[23] B. Biancardi, L. Maisonnave-Couterou, P. Renault, B. Ravenet, M. Mancini, and G. Varni, "The wonowa dataset: Investigating the transactive memory system in small group interactions," in *Proceedings of the ICMI'20*, ser. ICMI '20. New York, NY, USA: ACM, 2020, p. 528–537. [Online]. Available: https://doi.org/10.1145/3382507.3418843

[24] A. M. Cleary, "Dependent measures in memory research," *Handbook of Research Methods in Human Memory*, pp. 19–35, 2018.

[25] E. R. Smith, C. R. Seger, and D. M. Mackie, "Can emotions be truly group level? evidence regarding four conceptual criteria," *Journal of Personality and Social Psychology*, vol. 93, no. 3, p. 431–446, Sep. 2007.

[26] R. L. Moreland, "Are dyads really groups?" *Small Group Research*, vol. 41, no. 2, pp. 251–267, 2010. [Online]. Available: https://doi.org/10.1177/1046496409358618

[27] L. Hu, J. Cao, G. Xu, L. Cao, Z. Gu, and W. Cao, "Deep modeling of group preferences for group-based recommendation," *Proceedings of the AAAI Conference on Artificial Intelligence*,

vol. 28, no. 1, Jun. 2014. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/9007

[28] S. Brown-Schmidt, C. B. Jaeger, K. Lord, and A. S. Benjamin, "Remembering conversation in group settings," *Memory & Cognition*, 2024.

[29] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain," *Image and Vision Computing*, vol. 27, no. 12, pp. 1743–1759, 2009, visual and multimodal analysis of human spontaneous behaviour:. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885608002485

[30] K. Mahajan and S. Shaikh, "On the need for thoughtful data collection for multi-party dialogue: A survey of available corpora and collection methods," in *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Singapore and Online: Association for Computational Linguistics, Jul. 2021, pp. 338–352. [Online]. Available: https://aclanthology.org/2021.sigdial-1.36

[31] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big?" in *Proceedings of the ACM FAccT '21*, ser. FAccT '21. New York, NY, USA: ACM, 2021, p. 610–623. [Online]. Available: https://doi.org/10.1145/3442188.3445922

[32] A. Paullada, I. D. Raji, E. M. Bender, E. Denton, and A. Hanna, "Data and its (dis)contents: a survey of dataset development and use in machine learning research," *Patterns*, vol. 2, p. 100336, 2021.

[33] L. Stafford, V. R. Waldron, and L. L. Infield, "Actor-Observer Differences in Conversational Memory," *Human Communication Research*, vol. 15, no. 4, pp. 590–611, Jun. 1989. [Online]. Available: https://academic.oup.com/hcr/article/15/4/590-611/4584140

[34] J. M. J. Murre and J. Dros, "Replication and analysis of ebbinghaus' forgetting curve," *PLOS ONE*, vol. 10, no. 7, pp. 1–23, 07 2015. [Online]. Available: https://doi.org/10.1371/journal.pone.0120644

[35] A. Newman, C. Fosco, V. Casser, A. Lee, B. McNamara, and A. Oliva, "Multimodal Memorability: Modeling Effects of Semantics and Decay on Video Memorability," in *Computer Vision – ECCV 2020*, ser. Lecture Notes in Computer Science, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham: Springer International Publishing, 2020, pp. 223–240.

[36] M. Tsfasman, B. Dudzik, K. Fenech, A. Lorincz, C. M. Jonker, and C. Oertel, "Introducing memo: A multimodal dataset for

memory modelling in multiparty conversations," *arXiv preprint arXiv:2409.13715*, 2024.

[37] J. L. McGaugh, "Making lasting memories: Remembering the significant," *Proceedings of the National Academy of Sciences*, vol. 110, no. supplement_2, pp. 10 402–10 407, Jun. 2013.

[38] F. Dolcos, K. S. LaBar, and R. Cabeza, "Interaction between the amygdala and the medial temporal lobe memory system predicts better memory for emotional events," *Neuron*, vol. 42, no. 5, p. 855–863, Jun. 2004.

[39] L. Cahill and J. L. McGaugh, "Modulation of memory storage," *Current Opinion in Neurobiology*, vol. 6, no. 2, p. 237–242, Apr. 1996. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095943889680078X

[40] T. Sharot and E. A. Phelps, "How arousal modulates memory: Disentangling the effects of attention and retention," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 4, no. 3, pp. 294–306, Sep 2004. [Online]. Available: https://doi.org/10.3758/CABN.4.3.294

[41] Z. Kasap and N. Magnenat-Thalmann, "Interacting with Emotion and Memory Enabled Virtual Characters and Social Robots," in *Modeling Machine Emotions for Realizing Intelligence: Foundations and Applications*, ser. Smart Innovation, Systems and Technologies, T. Nishida, L. C. Jain, and C. Faucher, Eds. Berlin, Heidelberg: Springer, 2010, pp. 209–224. [Online]. Available: https://doi.org/10.1007/978-3-642-12604-8_10

[42] L. Martin, J.-H. Rosales, K. Jaime, and F. F. Ramos, "Affective episodic memory system for virtual creatures: The first step of emotion-oriented memory," *Computational Intelligence and Neuroscience*, vol. 2021, 2021.

[43] C. Brom and J. Lukavský, "Towards More Human-Like Episodic Memory for More Human-Like Agents," in *Intelligent Virtual Agents*, ser. Lecture Notes in Computer Science, Z. Ruttkay, M. Kipp, A. Nijholt, and H. H. Vilhjálmsson, Eds. Berlin, Heidelberg: Springer, 2009, pp. 484–485.

[44] S. K. D'mello and J. Kory, "A review and meta-analysis of multimodal affect detection systems," *ACM Comput. Surv.*, vol. 47, no. 3, feb 2015. [Online]. Available: https://doi.org/10.1145/2682899

[45] D. Shohamy and R. A. Adcock, "Dopamine and adaptive memory," *Trends in Cognitive Sciences*, vol. 14, no. 10, p. 464–472, Oct.

2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364661310001865

[46] S. Mariooryad and C. Busso, "Correcting Time-Continuous Emotional Labels by Modeling the Reaction Lag of Evaluators," *IEEE Transactions on Affective Computing*, vol. 6, no. 2, pp. 97–108, Apr. 2015.

[47] T. Zhang, A. El Ali, C. Wang, A. Hanjalic, and P. Cesar, "Rcea: Real-time, continuous emotion annotation for collecting precise mobile video ground truth labels," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu HI USA: ACM, Apr. 2020, pp. 1–15.

[48] E. A. Veltmeijer, C. Gerritsen, and K. V. Hindriks, "Automatic emotion recognition for groups: A review," *IEEE Transactions on Affective Computing*, vol. 14, pp. 89–107, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:234225582

[49] J. Cassell, Y. I. Nakano, T. W. Bickmore, C. L. Sidner, and C. Rich, "Non-verbal cues for discourse structure," in *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, 2001, pp. 114–123.

[50] B. Wrede and E. Shriberg, "Spotting "hot spots" in meetings: human judgments and prosodic cues," in *INTERSPEECH*, 2003.

[51] O. Aran and D. Gatica-Perez, "Fusing audio-visual nonverbal cues to detect dominant people in group conversations," in *2010 20th International Conference on Pattern Recognition*, 2010, pp. 3687–3690.

[52] L. Otten, A. Quayle, S. Akram, T. A. Ditewig, and M. Rugg, "Brain activity before an event predicts later recollection," *Nature Neuroscience*, vol. 9, pp. 489–491, 2006.

[53] P. Isola, J. Xiao, D. Parikh, A. Torralba, and A. Oliva, "What makes a photograph memorable?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, pp. 1469–1482, 2014.

[54] B. Dudzik, H. Hung, M. Neerincx, and J. Broekens, "Collecting mementos: A multimodal dataset for context-sensitive modeling of affect and memory processing in responses to videos," *IEEE Transactions on Affective Computing*, vol. 14, no. 2, pp. 1249–1266, 2023, accepted author manuscript.

[55] J. O'Dwyer, N. Murray, and R. Flynn, "Eye-based continuous affect prediction," *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 137–143, 2019.

[56] C. Oertel and G. Salvi, "A gaze-based method for relating group involvement to individual engagement in multimodal multiparty dialogue," in *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, ser. ICMI '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 99–106. [Online]. Available: https://doi-org.tudelft.idm.oclc.org/10.1145/2522848.2522865

[57] F. Capozzi and J. Ristic, "Attentional gaze dynamics in group interactions," *Visual Cognition*, vol. 30, pp. 135–150, 2021.

[58] G. Underwood, *Attention and memory*. Elsevier, 2013.

[59] J. Q. Sargent, J. M. Zacks, D. Z. Hambrick, R. T. Zacks, C. A. Kurby, H. R. Bailey, M. L. Eisenberg, and T. M. Beck, "Event segmentation ability uniquely predicts event memory," *Cognition*, vol. 129, no. 2, pp. 241–255, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010027713001352

[60] M. Tsfasman, K. Fenech, M. Tarvirdians, A. Lorincz, C. Jonker, and C. Oertel, "Towards creating a conversational memory for long-term meeting support: predicting memorable moments in multi-party conversations through eye-gaze," in *Proceedings of the 2022 ICMI*. Bengaluru India: ACM, Nov. 2022, p. 94–104. [Online]. Available: https://dl.acm.org/doi/10.1145/3536221.3556613

[61] R. Stickgold and M. P. Walker, "Sleep-dependent memory triage: evolving generalization through selective processing," *Nature Neuroscience*, vol. 16, no. 2, pp. 139–145, Feb. 2013. [Online]. Available: https://www.nature.com/articles/nn.3303

[62] C. W. Morris, "On the importance of conversation," *Dialogue*, vol. 32, 1993.

[63] L. Bietti, "Sharing memories, family conversation and interaction," *Discourse & Society*, vol. 21, pp. 499–523, 2010.

[64] S. Rossi, F. Ferland, and A. Tapus, "User profiling and behavioral adaptation for hri: A survey," *Pattern Recognit. Lett.*, vol. 99, pp. 3–12, 2017.

[65] U. Rutishauser, L. Reddy, F. Mormann, and J. Sarnthein, "The architecture of human memory: Insights from human single-neuron recordings," *The Journal of Neuroscience*, vol. 41, pp. 883–890, 2020.

[66] J. Holt-Lunstad, "The potential public health relevance of social isolation and loneliness: Prevalence, epidemiology, and risk factors," *Public Policy & Aging Report*, vol. 27, p. 127–130, 2017.

[67] W. H. Organisation, "Social isolation and loneliness among older people: advocacy brief," 2021. [Online]. Available: https://www.who.int/publications-detail-redirect/9789240030749

[68] D. Jeste, E. E. Lee, and S. Cacioppo, "Battling the modern behavioral epidemic of loneliness: Suggestions for research and interventions." *JAMA psychiatry*, 2020.

[69] I. Lefter, D. D. Luxton, A. Baird, T. Chaspari, Z. Hammal, M. Mahmoud, and A. A. Salah, "Affective computing for mental wellbeing: Challenges, opportunities, and promising synergies," in *2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, 2023, pp. 1–2.

[70] H. T. Reis, K. M. Sheldon, S. L. Gable, J. A. Roscoe, and R. M. Ryan, "Daily well-being: the role of autonomy, competence, and relatedness," *Personality and Social Psychology Bulletin*, vol. 26, pp. 419–435, 2000.

[71] J. Raczaszek-Leonardi, A. Debska, and A. Sochanowicz, "Pooling the ground: understanding and coordination in collective sense making," *Front Psychol*, vol. 5, 2014.

[72] D. Richards and K. Bransky, "ForgetMeNot: What and how users expect intelligent virtual agents to recall and forget personal conversational content," *International Journal of Human-Computer Studies*, vol. 72, no. 5, pp. 460–476, May 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1071581914000147

[73] M. Elvir, A. J. Gonzalez, C. Walls, and B. Wilder, "Remembering a Conversation – A Conversational Memory Architecture for Embodied Conversational Agents," *Journal of Intelligent Systems*, vol. 26, no. 1, pp. 1–21, Jan. 2017. [Online]. Available: https://www.degruyter.com/document/doi/10.1515/jisys-2015-0094/html

[74] J. Campos, J. Kennedy, and J. F. Lehman, "Challenges in Exploiting Conversational Memory in Human-Agent Interaction," in *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, ser. AAMAS '18. Stockholm, Sweden: International Foundation for Autonomous Agents and Multiagent Systems, Jul. 2018, pp. 1649–1657. [Online]. Available: https://dl-acm-org.tudelft.idm.oclc.org/doi/10.5555/3237383.3237945

[75] F. Nihei and Y. I. Nakano, "Exploring Methods for Predicting Important Utterances Contributing to Meeting Summarization," *Multimodal Technologies and Interaction*, vol. 3, no. 3, p. 50, Sep. 2019, number: 3 Publisher: Multidisciplinary Digital Publishing

Institute. [Online]. Available: https://www.mdpi.com/2414-4088/3/3/50

[76] S. Kashmira, J. L. Dantanarayana, J. Brodsky, A. Mahendra, Y. Kang, K. Flautner, L. Tang, and J. Mars, "A graph-based approach for conversational ai-driven personal memory capture and retrieval in a real-world application," *ArXiv*, vol. abs/2412.05447, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:274597659

[77] S. A. Bahrainian and F. A. Crestani, "Predicting the topics to review in preparation of your next meeting," in *Italian Information Retrieval Workshop*, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:2662007

[78] M. Kopelman, "Disorders of memory." *Brain : a journal of neurology*, vol. 125 Pt 10, pp. 2152–90, 2002.

[79] M. Tsfasman, R. Ghorbani, C. M. Jonker, and B. Dudzik, "The emotion-memory link: Do memorability annotations matter for intelligent systems?" 2025. [Online]. Available: https://arxiv.org/abs/2507.14084

[80] P. Rotshtein, Ed., *Encyclopedia of Behavioral Neuroscience, 2nd edition*. Netherlands: Elsevier, 2021.

[81] M. A. Conway, *Autobiographical Memory*. Elsevier, 1996, p. 165–194. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/B9780121025700500082

[82] S. Brown-Schmidt and M. C. Duff, "Memory and common ground processes in language use," *Topics in Cognitive Science*, vol. 8, no. 4, p. 722–736, Oct. 2016.

[83] P. J. Benoit and W. L. Benoit, "Anticipated future interaction and conversational memory using participants and observers," *Communication Quarterly*, vol. 42, no. 3, pp. 274–286, Jun. 1994. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/01463379409369934

[84] L. Stafford, C. S. Burggraf, and W. F. Sharkey, "Conversational Memory The Effects of Time, Recall, Mode, and Memory Expectancies on Remembrances of Natural Conversations," *Human Communication Research*, vol. 14, no. 2, pp. 203–229, Dec. 1987. [Online]. Available: https://academic.oup.com/hcr/article/14/2/203-229/4587713

[85] E. Cambria and B. White, "Jumping nlp curves: A review of natural language processing research," *IEEE Comput Intell Mag*, vol. 9, no. 2, 2014.

[86] B. A. Nosek, T. E. Hardwicke, H. Moshontz, A. Allard, K. S. Corker, A. Dreber, F. Fidler, J. Hilgard, M. Kline Struhl, M. B. Nuijten *et al.*, "Replicability, robustness, and reproducibility in psychological science," *Ann Rev Psych*, vol. 73, 2022.

[87] M. Hutson, "Artificial intelligence faces reproducibility crisis," 2018.

[88] J. Holt-Lunstad, "The major health implications of social connection," *Curr Dir Psychol Sci*, vol. 30, 2021.

[89] A. Garcia, "Dispute resolution without disputing: How the interactional organization of mediation hearings minimizes argument," *Am Soc Rev*, vol. 56, 1991.

[90] C. A. Picard and M. Jull, "Learning through deepening conversations: A key strategy of insight mediation," *Confl Res Quart*, vol. 29, 2011.

[91] K. N. Dillard, "Envisioning the role of facilitation in public deliberation," *J Appl Commun Res*, vol. 41, 2013.

[92] C. Bruce, A. D. Newell, J. H. Brewer, D. O. Timme, E. Cherry, J. Moore, J. Carrettin, E. Landeck, R. Axline, A. Millette, R. Taylor, A. Downey, F. Uddin, D. Gotur, F. Masud, and D. Zhukovsky, "Developing and testing a comprehensive tool to assess family meetings: Empirical distinctions between high- and low-quality meetings," *Journal of Critical Care*, vol. 42, p. 223–230, 2017.

[93] S. C. Hayne, "The facilitators perspective on meetings and implications for group support systems design," *SIGMIS Database*, vol. 30, no. 3–4, p. 72–91, sep 1999. [Online]. Available: https://doi.org/10.1145/344241.344246

[94] L. Phillips and M. C. Phillips, "Faciliated work groups: Theory and practice," *Journal of the Operational Research Society*, vol. 44, pp. 533–549, 1993.

[95] J. V. Li, M. Kreminski, S. M. Fernandes, A. Osborne, J. McVeigh-Schultz, and K. Isbister, "Conversation balance: A shared vr visualization to support turn-taking in meetings," in *Extended Abstracts of the 2022 CHI EA '22*, ser. CHI EA '22. New York, NY, USA: ACM, 2022. [Online]. Available: https://doi.org/10.1145/3491101.3519879

[96] G. Schiavo, A. Cappelletti, E. Mencarini, O. Stock, and M. Zancanaro, "Overt or subtle? supporting group conversations with automatically targeted directives," *Proceedings of the 19th ACM IUI*, 2014.

[97] O. A. Kulyk, J. Wang, and J. M. B. Terken, "Real-time feedback on nonverbal behaviour to enhance social dynamics in small group meetings," in *MLMI*, 2005. [Online]. Available: https://api.semanticscholar.org/CorpusID:16641435

[98] T. J. Kim, A. Chang, L. Holland, and A. S. Pentland, "Meeting mediator: enhancing group collaboration using sociometric feedback," *Proceedings of the 2008 ACM CSCW*, 2008. [Online]. Available: https://api.semanticscholar.org/CorpusID:52798570

[99] K. Nowak, L. Tankelevitch, J. Tang, and S. Rintel, "Hear we are: Spatial audio benefits perceptions of turn-taking and social presence in video meetings," *Proceedings of the 2nd Annual Meeting of the Symposium on Human-Computer Interaction for Work*, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:259267028

[100] T. Okada, S. Okamoto, and Y. Yamada, "Affective dynamics: Causality modeling of temporally evolving perceptual and affective responses," *IEEE Transactions on Affective Computing*, vol. 13, pp. 628–639, 2019.

[101] D. A. Norman, "The way i see it memory is more important than actuality," *Interactions*, vol. 16, no. 2, p. 24–26, mar 2009. [Online]. Available: https://doi-org.tudelft.idm.oclc.org/10.1145/1487632.1487638

[102] S. Nørby, "Why forget? on the adaptive value of memory loss," *Perspectives on Psychological Science*, vol. 10, no. 5, pp. 551–578, 2015, pMID: 26385996. [Online]. Available: https://doi.org/10.1177/1745691615596787

[103] E. Niforatos, M. Laporte, A. Bexheti, and M. Langheinrich, "Augmenting memory recall in work meetings: Establishing a quantifiable baseline," in *Proceedings of the 9th Augmented Human International Conference*, ser. AH '18. New York, NY, USA: Association for Computing Machinery, 2018. [Online]. Available: https://doi.org/10.1145/3174910.3174920

[104] S. A. Bahrainian and F. Crestani, "Augmentation of human memory: Anticipating topics that continue in the next meeting," in *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, ser. CHIIR '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 150–159. [Online]. Available: https://doi.org/10.1145/3176349.3176399

[105] W. A. Bainbridge, D. D. Dilks, and A. Oliva, "Memorability: A stimulus-driven perceptual neural signature distinctive from memory," *NeuroImage*, vol. 149, pp. 141–152, 2017.

[106] N.-A. H. Tan, H. Sha, E. Celen, P. Tran, K. Wang, G. Cheung, P. Hinch, and J. Huang, “Rewind: Automatically reconstructing everyday memories with first-person perspectives,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 4, pp. 191:1–191:20, 2018.

[107] F. M. Li, D. L. Chen, M. Fan, and K. N. Truong, “Fmt: A wearable camera-based object tracking memory aid for older adults,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 3, no. 3, pp. 95:1–95:25, 2019.

[108] R. N. Brewer, M. R. Morris, and S. E. Lindley, “How to remember what to remember: Exploring possibilities for digital reminder systems,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 3, pp. 38:1–38:20, 2017.

[109] S. W. T. Chan, T. Buddhika, H. Zhang, and S. Nanayakkara, “Prospecfit: In situ evaluation of digital prospective memory training for older adults,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 3, no. 3, pp. 77:1–77:20, 2019.

[110] M. Laporte, M. Gjoreski, and M. Langheinrich, “Laureate: A dataset for supporting research in affective computing and human memory augmentation,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 7, no. 3, Sep. 2023. [Online]. Available: https://doi.org/10.1145/3610892

[111] M. L. Knapp and J. A. Hall, *Nonverbal communication in human interaction*. Boston, MA: Wadsworth, Cengage Learning, 2010, oCLC: 244767251.

[112] W. L. Benoit and P. J. Benoit, “Memory for conversational behavior,” *Southern Communication Journal*, vol. 56, no. 1, pp. 24–33, Dec. 1990. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/10417949009372813

[113] R. Heale and A. Twycross, “Validity and reliability in quantitative studies,” *Evidence-Based Nursing*, vol. 18, pp. 66–67, 2015.

[114] U. Rumpf, I. Menze, N. G. Müller, and M. Schmicker, “Investigating the potential role of ecological validity on change-detection memory tasks and distractor processing in younger and older adults,” *Frontiers in Psychology*, vol. 10, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:162183455

[115] J. E. Dunsmoor, V. P. Murty, D. Clewett, E. A. Phelps, and L. Davachi, “Tag and capture: how salient experiences target and rescue nearby events in memory,” *Trends in Cognitive Sciences*,

vol. 26, no. 9, pp. 782–795, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364661322001401

[116] K. M. Schnitzspahn, L. Kvavilashvili, and M. Altgassen, "Redefining the pattern of age-prospective memory-paradox: new insights on age effects in lab-based, naturalistic, and self-assigned tasks," *Psychological Research*, vol. 84, pp. 1370–1386, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:58544411

[117] H. Jiang, X. Zhang, and J. D. Choi, "Automatic text-based personality recognition on monologues and multiparty dialogues using attentive networks and contextual embeddings," in *Proceedings of the CAI'20*, vol. 34, no. 10. AAAI, Apr. 2020, p. 13821–13822. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/7182

[118] M. Daneman and I. Green, "Individual differences in comprehending and producing words in context," *Journal of Memory and Language*, vol. 25, pp. 1–18, 1986.

[119] J. Erba, P. S. Bobkowski, B. Ternes, Y. Liu, and T. Logan, "Who are the "masses" in mass communication research? exploring participants' demographic characteristics between 2000 and 2014," *Howard Journal of Communications*, vol. 33, pp. 233–249, 2021.

[120] C. Infante-Rivard and A. Cusson, "Reflection on modern methods: selection bias-a review of recent developments." *International journal of epidemiology*, vol. 47 5, pp. 1714–1722, 2018.

[121] N. Qureshi, M. Edelen, L. Hilton, A. Rodriguez, R. D. Hays, and P. Herman, "Comparing data collected on amazon's mechanical turk to national surveys." *American journal of health behavior*, vol. 46 5, pp. 497–502, 2022.

[122] R. Cohendet, A.-L. Gilet, M. Perreira Da Silva, and P. Le Callet, "Using individual data to characterize emotional user experience and its memorability: Focus on gender factor," in *2016 Eighth QoMEX*. IEEE, 06 2016, pp. 1–6.

[123] A. Mahr, M. Cichon, S. Mateo, C. Grajeda, and I. Baggili, "Zooming into the pandemic! a forensic analysis of the zoom application," *Forensic Science International: Digital Investigation*, vol. 36, pp. 301 107 – 301 107, 2021.

[124] B. Dudzik, J. Broekens, M. Neerincx, J. Olenick, C.-H. Chang, S. W. J. Kozlowski, and H. Hung, "Discovering digital representations for remembered episodes from lifelog data," in *Proceedings of the Workshop on Modeling Cognitive Processes from Multimodal Data*, ser. MCPMD '18. New York, NY, USA: Association

for Computing Machinery, 2018. [Online]. Available: https://doi.org/10.1145/3279810.3279850

[125] C. Patino and J. Ferreira, "Internal and external validity: can you apply research study results to your patients?" *Jornal Brasileiro de Pneumologia*, vol. 44, no. 3, p. 183, Jun. 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6188693/

[126] T. Jd and R. Ml, "Differential effects of induced mood on the recall of positive, negative and neutral words," *British Journal of Clinical Psychology*, vol. 22, pp. 163–171, 1983.

[127] G. Matt, C. Vázquez, and W. K. Campbell, "Mood-congruent recall of affectively toned stimuli: A meta-analytic review," *Clinical Psychology Review*, vol. 12, pp. 227–255, 1992.

[128] P. R. Mayo, "A further study of the personality-congruent recall effect," *Personality and Individual Differences*, vol. 10, pp. 247–252, 1989.

[129] J. J. V. nor, A. Sklenar, A. Frankenstein, P. U. Levy, M. P. McCurdy, and E. Leshikar, "Value-directed memory effects on item and context memory," *Memory & Cognition*, vol. 49, pp. 1082–1100, 2021.

[130] K. M. Woosnam, "The inclusion of other in the self (ios) scale," *Annals of Tourism Research*, vol. 37, no. 3, pp. 857–860, 2010.

[131] C. R. Evans and K. Dion, "Group cohesion and performance," *Small Group Research*, vol. 22, pp. 175–186, 1991.

[132] S.-Y. Kim and E. Yang, "Does group cohesion foster self-directed learning for medical students? a longitudinal study," *BMC Medical Education*, vol. 20, 2020.

[133] R. Shaw and C. Kitzinger, "Memory in interaction: an analysis of repeat calls to a home birth helpline," *Research on Language & Social Interaction*, vol. 40, pp. 117–144, 2007.

[134] E. Tulving and Z. Pearlstone, "Availability versus accessibility of information in memory for words," *Journal of Verbal Learning and Verbal Behavior*, vol. 5, pp. 381–391, 1966.

[135] W. S. Horton and R. J. Gerrig, "Conversational common ground and memory processes in language production," *Discourse Processes*, vol. 40, no. 1, pp. 1–35, 2005.

[136] H. H. Clark and D. Wilkes-Gibbs, "Referring as a collaborative process," *Cognition*, vol. 22, no. 1, pp. 1–39, 1986.

[137] R. Cohendet, K. Yadati, N. Q. K. Duong, and C.-H. Demarty, "Annotating, Understanding, and Predicting Long-term Video Memorability," in *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*, ser. ICMR '18. New York, NY, USA: Association for Computing Machinery, Jun. 2018, pp. 178–186. [Online]. Available: http://doi.org/10.1145/3206025.3206056

[138] O. Raccah, P. Chen, T. M. Gureckis, D. Poeppel, and V. A. Vo, "The "naturalistic free recall" dataset: four stories, hundreds of participants, and high-fidelity transcriptions," *Scientific Data*, vol. 11, no. 1, p. 1317, Dec. 2024. [Online]. Available: https://www.nature.com/articles/s41597-024-04082-6

[139] Prolific, "Prolific: quickly find research participants you can trust." 2022. [Online]. Available: http://www.prolific.co/

[140] S. Wheelan, "Group size, group development, and group productivity," *Small Group Research*, vol. 40, pp. 247–262, 2009.

[141] R. E. de Vries, "The 24-item brief hexaco inventory (bhi)," *Journal of Research in Personality*, vol. 47, no. 6, pp. 871–880, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0092656613001220

[142] M. Lindeman and M. Verkasalo, "Measuring values with the short schwartz's value survey," *Journal of Personality Assessment*, vol. 85, no. 2, pp. 170–178, 2005, pMID: 16171417. [Online]. Available: https://doi.org/10.1207/s15327752jpa8502_09

[143] Qualtrics, "1 xm platform | powerful experience analytics," Oct 2021. [Online]. Available: https://www.qualtrics.com/

[144] J. Broekens and W.-P. Brinkman, "Affectbutton: A method for reliable and valid affective self-report," *International Journal of Human-Computer Studies*, vol. 71, no. 6, pp. 641–667, 2013.

[145] A. Aron, E. N. Aron, and D. Smollan, "Inclusion of other in the self scale and the structure of interpersonal closeness." *J Pers Soc Psychol*, vol. 63, no. 4, 1992.

[146] M. T. Braun, S. W. Kozlowski, T. A. Brown, and R. P. DeShon, "Exploring the dynamic team cohesion–performance and coordination–performance relationships of newly formed teams," *Small Group Res*, vol. 51, no. 5, 2020.

[147] N. Koudenburg, T. Postmes, and E. H. Gordijn, "Conversational flow and entitativity: The role of status," *Br J Clin Psychol*, vol. 53, no. 2, 2014.

[148] F. H. Gerpott, D. Balliet, S. Columbus, C. Molho, and R. E. de Vries, "How do people think about interdependence? a multidimensional model of subjective outcome interdependence." *J Pers Soc Psychol*, vol. 115, no. 4, 2018.

[149] J. F. Rauthmann and R. A. Sherman, "Ultra-brief measures for the situational eight diamonds domains." *Eur J Psychol Assess*, vol. 32, no. 2, 2016.

[150] C. A. Kurby and J. M. Zacks, "Segmentation in the perception and memory of events," *Trends Cogn Sci*, vol. 12, no. 2, 2008.

[151] Zoom, "Zoom video conferencing platform." [Online]. Available: https://zoom.us

[152] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely, "The kaldi speech recognition toolkit," in *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society, Dec. 2011, iEEE Catalog No.: CFP11SRW-USB.

[153] GazeSense, "3d eye tracking software for depth-sensing cameras," Jul 2022. [Online]. Available: https://eyeware.tech/gazesense/

[154] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "OpenFace: A general-purpose face recognition library with mobile applications," CMU-CS-16-118, CMU School of Computer Science, Tech. Rep., 2016.

[155] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, and M. Grundmann, "Mediapipe: A framework for building perception pipelines," *ArXiv*, vol. abs/1906.08172, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:195069430

[156] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "Openface: an open source facial behavior analysis toolkit," in *2016 IEEE WACV*. IEEE, 2016, pp. 1–10.

[157] R. L. Cohen, "On the generality of some memory laws," *Scandinavian Journal of Psychology*, vol. 22, no. 1, pp. 267–281, 1981. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9450.1981.tb00402.x

[158] S. Gächter, C. Starmer, and F. Tufano, "Measuring the closeness of relationships: A comprehensive evaluation of the 'inclusion of the other in the self' scale," *PLoS ONE*, vol. 10, 2015.

[159] J. Mu, W. Wang, W. Liu, T. Yan, and G. Wang, "Multimodal large language model with lora fine-tuning for multimodal sentiment analysis," *ACM Transactions on Intelligent Systems and Technology*, 2024.

[160] D. Jiachu, L. Luo, M. Xie, X. Xie, J. Guo, H. Ye, K. Cai, L. Zhou, G. Song, F. Jiang, D. Huang, M. Zhang, and C. Zheng, "A meta-learning approach for classifying multimodal retinal images of retinal vein occlusion with limited data," *Translational Vision Science Technology*, vol. 13, 2024.

[161] B. Dong, R. Wang, J. Yang, and L. Xue, "Multi-scale feature self-enhancement network for few-shot learning," *Multimedia Tools and Applications*, vol. 80, pp. 33 865 – 33 883, 2021.

[162] L. Clark, N. Pantidi, O. Cooney, P. R. Doyle, D. Garaialde, J. Edwards, B. Spillane, C. Murad, C. Munteanu, V. Wade, and B. R. Cowan, "What makes a good conversation?: Challenges in designing truly conversational agents," *Proceedings of the 2019 CHI*, 2019.

[163] S. C. Akhter-Khan and R. Au, "Why loneliness interventions are unsuccessful: A call for precision health," *Adv Geriatr Med Res*, vol. 2, 2020.

[164] E. Croes and M. L. Antheunis, "Can we be friends with mitsuku? a longitudinal study on the process of relationship formation between humans and a social chatbot," *Journal of Social and Personal Relationships*, vol. 38, pp. 279–300, 2020.

[165] C. BROM, J. LUKAVSKÝ, and R. KADLEC, "EPISODIC MEMORY FOR HUMAN-LIKE AGENTS AND HUMAN-LIKE AGENTS FOR EPISODIC MEMORY," *International Journal of Machine Consciousness*, vol. 02, no. 02, pp. 227–244, Dec. 2010. [Online]. Available: http://www.worldscientific.com/doi/abs/10.1142/S1793843010000461

[166] Z. Kasap and N. Magnenat-Thalmann, "Building long-term relationships with virtual and robotic characters: the role of remembering," *The Visual Computer*, vol. 28, no. 1, pp. 87–97, Jan. 2012. [Online]. Available: https://doi.org/10.1007/s00371-011-0630-7

[167] M. Tsfasman, A. Saravanan, D. Viner, D. Goslinga, S. De Wolf, C. Raman, C. M. Jonker, and C. Oertel, "Towards a real-time measure of the perception of anthropomorphism in human-robot interaction," in *Proceedings of the 2nd ACM Multimedia Workshop on Multimodal Conversational AI*, 2021, pp. 13–18.

[168] Y. Tian, S. Liu, and J. Wang, "A corpus study on the difference of turn-taking in online audio, online video, and face-to-face conversation." *Language and speech*, p. 238309231176768, 2023.

[169] C. S. L. Ng and W. Cheung, "Comparing face to face, tutor led discussion and online discussion in the classroom," *Australasian Journal of Educational Technology*, vol. 23, pp. 455–469, 2007.

[170] D. G. Ray, S. Gomillion, A. I. Pintea, and I. Hamlin, "On being forgotten: Memory and forgetting serve as signals of interpersonal importance," *J Pers Soc Psychol*, vol. 116, 2019.

[171] R. Briker, S. Hohmann, F. Walter, C. K. Lam, and Y. Zhang, "Formal supervisors' role in stimulating team members' informal leader emergence: Supervisor and member status as critical moderators," *Journal of Organizational Behavior*, 2021.

[172] V. Maljkovic and P. Martini, "Short-term memory for scenes with affective content," *Journal of Vision*, vol. 5, no. 3, p. 6, Mar. 2005. [Online]. Available: http://jov.arvojournals.org/article.aspx?doi=10.1167/5.3.6

[173] S. Erk, M. Kiefer, J. o. Grothe, A. P. Wunderlich, M. Spitzer, and H. Walter, "Emotional context modulates subsequent memory effect," *NeuroImage*, vol. 18, no. 2, p. 439–447, Feb. 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811902000150

[174] C. F. A. Gomes, C. J. Brainerd, and L. M. Stein, "Effects of emotional valence and arousal on recollective and nonrecollective recall." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 39, no. 3, p. 663–677, May 2013. [Online]. Available: https://doi.apa.org/doi/10.1037/a0028578

[175] L. Olteanu, S. Golani, B. Eitam, and A. Kron, "The effect of relevance appraisal on the emotional response." *Emotion*, vol. 19, no. 4, p. 715–725, Jun. 2019. [Online]. Available: https://doi.apa.org/doi/10.1037/emo0000473

[176] Z. Kasap and N. Magnenat-Thalmann, "Towards episodic memory-based long-term affective interaction with a human-like robot," in *19th International Symposium in Robot and Human Interactive Communication*. IEEE, Sep. 2010, pp. 452–457, iSSN: 1944-9445. [Online]. Available: http://ieeexplore.ieee.org/document/5598644/

[177] M. Ahmad, Y. Gao, F. Alnajjar, S. Shahid, and O. Mubin, "Emotion and memory model for social robots: a reinforcement learning based behaviour selection," *Behaviour & Information Technology*, vol. 41, pp. 3210–3236, 2021.

[178] T. Moerland, J. Broekens, and C. Jonker, "Emotion in reinforcement learning agents and robots: a survey," *Machine Learning*, vol. 107, pp. 443–480, 2017.

[179] K. R. Scherer, “Emotion as a multicomponent process: A model and some cross-cultural data,” *Review of Personality & Social Psychology*, vol. 5, p. 37–63, 1984.

[180] M. Lewis, “Bridging emotion theory and neurobiology through dynamic systems modeling,” *The Behavioral and brain sciences*, vol. 28, pp. 169–94; discussion 194, 05 2005.

[181] N. R. Prabhu, M. Tsfasman, C. Oertel, T. Gerkmann, and N. Lehmann-Willenbrock, “Dynamics of collective group affect: Group-level annotations and the multimodal modeling of convergence and divergence,” 2024. [Online]. Available: https://arxiv.org/abs/2409.08578

[182] J. M. Talarico, K. S. LaBar, and D. C. Rubin, “Emotional intensity predicts autobiographical memory experience,” *Memory & Cognition*, vol. 32, no. 7, pp. 1118–1132, Oct. 2004.

[183] B. Dudzik and J. Broekens, “A Valid Self-Report is Never Late, Nor is it Early: On Considering the "Right" Temporal Distance for Assessing Emotional Experience,” *arXiv*, Jan. 2023.

[184] A. Triantafyllou and G. A. Tsihrintzis, “Group affect recognition: Completed databases and smart uses,” in *Proceedings of the 2019 3rd International Conference on E-Education, E-Business and E-Technology*, ser. ICEBT ’19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 38–42. [Online]. Available: https://doi.org/10.1145/3355166.3355965

[185] H. Järvenoja, S. Järvelä, and J. Malmberg, “Supporting groups’ emotion and motivation regulation during collaborative learning,” *Learning and Instruction*, vol. 70, p. 101090, 2017.

[186] A. Chohra, K. Madani, and C. N. van der Wal, “Group affect in complex decision-making: Theory and formalisms from psychology and computer science,” in *Computational Collective Intelligence*, N. T. Nguyen, E. Pimenidis, Z. Khan, and B. Trawiński, Eds. Cham: Springer International Publishing, 2018, pp. 222–233.

[187] S. G. Barsade and D. E. Gibson, “Group affect: Its influence on individual and group outcomes,” *Current Directions in Psychological Science*, vol. 21, no. 2, p. 119–123, Apr. 2012. [Online]. Available: https://journals.sagepub.com/doi/10.1177/0963721412438352

[188] A. Klep, B. M. Wisse, and H. van der Flier, “Interactive affective sharing versus non-interactive affective sharing in work groups: Comparative effects of group affect on work group performance and dynamics,” *European Journal of Social*

Psychology, vol. 41, pp. 312–323, 2011. [Online]. Available: https://api.semanticscholar.org/CorpusID:37625804

[189] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, p. 1161–1178, 1980.

[190] T. Wolf, J. Pociunaite, S. Hoehne, and D. Zimprich, "The valence and the functions of autobiographical memories: Does intensity matter?" *Consciousness and Cognition*, vol. 91, p. 103119, 2021.

[191] R. Reisenzein, "Pleasure-arousal theory and the intensity of emotions." *Journal of Personality and Social Psychology*, vol. 67, no. 3, p. 525–539, Sep. 1994. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.67.3.525

[192] N. Goto and A. Schaefer, *Emotional Intensity*. Cham: Springer International Publishing, 2017, pp. 1–9. [Online]. Available: https://doi.org/10.1007/978-3-319-28099-8_509-1

[193] R. Ghorbani, M. J. T. Reinders, and D. M. J. Tax, "Pate: Proximity-aware time series anomaly evaluation," 2024.

[194] O. Jeunehomme and A. D'Argembeau, "Event segmentation and the temporal compression of experience in episodic memory," *Psychological Research*, vol. 84, no. 2, p. 481–490, Mar. 2020. [Online]. Available: https://doi.org/10.1007/s00426-018-1047-y

[195] Muller, *Dynamic Time Warping*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 69–84. [Online]. Available: https://doi.org/10.1007/978-3-540-74048-3_4

[196] L. Cameron and N. Overall, "Suppression and expression as distinct emotion-regulation processes in daily interactions: Longitudinal and meta-analyses," *Emotion*, vol. 18, p. 465–480, 2018.

[197] J. Yoon, S. R. Thye, and E. Lawler, "Exchange and cohesion in dyads and triads: A test of simmel's hypothesis." *Social science research*, vol. 42 6, pp. 1457–66, 2013.

[198] A. J. Moye and M. K. van Vugt, "A computational model of focused attention meditation and its transfer to a sustained attention task," *IEEE TAffC*, vol. 12, no. 2, 2021.

[199] K. Oberauer, "Working memory and attention - a conceptual analysis and review," *Journal of cognition*, vol. 2, no. 1, pp. 36–36, Aug 2019, 31517246[pmid]. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/31517246

[200] M. T. deBettencourt, K. A. Norman, and N. B. Turk-Browne, "Forgetting from lapses of sustained attention," *Psychonomic Bulletin & Review*, vol. 25, no. 2, p. 605–611, Apr. 2018. [Online]. Available: https://doi.org/10.3758/s13423-017-1309-5

[201] L. Baker-Ward, T. M. Hess, and D. A. Flannagan, "The effects of involvement on children's memory for events," *Cognitive Development*, vol. 5, no. 1, p. 55–69, Jan. 1990. [Online]. Available: https://www.sciencedirect.com/science/article/pii/088520149090012I

[202] S. Ho, T. Foulsham, and A. Kingstone, "Speaking and listening with the eyes: Gaze signaling during dyadic interactions," *PLOS ONE*, vol. 10, no. 8, pp. 1–18, 08 2015. [Online]. Available: https://doi.org/10.1371/journal.pone.0136905

[203] N. Garg, B. Favre, K. Riedhammer, and D. Z. Hakkani-Tür, "Clusterrank: a graph based method for meeting summarization," in *INTERSPEECH*, 2009.

[204] V. Rennard, G. Shang, J. Hunter, and M. Vazirgiannis, "Abstractive meeting summarization: A survey," *Transactions of the Association for Computational Linguistics*, vol. 11, pp. 861–884, 2022.

[205] A. Kumbhar, H. Kulkarni, A. Mali, S. Sonawane, and P. Mulay, "The current landscape of multimodal summarization," in *Proceedings of the 20th International Conference on Natural Language Processing (ICON)*, J. D. Pawar and S. Lalitha Devi, Eds. Goa University, Goa, India: NLP Association of India (NLPAI), Dec. 2023, p. 797–806. [Online]. Available: https://aclanthology.org/2023.icon-1.82/

[206] M. A. A. Dewan, M. Murshed, and F. Lin, "Engagement detection in online learning: a review," *Smart Learning Environments*, vol. 6, no. 1, p. 1, Jan. 2019. [Online]. Available: https://doi.org/10.1186/s40561-018-0080-z

[207] M. Wang, L. Jiang, and H. Luo, "Dyads or quads? impact of group size and learning context on collaborative learning," *Frontiers in Psychology*, vol. 14, 2023.

[208] S. Glucksberg and G. N. Cowen, "Memory for nonattended auditory material," *Cognitive Psychology*, vol. 1, no. 2, pp. 149–156, 1970. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0010028570900101

[209] F. Gobet and H. A. Simon, "Five seconds or sixty? presentation time in expert memory," *Cognitive Science*, vol. 24, no. 4, p. 651–682, 2000.

[210] H. Sloetjes and P. Wittenburg, "Annotation by category - elan and iso dcr," in *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*. Marrakech, Morocco: European Language Resources Association (ELRA), 2008.

[211] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *biometrics*, pp. 159–174, 1977.

[212] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[213] E. Zormpa, *Memory for speaking and listening*, 2020. [Online]. Available: https://repository.ubn.ru.nl/handle/2066/227383

[214] G. Doherty-Sneddon and F. G. Phelps, "Gaze aversion: A response to cognitive or social difficulty?" *Memory & Cognition*, vol. 33, no. 4, pp. 727–733, Jun 2005. [Online]. Available: https://doi.org/10.3758/BF03195338

[215] A. M. Glenberg, J. L. Schroeder, and D. A. Robertson, "Averting the gaze disengages the environment and facilitates remembering," *Memory & Cognition*, vol. 26, no. 4, pp. 651–658, Jul 1998. [Online]. Available: https://doi.org/10.3758/BF03211385

[216] K. Truong and D. Heylen, "Disambiguating the functions of conversational sounds with prosody: The case of 'yeah'," in *Clinical Physics and Physiological Measurement*, 2010, pp. 2554–2557.

[217] J. Vargas-Quiros, L. Cabrera-Quiros, C. Oertel, and H. Hung, "Impact of annotation modality on label quality and model performance in the automatic assessment of laughter in-the-wild," *IEEE Transactions on Affective Computing*, vol. 15, no. 2, pp. 519–534, 2023, green Open Access added to TU Delft Institutional Repository 'You share, we take care!' – Taverne project https://www.openaccess.nl/en/you-share-we-take-care Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

[218] S. Dhamija and T. E. Boult, "Automated mood-aware engagement prediction," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, Oct. 2017, pp. 1–8, iSSN: 2156-8111.

[219] P. Eberle, C. Schwarzinger, and C. Stary, "User modelling and cognitive user support: towards structured development," *Universal Access in the Information Society*, vol. 10, pp. 275–293, 2011.

[220] M. Barr, A. Kabir, B. Harris-Roxas, E. Comino, T. Jackson, A.-M. Crozier, B. Goodger, J. Finch, and M. Harris, "1034all-cause mortality in australia: impact of social isolation and living alone," *International Journal of Epidemiology*, 2021.

[221] J. Beller and A. Wagner, "Loneliness, social isolation, their synergistic interaction, and mortality." *Health Psychology*, vol. 37, no. 9, p. 808, 2018.

[222] T. J. Holwerda, D. J. Deeg, A. T. Beekman, T. G. Van Tilburg, M. L. Stek, C. Jonker, and R. A. Schoevers, "Feelings of loneliness, but not social isolation, predict dementia onset: results from the amsterdam study of the elderly (amstel)," *Journal of Neurology, Neurosurgery & Psychiatry*, 2012.

[223] Y. Lee and Y. Ko, "Feeling lonely when not socially isolated," *Journal of Social and Personal Relationships*, vol. 35, pp. 1340–1355, 2017.

[224] K. Bosman, T. Bosse, and D. Formolo, *Virtual Agents for Professional Social Skills Training: An Overview of the State-of-the-Art*. Cham: Springer International Publishing, 2019, vol. 273, p. 75–84. [Online]. Available: http://link.springer.com/10.1007/978-3-030-16447-8_8

[225] A. Bagmar, K. Hogan, D. Shalaby, and J. Purtilo, "Analyzing the effectiveness of an extensible virtual moderator," *Proc. ACM Hum.-Comput. Interact.*, vol. 6, no. GROUP, jan 2022. [Online]. Available: https://doi-org.tudelft.idm.oclc.org/10.1145/3492837

[226] K. Loveys, G. Fricchione, K. Kolappa, M. Sagar, and E. Broadbent, "Reducing patient loneliness with artificial agents: Design insights from evolutionary neuropsychiatry," *Journal of Medical Internet Research*, vol. 21, 2019.

[227] H. Trinh, A. Shamekhi, E. Kimani, and T. Bickmore, "Predicting user engagement in longitudinal interventions with virtual agents," *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, 2018.

[228] C. Clavel, A. Cafaro, S. Campano, and C. Pelachaud, "Fostering user engagement in face-to-face human-agent interactions: A survey," in *Toward Robotic Socially Believable Behaving Systems*, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:30577188

[229] I. Lopatovska, O. Turpin, J. Yoon, D. Brown, L. Vroom, C. Nielsen, K. Hayes, K. Roslund, M. Dickson, and D. Anger, "Measuring the impact of conversational technology interventions on adolescent wellbeing: Quantitative and qualitative approaches," *Proceedings*

of the Association for Information Science and Technology, vol. 59, 2022.

[230] A. Saravanan, M. Tsfasman, M. Neerincx, and C. Oertel, "Giving social robots a conversational memory for motivational experience sharing," in *Proceedings of 31st IEEE International Conference on Robot & Human Interactive Communication*. IEEE, 2022.

[231] B. Dudzik, H. Hung, M. Neerincx, and J. Broekens, "Artificial Empathic Memory: Enabling Media Technologies to Better Understand Subjective User Experience," in *Proceedings of the 2018 Workshop on Understanding Subjective Attributes of Data, with the Focus on Evoked Emotions - EE-USAD'18*. New York, New York, USA: ACM Press, 2018, pp. 1–8.

[232] ——, "Investigating the Influence of Personal Memories on Video-Induced Emotions," in *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*. New York, NY, USA: ACM, Jul. 2020, pp. 53–61.

[233] H. J. Escalante, H. Kaya, A. A. Salah, S. Escalera, Y. Güçlütürk, U. Güçlü, X. Baró, I. Guyon, J. C. S. J. Junior, M. Madadi, S. Ayache, E. Viegas, F. Gürpınar, A. S. Wicaksana, C. C. S. Liem, M. A. J. van Gerven, and R. van Lier, "Modeling, recognizing, and explaining apparent personality from videos," *IEEE Transactions on Affective Computing*, vol. 13, no. 2, pp. 894–911, 2022.

[234] M. Lea and P. Rogers, "Cohesion in online groups," *WIT Tran Inf Com Techn*, vol. 31, 2004.

[235] W. Piper, M. Marrache, R. Lacroix, A. Richardsen, and B. D. Jones, "Cohesion as a basic bond in groups," *Human Relations*, vol. 36, 1983.

[236] "Final cut pro." [Online]. Available: https://www.apple.com/nl/final-cut-pro/

ACKNOWLEDGEMENTS

A lot of people have been by my side throughout this PhD journey, and to them I am eternally grateful and will try to express the inexpressible here.

I thank my defence committee members - **Alan Hanjalic, Dirk Heylen, Mark Neerincx, Albert Salah, Odette Scharenborg** - for taking the time to read this thesis and prepare questions for my defence. Thank you, **Willem-Paul**, for stepping in for Catholijn for the purpose of my defence.

I thank **Catholijn** for making sure I didn't need to worry about funding and that my research and experiments were always supported financially, it was security that not many PhDs and postdocs can have and I don't take it for granted. **Bernd**, thank you for stepping in as my supervisor at the moment when I was ready to quit. I have learned more from you in the last 2 years than since the beginning of my PhD. I know it has not been easy to come in at that time, but you managed it brilliantly, and I hope we can go back to being friends (no hierarchy in the way) and laughing at the situations we've navigated through together after this is finished.

Thanks to my collaborators - **Navin, Kristian, and Andras**, to name a few. Kristian and Andras, without you, MeMo would not be the same, it was lovely working with you!

Thank you **Karin van Nispen**, for organising PhD writing sessions and sparking the idea for my PhD co-writing group. My PhD would not be complete without your dissertation writing course and individual coaching sessions. I owe approximately 90% of my writing progress to my PhD co-writing group. I, therefore, thank my writing partners - **Ellen, Seline, Aashna, Jinke, Katerina, Nanda, Salwa, David, Sacha, Suriya, Figen, Burcu, and others** - for keeping me accountable, engaged and motivated. I wrote most of my thesis with you by my side and you made it fun!

I am also grateful to TU Delft confidential advisors - **Ada van Gulik**, you were the first person of power at TU Delft who truly listened and took action when I didn't think I could go on. Your empathy and listening ear truly meant a lot when no one else involved would take my problems seriously. Thank you for suggesting Leo van Iersel as my mentor.

Leo, thank you so much for being on my side and advocating for me when I needed it the most. You are a great mentor and without you I would probably have quit. Thank you for giving me the courage to stand

up for myself and bring Bernd onto the supervisory team.

Thank you Nele and Jose for being absolutely incredible paranymphs and helping me so much in the organisation of celebrations and finalising the thesis for printing!

Humongous thanks go to all my friends! Chronologically, my first thanks go to **Joanna, Nele and Dima** - we started at the exact same time and having you two to figure out what we are doing with has been lovely. I've started this PhD full of curiosity and love for science and academia, and even though I am coming out on the other side feeling chewed up and digested by the academia machine, I have gained lots of friends along the way and you were the first ones! **Joanna**, I'm glad I have started this PhD with you by my side, you are such a lovely person! I wish I had listened to you more and realised that the issues you raised were not unique and showed a pattern that would emerge throughout the months after you left, which would cause me and many other students more suffering.

Nele, you are my highlight and my guiding star during my time in Delft from the very first day of being in Delft. From inviting me to your birthday party, you have been my rock and my foundation, you have been by my side, supporting me in bad and good times, crying and laughing with me, and I can't remember life without you anymore <3.

Dima, I'm glad I met you on my very first train to Delft, not realising how aligned we are in our thinking and patterns of creative obsessions! It was great knowing that you would always be there at the photography studio whenever I came to do some ceramics at X. You kept me active with your tennis and then badminton and kept me curious about the next funny and ridiculous life stories you'd bring to the next meetup.

Jose, even though we never ended up publishing together, those several conversations with you at the very beginning were true inspiration for my MeMo corpus and the subsequent work in this thesis. I'm glad that our friendship worked out better than our never-published joint research project :D. Still hoping to join you at one handmade market one day and to propel each other's creative endeavours into a full-on business. Also, I owe (some of) my thesis progress to our co-working sessions:). Still waiting on that startup idea by the way!

Carolina and Enrico, thank you for being on my side, listening to me complain, and being a shoulder to cry on when coming into the office felt like an unbearable effort. Thank you for being the social centres of our group, organising ice cream outings, and board game/ cooking nights. Carolina, thank you for being my advocate, my running partner (I really didn't know I could run that fast><), and always being ready to dive deep into a psychological debate.

Mani, thank you for being a safe space for me any time we talked, I loved having your warm and welcoming energy around the office, and I just love discussing life with you. Cheers to more conversations to

come!

Micha, thanks for keeping me procrastinating at the office, our gym training sessions and an 80s disco workout (still wish you'd dress up for it xD). Hope there are more Hyrox-es to come with you and that we find that golden startup idea soon!

Mo, thanks for being a calming presence, my board game pal, and being always up for talking about Japan! **Zuzanna**, thank you for making me fitter with our fun gym & chat sessions, thanks for keeping me up to date on what's happening and thanks for your fun energy around the office! Thank you **Morita and Deborah** for keeping me company in our office. I hope that you can stand up for yourselves and acknowledge the issues earlier than I did. Thank you, **Stephanie** for sharing my office for a bit, too, and getting me out of a panic attack when it happened. I loved our scientific conversations and wish we'd met earlier throughout my PhD to bring one of our ideas to publication!

Tiffany, Chenxu, thanks for our walks and meetups. Tiffany, it was my honour to be your paranymph, and thank you for all the SSP discussions and keeping me interested in social computing. Conversations with you definitely inspired me and made my thesis better. Also loved exploring Bangalore with you and Bernd!

Thanks to **Wendy van Aartsen** for organising fun Hybrid Intelligence meet ups and a lovely writing retreat. It definitely kept my spirits up throughout these years. Thank you to all **HI PhDs** I haven't yet mentioned for fun and inspiring discussions. **Merle**, thanks for your cheerfulness and lucky clovers!

Anita, thank you for your support along the way, you were always in touch to answer my many administrative and organisational questions. Thank you, **Ruud**, for keeping my devices backed up, fixing my laptop whenever it broke down, and for your friendly face at the office. You two really keep our group running with your organisational and tech support!

Senthil, Eric, Shambhawi, Rolf, Agnes, Myrthe, Willem-Paul, Pradeep, Luciano, Wouter, Linyun, Laxmi, Yanzhe, Ruben, Pei-Yu, Paul, Antonio, Amir, Miguel, Elena, Sietze, Emma, Sid, Frank, and Davide, thank you for fun lunches, useful advice, and your friendly smiles!

Thanks to all the MSc and BSc students I have worked with. With your enthusiasm and curiosity, you'll go a long way!

Thanks to **Hayley Hung, Chirag, Vandana, Zonghuan, Ojas, Ivan, and others** from the Socially Perceptive Computing Lab for fun reading groups and discussions at our biweekly meetings.

Thanks to my friends outside the PhD world, such as **Anechka, Daphne, Jochem, and Gabriel**. You are always great to be around and I'm so proud to be your friend! **Daphne**, creative and co-working sessions with you were always a highlight of my month:). Cheers to more foraging, camping, and creative obsessions! **Jochem and**

Gabriel, you have been so lovely to come home to, best neighbours ever. Loved popping by whenever I needed someone to talk to, and loved our movie nights, board game evenings, and coffee talks. You have been my family in Delft and kept my spirits up throughout this thesis!

Dear-dear **Jay**, even though you have always been the voice that said "Please quit, this is not worth the amount of suffering in your made-up academia bubble", I owe the completion of this thesis to you! Even on the worst days, I always had your embrace to come back home to and this is worth everything. Thank you for tolerating my complaining and getting me through countless breakdowns. Also, of course, thank you for designing the thesis cover. Love you!

Thank you to my family - **mum, dad, Tanyushka, Liska, Fedyushka** (and your partners) - even though you are far and spread out around the world I feel your presence and support wherever I go, you bring a lot of joy into my life! **Vincent, Nael, Nina, Elemelech, Solomon,** I wish I abandoned this PhD ages ago to be your stay-at-home full-time auntie ;D.

Thanks to all open studio assistants and organisers at X for doing ceramics by my side and keeping this amazing space running. Thank you **Joris and Cora** from 3Trees for providing me with a kiln and fun ceramics classes, and a space to be unapologetically me with wine and tea.

At last, I would like to thank all the places that kept me going throughout these years and allowed me to work there from time to time - **LOT, Piada, Open, TU Delft library**. Thanks to all the X classes I attended (to name a few, Pilates with **Shamangi** and Yoga with **Astrid and Jane** have gotten me out of a freeze/ panic state to calm and hopefulness countless times). Thanks to the **Je van het loket** burrito truck for serving me the best lunch ever and fueling this thesis!

CURRICULUM VITAE

Maria Tsfasman

EDUCATION

2020–2026	**PhD in Computer Science** Delft University of Technology, Delft, the Netherlands
2018–2020	**MSc in Artificial Intelligence, Cum Laude** Radboud University, Nijmegen, the Netherlands
2014–2018	**BSc in Fundamental and Computational Linguistics** HSE, Moscow, Russia
2017	Exchange Semester with Cognitive Linguistics Major, UiT The Arctic University of Norway, Tromso, Norway

EXPERIENCE

2020–2024	**PhD Researcher in Computer Science** Delft University of Technology, Delft, the Netherlands
2020	**Research Intern, Developmental Cognitive Robotics Lab** IRCN, University of Tokyo, Japan
2019	**Teaching Assistant, MSc "RobotLab Practical" Course** Radboud University, Nijmegen, the Netherlands
2019	**Research Intern, "Architectures and Models for Adaptation and Cognition"** Institut des Systèmes Intelligents et de Robotique, Paris, France
2017–2018	**Research Intern, Laboratory of Neurocognitive Technology** Kurchatov Institute, Moscow, Russia

LIST OF PUBLICATIONS

2025

1. *[pre-print]* **M Tsfasman**, R Ghorbani, CM Jonker, B Dudzik, "The Emotion-Memory Link: Do Memorability Annotations Matter for Intelligent Systems?" *arXiv preprint arXiv:2507.14084*, 2025.

2024

2. *[pre-print]* NR Prabhu, **M Tsfasman**, C Oertel, T Gerkmann, N Lehmann-Willenbrock, "Dynamics of collective group affect: Group-level annotations and the multimodal modeling of convergence and divergence," *arXiv preprint arXiv:2409.08578*, 2024.

3. *[pre-print]* **M Tsfasman**, B Dudzik, K Fenech, A Lorincz, CM Jonker, C Oertel, "Introducing MeMo: A Multimodal Dataset for Memory Modelling in Multiparty Conversations," *arXiv preprint arXiv:2409.13715*, 2024.

4. B Dudzik, T Matej Hrkalovic, C Hao, C Raman, **M Tsfasman**, "Indeterminacy in Affective Computing: Considering Meaning and Context in Data Collection Practices," in *Proc. 12th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pp. 181–185, IEEE, 2024. 10.1109/ACIIW63320.2024.00036

2022

5. **M Tsfasman**, A Saravanan, MA Neerincx, C Oertel, "Giving social robots a conversational memory for motivational experience sharing," in *Proc. 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2022.

6. **M Tsfasman**, K Fenech, M Tarvirdians, A Lorincz, C Jonker, C Oertel, "Towards creating a conversational memory for long-term meeting support: predicting memorable moments in multi-party conversations through eye-gaze," in *Proc. International Conference on Multimodal Interaction (ICMI)*, pp. 94–104, 2022.

7. **M Tsfasman**, A Philippsen, C Mazzola, S Thill, A Sciutti, Y Nagai, “The world seems different in a social context: A neural network analysis of human experimental data,” *PLoS ONE*, vol. 17, no. 8, 2022.

8. **M Tsfasman**, A Saravanan, D Viner, D Goslinga, S De Wolf, C Raman, CM Jonker, C Oertel, “Towards a Real-time Measure of the Perception of Anthropomorphism in Human-robot Interaction,” in *Proc. 2nd ACM Multimedia Workshop on Multimodal Conversational AI*, 2021.

2019

9. M Daniel, R von Waldenfels, A Ter-Avanesova, P Kazakova, I Schurov, E Gerasimenko, D Ignatenko, E Makhlina, **M Tsfasman** and S Verhees, “Dialect loss in the Russian North: Modeling change across variables,” *Language Variation and Change*, vol. 31, no. 3, pp. 353–376, 2019.

2017

10. **M Tsfasman**, “Determining Children’s Level of Acquisition through Grammatical Profiles: Evidence from a Bilingual Russian-English Child Acquiring Verbs,” *Poljarnyj Vestnik*, no. 20, pp. 45–55, 2017.

SIKS DISSERTATIONS

2016 01 Syed Saiden Abbas (RUN), Recognition of Shapes by Humans and Machines

02 Michiel Christiaan Meulendijk (UU), Optimizing medication reviews through decision support: prescribing a better pill to swallow

03 Maya Sappelli (RUN), Knowledge Work in Context: User Centered Knowledge Worker Support

04 Laurens Rietveld (VUA), Publishing and Consuming Linked Data

05 Evgeny Sherkhonov (UvA), Expanded Acyclic Queries: Containment and an Application in Explaining Missing Answers

06 Michel Wilson (TUD), Robust scheduling in an uncertain environment

07 Jeroen de Man (VUA), Measuring and modeling negative emotions for virtual training

08 Matje van de Camp (TiU), A Link to the Past: Constructing Historical Social Networks from Unstructured Data

09 Archana Nottamkandath (VUA), Trusting Crowdsourced Information on Cultural Artefacts

10 George Karafotias (VUA), Parameter Control for Evolutionary Algorithms

11 Anne Schuth (UvA), Search Engines that Learn from Their Users

12 Max Knobbout (UU), Logics for Modelling and Verifying Normative Multi-Agent Systems

13 Nana Baah Gyan (VUA), The Web, Speech Technologies and Rural Development in West Africa - An ICT4D Approach

14 Ravi Khadka (UU), Revisiting Legacy Software System Modernization

15 Steffen Michels (RUN), Hybrid Probabilistic Logics - Theoretical Aspects, Algorithms and Experiments

16 Guangliang Li (UvA), Socially Intelligent Autonomous Agents that Learn from Human Reward

17 Berend Weel (VUA), Towards Embodied Evolution of Robot Organisms

18 Albert Meroño Peñuela (VUA), Refining Statistical Data on the Web

19 Julia Efremova (TU/e), Mining Social Structures from Genealogical Data

20 Daan Odijk (UvA), Context & Semantics in News & Web Search

21 Alejandro Moreno Célleri (UT), From Traditional to Interactive Playspaces: Automatic Analysis of Player Behavior in the Interactive Tag Playground

22 Grace Lewis (VUA), Software Architecture Strategies for Cyber-Foraging Systems

23 Fei Cai (UvA), Query Auto Completion in Information Retrieval

24 Brend Wanders (UT), Repurposing and Probabilistic Integration of Data; An Iterative and data model independent approach

25 Julia Kiseleva (TU/e), Using Contextual Information to Understand Searching and Browsing Behavior

26 Dilhan Thilakarathne (VUA), In or Out of Control: Exploring Computational Models to Study the Role of Human Awareness and Control in Behavioural Choices, with Applications in Aviation and Energy Management Domains

27 Wen Li (TUD), Understanding Geo-spatial Information on Social Media

28 Mingxin Zhang (TUD), Large-scale Agent-based Social Simulation - A study on epidemic prediction and control

29 Nicolas Höning (TUD), Peak reduction in decentralised electricity systems - Markets and prices for flexible planning

30 Ruud Mattheij (TiU), The Eyes Have It

31 Mohammad Khelghati (UT), Deep web content monitoring

32 Eelco Vriezekolk (UT), Assessing Telecommunication Service Availability Risks for Crisis Organisations

33 Peter Bloem (UvA), Single Sample Statistics, exercises in learning from just one example

34 Dennis Schunselaar (TU/e), Configurable Process Trees: Elicitation, Analysis, and Enactment

35 Zhaochun Ren (UvA), Monitoring Social Media: Summarization, Classification and Recommendation

36 Daphne Karreman (UT), Beyond R2D2: The design of nonverbal interaction behavior optimized for robot-specific morphologies

37 Giovanni Sileno (UvA), Aligning Law and Action - a conceptual and computational inquiry

38 Andrea Minuto (UT), Materials that Matter - Smart Materials meet Art & Interaction Design

39 Merijn Bruijnes (UT), Believable Suspect Agents; Response and Interpersonal Style Selection for an Artificial Suspect

40 Christian Detweiler (TUD), Accounting for Values in Design

41 Thomas King (TUD), Governing Governance: A Formal Framework for Analysing Institutional Design and Enactment Governance

42 Spyros Martzoukos (UvA), Combinatorial and Compositional Aspects of Bilingual Aligned Corpora

43 Saskia Koldijk (RUN), Context-Aware Support for Stress Self-Management: From Theory to Practice

44 Thibault Sellam (UvA), Automatic Assistants for Database Exploration

45 Bram van de Laar (UT), Experiencing Brain-Computer Interface Control

46 Jorge Gallego Perez (UT), Robots to Make you Happy

47 Christina Weber (UL), Real-time foresight - Preparedness for dynamic innovation networks

48 Tanja Buttler (TUD), Collecting Lessons Learned

49 Gleb Polevoy (TUD), Participation and Interaction in Projects. A Game-Theoretic Analysis

50 Yan Wang (TiU), The Bridge of Dreams: Towards a Method for Operational Performance Alignment in IT-enabled Service Supply Chains

2017 01 Jan-Jaap Oerlemans (UL), Investigating Cybercrime

02 Sjoerd Timmer (UU), Designing and Understanding Forensic Bayesian Networks using Argumentation

03 Daniël Harold Telgen (UU), Grid Manufacturing; A Cyber-Physical Approach with Autonomous Products and Reconfigurable Manufacturing Machines

04 Mrunal Gawade (CWI), Multi-core Parallelism in a Column-store

05 Mahdieh Shadi (UvA), Collaboration Behavior

06 Damir Vandic (EUR), Intelligent Information Systems for Web Product Search

07 Roel Bertens (UU), Insight in Information: from Abstract to Anomaly

08 Rob Konijn (VUA), Detecting Interesting Differences:Data Mining in Health Insurance Data using Outlier Detection and Subgroup Discovery

09 Dong Nguyen (UT), Text as Social and Cultural Data: A Computational Perspective on Variation in Text

10 Robby van Delden (UT), (Steering) Interactive Play Behavior

11 Florian Kunneman (RUN), Modelling patterns of time and emotion in Twitter #anticipointment

12 Sander Leemans (TU/e), Robust Process Mining with Guarantees

13 Gijs Huisman (UT), Social Touch Technology - Extending the reach of social touch through haptic technology

14 Shoshannah Tekofsky (TiU), You Are Who You Play You Are: Modelling Player Traits from Video Game Behavior

15 Peter Berck (RUN), Memory-Based Text Correction

16 Aleksandr Chuklin (UvA), Understanding and Modeling Users of Modern Search Engines

17 Daniel Dimov (UL), Crowdsourced Online Dispute Resolution

18 Ridho Reinanda (UvA), Entity Associations for Search

19 Jeroen Vuurens (UT), Proximity of Terms, Texts and Semantic Vectors in Information Retrieval

20 Mohammadbashir Sedighi (TUD), Fostering Engagement in Knowledge Sharing: The Role of Perceived Benefits, Costs and Visibility

21 Jeroen Linssen (UT), Meta Matters in Interactive Storytelling and Serious Gaming (A Play on Worlds)

22 Sara Magliacane (VUA), Logics for causal inference under uncertainty

23 David Graus (UvA), Entities of Interest — Discovery in Digital Traces

24 Chang Wang (TUD), Use of Affordances for Efficient Robot Learning

25 Veruska Zamborlini (VUA), Knowledge Representation for Clinical Guidelines, with applications to Multimorbidity Analysis and Literature Search

26 Merel Jung (UT), Socially intelligent robots that understand and respond to human touch

27 Michiel Joosse (UT), Investigating Positioning and Gaze Behaviors of Social Robots: People’s Preferences, Perceptions and Behaviors

28 John Klein (VUA), Architecture Practices for Complex Contexts

29 Adel Alhuraibi (TiU), From IT-BusinessStrategic Alignment to Performance: A Moderated Mediation Model of Social Innovation, and Enterprise Governance of IT"

30 Wilma Latuny (TiU), The Power of Facial Expressions

31 Ben Ruijl (UL), Advances in computational methods for QFT calculations

32 Thaer Samar (RUN), Access to and Retrievability of Content in Web Archives

33 Brigit van Loggem (OU), Towards a Design Rationale for Software Documentation: A Model of Computer-Mediated Activity

34 Maren Scheffel (OU), The Evaluation Framework for Learning Analytics

35 Martine de Vos (VUA), Interpreting natural science spreadsheets

36 Yuanhao Guo (UL), Shape Analysis for Phenotype Characterisation from High-throughput Imaging

37 Alejandro Montes Garcia (TU/e), WiBAF: A Within Browser Adaptation Framework that Enables Control over Privacy

38 Alex Kayal (TUD), Normative Social Applications

39 Sara Ahmadi (RUN), Exploiting properties of the human auditory system and compressive sensing methods to increase noise robustness in ASR

40 Altaf Hussain Abro (VUA), Steer your Mind: Computational Exploration of Human Control in Relation to Emotions, Desires and Social Support For applications in human-aware support systems

41 Adnan Manzoor (VUA), Minding a Healthy Lifestyle: An Exploration of Mental Processes and a Smart Environment to Provide Support for a Healthy Lifestyle

42 Elena Sokolova (RUN), Causal discovery from mixed and missing data with applications on ADHD datasets

43 Maaike de Boer (RUN), Semantic Mapping in Video Retrieval

44 Garm Lucassen (UU), Understanding User Stories - Computational Linguistics in Agile Requirements Engineering

45 Bas Testerink (UU), Decentralized Runtime Norm Enforcement

46 Jan Schneider (OU), Sensor-based Learning Support

47 Jie Yang (TUD), Crowd Knowledge Creation Acceleration

48 Angel Suarez (OU), Collaborative inquiry-based learning

2018 01 Han van der Aa (VUA), Comparing and Aligning Process Representations

02 Felix Mannhardt (TU/e), Multi-perspective Process Mining

03 Steven Bosems (UT), Causal Models For Well-Being: Knowledge Modeling, Model-Driven Development of Context-Aware Applications, and Behavior Prediction

04 Jordan Janeiro (TUD), Flexible Coordination Support for Diagnosis Teams in Data-Centric Engineering Tasks

05 Hugo Huurdeman (UvA), Supporting the Complex Dynamics of the Information Seeking Process

06 Dan Ionita (UT), Model-Driven Information Security Risk Assessment of Socio-Technical Systems

07 Jieting Luo (UU), A formal account of opportunism in multi-agent systems

08 Rick Smetsers (RUN), Advances in Model Learning for Software Systems

09 Xu Xie (TUD), Data Assimilation in Discrete Event Simulations

10 Julienka Mollee (VUA), Moving forward: supporting physical activity behavior change through intelligent technology

11 Mahdi Sargolzaei (UvA), Enabling Framework for Service-oriented Collaborative Networks

12 Xixi Lu (TU/e), Using behavioral context in process mining

13 Seyed Amin Tabatabaei (VUA), Computing a Sustainable Future

14 Bart Joosten (TiU), Detecting Social Signals with Spatiotemporal Gabor Filters

15 Naser Davarzani (UM), Biomarker discovery in heart failure

16 Jaebok Kim (UT), Automatic recognition of engagement and emotion in a group of children

17 Jianpeng Zhang (TU/e), On Graph Sample Clustering

18 Henriette Nakad (UL), De Notaris en Private Rechtspraak

19 Minh Duc Pham (VUA), Emergent relational schemas for RDF

20 Manxia Liu (RUN), Time and Bayesian Networks

21 Aad Slootmaker (OU), EMERGO: a generic platform for authoring and playing scenario-based serious games

22 Eric Fernandes de Mello Araújo (VUA), Contagious: Modeling the Spread of Behaviours, Perceptions and Emotions in Social Networks

23 Kim Schouten (EUR), Semantics-driven Aspect-Based Sentiment Analysis

24 Jered Vroon (UT), Responsive Social Positioning Behaviour for Semi-Autonomous Telepresence Robots

25 Riste Gligorov (VUA), Serious Games in Audio-Visual Collections

26 Roelof Anne Jelle de Vries (UT),Theory-Based and Tailor-Made: Motivational Messages for Behavior Change Technology

27 Maikel Leemans (TU/e), Hierarchical Process Mining for Scalable Software Analysis

28 Christian Willemse (UT), Social Touch Technologies: How they feel and how they make you feel

29 Yu Gu (TiU), Emotion Recognition from Mandarin Speech

30 Wouter Beek (VUA), The "K" in "semantic web" stands for "knowledge": scaling semantics to the web

---

2019 01 Rob van Eijk (UL),Web privacy measurement in real-time bidding systems. A graph-based approach to RTB system classification

02 Emmanuelle Beauxis Aussalet (CWI, UU), Statistics and Visualizations for Assessing Class Size Uncertainty

03 Eduardo Gonzalez Lopez de Murillas (TU/e), Process Mining on Databases: Extracting Event Data from Real Life Data Sources

04 Ridho Rahmadi (RUN), Finding stable causal structures from clinical data

05 Sebastiaan van Zelst (TU/e), Process Mining with Streaming Data

06 Chris Dijkshoorn (VUA), Nichesourcing for Improving Access to Linked Cultural Heritage Datasets

07 Soude Fazeli (TUD), Recommender Systems in Social Learning Platforms

08 Frits de Nijs (TUD), Resource-constrained Multi-agent Markov Decision Processes

09 Fahimeh Alizadeh Moghaddam (UvA), Self-adaptation for energy efficiency in software systems

10 Qing Chuan Ye (EUR), Multi-objective Optimization Methods for Allocation and Prediction

11 Yue Zhao (TUD), Learning Analytics Technology to Understand Learner Behavioral Engagement in MOOCs

12 Jacqueline Heinerman (VUA), Better Together

13 Guanliang Chen (TUD), MOOC Analytics: Learner Modeling and Content Generation

14 Daniel Davis (TUD), Large-Scale Learning Analytics: Modeling Learner Behavior & Improving Learning Outcomes in Massive Open Online Courses

15 Erwin Walraven (TUD), Planning under Uncertainty in Constrained and Partially Observable Environments

16 Guangming Li (TU/e), Process Mining based on Object-Centric Behavioral Constraint (OCBC) Models

17 Ali Hurriyetoglu (RUN),Extracting actionable information from microtexts

18 Gerard Wagenaar (UU), Artefacts in Agile Team Communication

19 Vincent Koeman (TUD), Tools for Developing Cognitive Agents

20 Chide Groenouwe (UU), Fostering technically augmented human collective intelligence

21 Cong Liu (TU/e), Software Data Analytics: Architectural Model Discovery and Design Pattern Detection

22 Martin van den Berg (VUA),Improving IT Decisions with Enterprise Architecture

23 Qin Liu (TUD), Intelligent Control Systems: Learning, Interpreting, Verification

24 Anca Dumitrache (VUA), Truth in Disagreement - Crowdsourcing Labeled Data for Natural Language Processing

25 Emiel van Miltenburg (VUA), Pragmatic factors in (automatic) image description

26 Prince Singh (UT), An Integration Platform for Synchromodal Transport

27 Alessandra Antonaci (OU), The Gamification Design Process applied to (Massive) Open Online Courses

28 Esther Kuindersma (UL), Cleared for take-off: Game-based learning to prepare airline pilots for critical situations

29 Daniel Formolo (VUA), Using virtual agents for simulation and training of social skills in safety-critical circumstances

30 Vahid Yazdanpanah (UT), Multiagent Industrial Symbiosis Systems

31 Milan Jelisavcic (VUA), Alive and Kicking: Baby Steps in Robotics

32 Chiara Sironi (UM), Monte-Carlo Tree Search for Artificial General Intelligence in Games

33 Anil Yaman (TU/e), Evolution of Biologically Inspired Learning in Artificial Neural Networks

34 Negar Ahmadi (TU/e), EEG Microstate and Functional Brain Network Features for Classification of Epilepsy and PNES

35 Lisa Facey-Shaw (OU), Gamification with digital badges in learning programming

36 Kevin Ackermans (OU), Designing Video-Enhanced Rubrics to Master Complex Skills

37 Jian Fang (TUD), Database Acceleration on FPGAs

38 Akos Kadar (OU), Learning visually grounded and multilingual representations

---

2020 01 Armon Toubman (UL), Calculated Moves: Generating Air Combat Behaviour

02 Marcos de Paula Bueno (UL), Unraveling Temporal Processes using Probabilistic Graphical Models

03 Mostafa Deghani (UvA), Learning with Imperfect Supervision for Language Understanding

04 Maarten van Gompel (RUN), Context as Linguistic Bridges

05 Yulong Pei (TU/e), On local and global structure mining

06 Preethu Rose Anish (UT), Stimulation Architectural Thinking during Requirements Elicitation - An Approach and Tool Support

07 Wim van der Vegt (OU), Towards a software architecture for reusable game components

08 Ali Mirsoleimani (UL),Structured Parallel Programming for Monte Carlo Tree Search

09 Myriam Traub (UU), Measuring Tool Bias and Improving Data Quality for Digital Humanities Research

10 Alifah Syamsiyah (TU/e), In-database Preprocessing for Process Mining

11 Sepideh Mesbah (TUD), Semantic-Enhanced Training Data AugmentationMethods for Long-Tail Entity Recognition Models

12 Ward van Breda (VUA), Predictive Modeling in E-Mental Health: Exploring Applicability in Personalised Depression Treatment

13 Marco Virgolin (CWI), Design and Application of Gene-pool Optimal Mixing Evolutionary Algorithms for Genetic Programming

14 Mark Raasveldt (CWI/UL), Integrating Analytics with Relational Databases

15 Konstantinos Georgiadis (OU), Smart CAT: Machine Learning for Configurable Assessments in Serious Games

16 Ilona Wilmont (RUN), Cognitive Aspects of Conceptual Modelling

17 Daniele Di Mitri (OU), The Multimodal Tutor: Adaptive Feedback from Multimodal Experiences

18 Georgios Methenitis (TUD), Agent Interactions & Mechanisms in Markets with Uncertainties: Electricity Markets in Renewable Energy Systems

19 Guido van Capelleveen (UT), Industrial Symbiosis Recommender Systems

20 Albert Hankel (VUA), Embedding Green ICT Maturity in Organisations

21 Karine da Silva Miras de Araujo (VUA), Where is the robot?: Life as it could be

22 Maryam Masoud Khamis (RUN), Understanding complex systems implementation through a modeling approach: the case of e-government in Zanzibar

23 Rianne Conijn (UT), The Keys to Writing: A writing analytics approach to studying writing processes using keystroke logging

24 Lenin da Nóbrega Medeiros (VUA/RUN), How are you feeling, human? Towards emotionally supportive chatbots

25 Xin Du (TU/e), The Uncertainty in Exceptional Model Mining

26 Krzysztof Leszek Sadowski (UU), GAMBIT: Genetic Algorithm for Model-Based mixed-Integer opTimization

27 Ekaterina Muravyeva (TUD), Personal data and informed consent in an educational context

28 Bibeg Limbu (TUD), Multimodal interaction for deliberate practice: Training complex skills with augmented reality

29 Ioan Gabriel Bucur (RUN), Being Bayesian about Causal Inference

30 Bob Zadok Blok (UL), Creatief, Creatiever, Creatiefst

31 Gongjin Lan (VUA), Learning better – From Baby to Better

32 Jason Rhuggenaath (TU/e), Revenue management in online markets: pricing and online advertising

33 Rick Gilsing (TU/e), Supporting service-dominant business model evaluation in the context of business model innovation

34 Anna Bon (UM), Intervention or Collaboration? Redesigning Information and Communication Technologies for Development

35 Siamak Farshidi (UU), Multi-Criteria Decision-Making in Software Production

---

2021 01 Francisco Xavier Dos Santos Fonseca (TUD),Location-based Games for Social Interaction in Public Space

02 Rijk Mercuur (TUD), Simulating Human Routines: Integrating Social Practice Theory in Agent-Based Models

03 Seyyed Hadi Hashemi (UvA), Modeling Users Interacting with Smart Devices

04 Ioana Jivet (OU), The Dashboard That Loved Me: Designing adaptive learning analytics for self-regulated learning

05 Davide Dell'Anna (UU), Data-Driven Supervision of Autonomous Systems

06 Daniel Davison (UT), "Hey robot, what do you think?" How children learn with a social robot

07 Armel Lefebvre (UU), Research data management for open science

08 Nardie Fanchamps (OU), The Influence of Sense-Reason-Act Programming on Computational Thinking

09 Cristina Zaga (UT), The Design of Robothings. Non-Anthropomorphic and Non-Verbal Robots to Promote Children's Collaboration Through Play

10 Quinten Meertens (UvA), Misclassification Bias in Statistical Learning

11 Anne van Rossum (UL), Nonparametric Bayesian Methods in Robotic Vision

12 Lei Pi (UL), External Knowledge Absorption in Chinese SMEs

13 Bob R. Schadenberg (UT), Robots for Autistic Children: Understanding and Facilitating Predictability for Engagement in Learning

14 Negin Samaeemofrad (UL), Business Incubators: The Impact of Their Support

15 Onat Ege Adali (TU/e), Transformation of Value Propositions into Resource Re-Configurations through the Business Services Paradigm

16 Esam A. H. Ghaleb (UM), Bimodal emotion recognition from audio-visual cues

17 Dario Dotti (UM), Human Behavior Understanding from motion and bodily cues using deep neural networks

18 Remi Wieten (UU), Bridging the Gap Between Informal Sense-Making Tools and Formal Systems - Facilitating the Construction of Bayesian Networks and Argumentation Frameworks

19 Roberto Verdecchia (VUA), Architectural Technical Debt: Identification and Management

20 Masoud Mansoury (TU/e), Understanding and Mitigating Multi-Sided Exposure Bias in Recommender Systems

21 Pedro Thiago Timbó Holanda (CWI), Progressive Indexes

22 Sihang Qiu (TUD), Conversational Crowdsourcing

23 Hugo Manuel Proença (UL), Robust rules for prediction and description

24 Kaijie Zhu (TU/e), On Efficient Temporal Subgraph Query Processing

25 Eoin Martino Grua (VUA), The Future of E-Health is Mobile: Combining AI and Self-Adaptation to Create Adaptive E-Health Mobile Applications

26 Benno Kruit (CWI/VUA), Reading the Grid: Extending Knowledge Bases from Human-readable Tables

27 Jelte van Waterschoot (UT), Personalized and Personal Conversations: Designing Agents Who Want to Connect With You

28 Christoph Selig (UL), Understanding the Heterogeneity of Corporate Entrepreneurship Programs

2022 01 Judith van Stegeren (UT), Flavor text generation for role-playing video games

02 Paulo da Costa (TU/e), Data-driven Prognostics and Logistics Optimisation: A Deep Learning Journey

03 Ali el Hassouni (VUA), A Model A Day Keeps The Doctor Away: Reinforcement Learning For Personalized Healthcare

04 Ünal Aksu (UU), A Cross-Organizational Process Mining Framework

05 Shiwei Liu (TU/e), Sparse Neural Network Training with In-Time Over-Parameterization

06 Reza Refaei Afshar (TU/e), Machine Learning for Ad Publishers in Real Time Bidding

07 Sambit Praharaj (OU), Measuring the Unmeasurable? Towards Automatic Co-located Collaboration Analytics

08 Maikel L. van Eck (TU/e), Process Mining for Smart Product Design

09 Oana Andreea Inel (VUA), Understanding Events: A Diversity-driven Human-Machine Approach

10 Felipe Moraes Gomes (TUD), Examining the Effectiveness of Collaborative Search Engines

11 Mirjam de Haas (UT), Staying engaged in child-robot interaction, a quantitative approach to studying preschoolers' engagement with robots and tasks during second-language tutoring

12 Guanyi Chen (UU), Computational Generation of Chinese Noun Phrases

13 Xander Wilcke (VUA), Machine Learning on Multimodal Knowledge Graphs: Opportunities, Challenges, and Methods for Learning on Real-World Heterogeneous and Spatially-Oriented Knowledge

14 Michiel Overeem (UU), Evolution of Low-Code Platforms

15 Jelmer Jan Koorn (UU), Work in Process: Unearthing Meaning using Process Mining

16 Pieter Gijsbers (TU/e), Systems for AutoML Research

17 Laura van der Lubbe (VUA), Empowering vulnerable people with serious games and gamification

18 Paris Mavromoustakos Blom (TiU), Player Affect Modelling and Video Game Personalisation

19 Bilge Yigit Ozkan (UU), Cybersecurity Maturity Assessment and Standardisation

20 Fakhra Jabeen (VUA), Dark Side of the Digital Media - Computational Analysis of Negative Human Behaviors on Social Media

21 Seethu Mariyam Christopher (UM), Intelligent Toys for Physical and Cognitive Assessments

22 Alexandra Sierra Rativa (TiU), Virtual Character Design and its potential to foster Empathy, Immersion, and Collaboration Skills in Video Games and Virtual Reality Simulations

23 Ilir Kola (TUD), Enabling Social Situation Awareness in Support Agents

24 Samaneh Heidari (UU), Agents with Social Norms and Values - A framework for agent based social simulations with social norms and personal values

25 Anna L.D. Latour (UL), Optimal decision-making under constraints and uncertainty

26 Anne Dirkson (UL), Knowledge Discovery from Patient Forums: Gaining novel medical insights from patient experiences

27 Christos Athanasiadis (UM), Emotion-aware cross-modal domain adaptation in video sequences

28 Onuralp Ulusoy (UU), Privacy in Collaborative Systems

29 Jan Kolkmeier (UT), From Head Transform to Mind Transplant: Social Interactions in Mixed Reality

30 Dean De Leo (CWI), Analysis of Dynamic Graphs on Sparse Arrays

31 Konstantinos Traganos (TU/e), Tackling Complexity in Smart Manufacturing with Advanced Manufacturing Process Management

32 Cezara Pastrav (UU), Social simulation for socio-ecological systems

33 Brinn Hekkelman (CWI/TUD), Fair Mechanisms for Smart Grid Congestion Management

34 Nimat Ullah (VUA), Mind Your Behaviour: Computational Modelling of Emotion & Desire Regulation for Behaviour Change

35 Mike E.U. Ligthart (VUA), Shaping the Child-Robot Relationship: Interaction Design Patterns for a Sustainable Interaction

---

2023 01 Bojan Simoski (VUA), Untangling the Puzzle of Digital Health Interventions

02 Mariana Rachel Dias da Silva (TiU), Grounded or in flight? What our bodies can tell us about the whereabouts of our thoughts

03 Shabnam Najafian (TUD), User Modeling for Privacy-preserving Explanations in Group Recommendations

04 Gineke Wiggers (UL), The Relevance of Impact: bibliometric-enhanced legal information retrieval

05 Anton Bouter (CWI), Optimal Mixing Evolutionary Algorithms for Large-Scale Real-Valued Optimization, Including Real-World Medical Applications

06 António Pereira Barata (UL), Reliable and Fair Machine Learning for Risk Assessment

07 Tianjin Huang (TU/e), The Roles of Adversarial Examples on Trustworthiness of Deep Learning

08 Lu Yin (TU/e), Knowledge Elicitation using Psychometric Learning

09 Xu Wang (VUA), Scientific Dataset Recommendation with Semantic Techniques

10 Dennis J.N.J. Soemers (UM), Learning State-Action Features for General Game Playing

11 Fawad Taj (VUA), Towards Motivating Machines: Computational Modeling of the Mechanism of Actions for Effective Digital Health Behavior Change Applications

12 Tessel Bogaard (VUA), Using Metadata to Understand Search Behavior in Digital Libraries

13 Injy Sarhan (UU), Open Information Extraction for Knowledge Representation

14 Selma Čaušević (TUD), Energy resilience through self-organization

15 Alvaro Henrique Chaim Correia (TU/e), Insights on Learning Tractable Probabilistic Graphical Models

16 Peter Blomsma (TiU), Building Embodied Conversational Agents: Observations on human nonverbal behaviour as a resource for the development of artificial characters

17 Meike Nauta (UT), Explainable AI and Interpretable Computer Vision – From Oversight to Insight

18 Gustavo Penha (TUD), Designing and Diagnosing Models for Conversational Search and Recommendation

19 George Aalbers (TiU), Digital Traces of the Mind: Using Smartphones to Capture Signals of Well-Being in Individuals

20 Arkadiy Dushatskiy (TUD), Expensive Optimization with Model-Based Evolutionary Algorithms applied to Medical Image Segmentation using Deep Learning

21 Gerrit Jan de Bruin (UL), Network Analysis Methods for Smart Inspection in the Transport Domain

22 Alireza Shojaifar (UU), Volitional Cybersecurity

23 Theo Theunissen (UU), Documentation in Continuous Software Development

24 Agathe Balayn (TUD), Practices Towards Hazardous Failure Diagnosis in Machine Learning

25 Jurian Baas (UU), Entity Resolution on Historical Knowledge Graphs

26 Loek Tonnaer (TU/e), Linearly Symmetry-Based Disentangled Representations and their Out-of-Distribution Behaviour

27 Ghada Sokar (TU/e), Learning Continually Under Changing Data Distributions

28 Floris den Hengst (VUA), Learning to Behave: Reinforcement Learning in Human Contexts

29 Tim Draws (TUD), Understanding Viewpoint Biases in Web Search Results

---

2024 01 Daphne Miedema (TU/e), On Learning SQL: Disentangling concepts in data systems education

02 Emile van Krieken (VUA), Optimisation in Neurosymbolic Learning Systems

03 Feri Wijayanto (RUN), Automated Model Selection for Rasch and Mediation Analysis

04 Mike Huisman (UL), Understanding Deep Meta-Learning

05 Yiyong Gou (UM), Aerial Robotic Operations: Multi-environment Cooperative Inspection & Construction Crack Autonomous Repair

06 Azqa Nadeem (TUD), Understanding Adversary Behavior via XAI: Leveraging Sequence Clustering to Extract Threat Intelligence

07 Parisa Shayan (TiU), Modeling User Behavior in Learning Management Systems

08 Xin Zhou (UvA), From Empowering to Motivating: Enhancing Policy Enforcement through Process Design and Incentive Implementation

09 Giso Dal (UT), Probabilistic Inference Using Partitioned Bayesian Networks

10 Cristina-Iulia Bucur (VUA), Linkflows: Towards Genuine Semantic Publishing in Science

11 withdrawn

12 Peide Zhu (TUD), Towards Robust Automatic Question Generation For Learning

13 Enrico Liscio (TUD), Context-Specific Value Inference via Hybrid Intelligence

14 Larissa Capobianco Shimomura (TU/e), On Graph Generating Dependencies and their Applications in Data Profiling

15 Ting Liu (VUA), A Gut Feeling: Biomedical Knowledge Graphs for Interrelating the Gut Microbiome and Mental Health

16 Arthur Barbosa Câmara (TUD), Designing Search-as-Learning Systems

17 Razieh Alidoosti (VUA), Ethics-aware Software Architecture Design

18 Laurens Stoop (UU), Data Driven Understanding of Energy-Meteorological Variability and its Impact on Energy System Operations

19 Azadeh Mozafari Mehr (TU/e), Multi-perspective Conformance Checking: Identifying and Understanding Patterns of Anomalous Behavior

20 Ritsart Anne Plantenga (UL), Omgang met Regels

21 Federica Vinella (UU), Crowdsourcing User-Centered Teams

22 Zeynep Ozturk Yurt (TU/e), Beyond Routine: Extending BPM for Knowledge-Intensive Processes with Controllable Dynamic Contexts

23 Jie Luo (VUA), Lamarck's Revenge: Inheritance of Learned Traits Improves Robot Evolution

24 Nirmal Roy (TUD), Exploring the effects of interactive interfaces on user search behaviour

25 Alisa Rieger (TUD), Striving for Responsible Opinion Formation in Web Search on Debated Topics

26 Tim Gubner (CWI), Adaptively Generating Heterogeneous Execution Strategies using the VOILA Framework

27 Lincen Yang (UL), Information-theoretic Partition-based Models for Interpretable Machine Learning

28 Leon Helwerda (UL), Grip on Software: Understanding development progress of Scrum sprints and backlogs

29 David Wilson Romero Guzman (VUA), The Good, the Efficient and the Inductive Biases: Exploring Efficiency in Deep Learning Through the Use of Inductive Biases

30 Vijanti Ramautar (UU), Model-Driven Sustainability Accounting

31 Ziyu Li (TUD), On the Utility of Metadata to Optimize Machine Learning Workflows

32 Vinicius Stein Dani (UU), The Alpha and Omega of Process Mining

33 Siddharth Mehrotra (TUD), Designing for Appropriate Trust in Human-AI interaction

34 Robert Deckers (VUA), From Smallest Software Particle to System Specification - MuDForM: Multi-Domain Formalization Method

35 Sicui Zhang (TU/e), Methods of Detecting Clinical Deviations with Process Mining: a fuzzy set approach

36 Thomas Mulder (TU/e), Optimization of Recursive Queries on Graphs

37 James Graham Nevin (UvA), The Ramifications of Data Handling for Computational Models

38 Christos Koutras (TUD), Tabular Schema Matching for Modern Settings

39 Paola Lara Machado (TU/e), The Nexus between Business Models and Operating Models: From Conceptual Understanding to Actionable Guidance

40 Montijn van de Ven (TU/e), Guiding the Definition of Key Performance Indicators for Business Models

41 Georgios Siachamis (TUD), Adaptivity for Streaming Dataflow Engines

42 Emmeke Veltmeijer (VUA), Small Groups, Big Insights: Understanding the Crowd through Expressive Subgroup Analysis

43 Cedric Waterschoot (KNAW Meertens Instituut), The Constructive Conundrum: Computational Approaches to Facilitate Constructive Commenting on Online News Platforms

44 Marcel Schmitz (OU), Towards learning analytics-supported learning design

45 Sara Salimzadeh (TUD), Living in the Age of AI: Understanding Contextual Factors that Shape Human-AI Decision-Making

46 Georgios Stathis (Leiden University), Preventing Disputes: Preventive Logic, Law & Technology

47 Daniel Daza (VUA), Exploiting Subgraphs and Attributes for Representation Learning on Knowledge Graphs

48 Ioannis Petros Samiotis (TUD), Crowd-Assisted Annotation of Classical Music Compositions

---

2025 01 Max van Haastrecht (UL), Transdisciplinary Perspectives on Validity: Bridging the Gap Between Design and Implementation for Technology-Enhanced Learning Systems

02 Jurgen van den Hoogen (JADS), Time Series Analysis Using Convolutional Neural Networks

03 Andra-Denis Ionescu (TUD), Feature Discovery for Data-Centric AI

04 Rianne Schouten (TU/e), Exceptional Model Mining for Hierarchical Data

05 Nele Albers (TUD), Psychology-Informed Reinforcement Learning for Situated Virtual Coaching in Smoking Cessation

06 Daniël Vos (TUD), Decision Tree Learning: Algorithms for Robust Prediction and Policy Optimization

07 Ricky Maulana Fajri (TU/e), Towards Safer Active Learning: Dealing with Unwanted Biases, Graph-Structured Data, Adversary, and Data Imbalance

08 Stefan Bloemheuvel (TiU), Spatio-Temporal Analysis Through Graphs: Predictive Modeling and Graph Construction

09 Fadime Kaya (VUA), Decentralized Governance Design - A Model-Based Approach

10 Zhao Yang (UL), Enhancing Autonomy and Efficiency in Goal-Conditioned Reinforcement Learning

11 Shahin Sharifi Noorian (TUD), From Recognition to Understanding: Enriching Visual Models Through Multi-Modal Semantic Integration

12 Lijun Lyu (TUD), Interpretability in Neural Information Retrieval

13 Fuda van Diggelen (VUA), Robots Need Some Education: on the complexity of learning in evolutionary robotics

14 Gennaro Gala (TU/e), Probabilistic Generative Modeling with Latent Variable Hierarchies

15 Michiel van der Meer (UL), Opinion Diversity through Hybrid Intelligence

16 Monika Grewal (TU Delft), Deep Learning for Landmark Detection, Segmentation, and Multi-Objective Deformable Registration in Medical Imaging

17 Matteo De Carlo (VUA), Real Robot Reproduction: Towards Evolving Robotic Ecosystems

18 Anouk Neerincx (UU), Robots That Care: How Social Robots Can Boost Children’s Mental Wellbeing

19 Fang Hou (UU), Trust in Software Ecosystems

20 Alexander Melchior (UU), Modelling for Policy is More Than Policy Modelling (The Useful Application of Agent-Based Modelling in Complex Policy Processes)

21 Mandani Ntekouli (UM), Bridging Individual and Group Perspectives in Psychopathology: Computational Modeling Approaches using Ecological Momentary Assessment Data

22 Hilde Weerts (TU/e), Decoding Algorithmic Fairness: Towards Interdisciplinary Understanding of Fairness and Discrimination in Algorithmic Decision-Making

23 Roderick van der Weerdt (VUA), IoT Measurement Knowledge Graphs: Constructing, Working and Learning with IoT Measurement Data as a Knowledge Graph

24 Zhong Li (UL), Trustworthy Anomaly Detection for Smart Manufacturing

25 Kyana van Eijndhoven (TiU), A Breakdown of Breakdowns: Multi-Level Team Coordination Dynamics under Stressful Conditions

26 Tom Pepels (UM), Monte-Carlo Tree Search is Work in Progress

27 Danil Provodin (JADS, TU/e), Sequential Decision Making Under Complex Feedback

28 Jinke He (TU Delft), Exploring Learned Abstract Models for Efficient Planning and Learning

29 Erik van Haeringen (VUA), Mixed Feelings: Simulating Emotion Contagion in Groups

30 Myrthe Reuver (VUA), A Puzzle of Perspectives: Interdisciplinary Language Technology for Responsible News Recommendation

31 Gebrekirstos Gebreselassie Gebremeskel (RUN), Spotlight on Recommender Systems: Contributions to Selected Components in the Recommendation Pipeline

32 Ryan Brate (UU), Words Matter: A Computational Toolkit for Charged Terms

33 Merle Reimann (VUA), Speaking the Same Language: Spoken Capability Communication in Human-Agent and Human-Robot Interaction

34 Eduard C. Groen (UU), Crowd-Based Requirements Engineering

35 Urja Khurana (VUA), From Concept To Impact: Toward More Robust Language Model Deployment

36 Anna Maria Wegmann (UU), Say the Same but Differently: Computational Approaches to Stylistic Variation and Paraphrasing

37 Chris Kamphuis (RUN), Exploring Relations and Graphs for Information Retrieval

38 Valentina Maccatrozzo (VUA), Break the Bubble: Semantic Patterns for Serendipity

39 Dimitrios Alivanistos (VUA), Knowledge Graphs & Transformers for Hypothesis Generation: Accelerating Scientific Discovery in the Era of Artificial Intelligence

40 Stefan Grafberger (UvA), Declarative Machine Learning Pipeline Management via Logical Query Plans

41 Mozhgan Vazifehdoostirani (TU/e), Leveraging Process Flexibility to Improve Process Outcome - From Descriptive Analytics to Actionable Insights

42 Margherita Martorana (VUA), Semantic Interpretation of Dataless Tables: a metadata-driven approach for findable, accessible, interoperable and reusable restricted access data

43 Krist Shingjergji (OU), Sense the Classroom - Using AI to Detect and Respond to Learning-Centered Affective States in Online Education

44 Robbert Reijnen (TU/e), Dynamic Algorithm Configuration for Machine Scheduling Using Deep Reinforcement Learning

45 Anjana Mohandas Sheeladevi (VUA), Occupant-Centric Energy Management: Balancing Privacy, Well-being and Sustainability in Smart Buildings

46 Ya Song (TU/e), Graph Neural Networks for Modeling Temporal and Spatial Dimensions in Industrial Decision-making

47 Tom Kouwenhoven (UL), Collaborative Meaning-Making. The Emergence of Novel Languages in Humans, Machines, and Human-Machine Interactions

48 Evy van Weelden (TiU), Integrating Virtual Reality and Neurophysiology in Flight Training

49 Selene Báez Santamaría (VUA), Knowledge-centered conversational agents with a drive to learn

50 Lea Krause (VUA), Contextualising Conversational AI

51 Jiaxu Zhao (TU/e), Understanding and Mitigating Unwanted Biases in Generative Language Models

52 Qiao Xiao (TU/e), Model, Data and Communication Sparsity for Efficient Training of Neural Networks

53 Gaole He (TUD), Towards Effective Human-AI Collaboration: Promoting Appropriate Reliance on AI Systems

54 Go Sugimoto (VUA), MISSING LINKS Investigating the Quality of Linked Data and its Tools in Cultural Heritage and Digital Humanities

55 Sietze Kai Kuilman (TUD), AI that Glitters is Not Gold: Requirements for Meaningful Control of AI Systems

56 Wijnand van Woerkom (UU), A Fortiori Case-Based Reasoning: Formal Studies with Applications in Artificial Intelligence and Law

57 Syeda Amna Sohail (UT), Privacy-Utility Trade-Off in Healthcare Metadata Sharing and Beyond: A Normative and Empirical Evaluation at Inter and Intra Organizational Levels

58 Junhan Wen (TUD), "From iMage to Market": Machine-Learning-Empowered Fruit Supply

59 Mohsen Abbaspour Onari (TU/e), From Explanation to Trust: Modeling and Measuring Trust in Explainable Decision Support

60 Marcel Jurriaan Robeer (UU), Beyond Trust: A Causal Approach to Explainable AI in Law Enforcement

61 Shuai Wang (VUA), Links in Large Integrated Knowledge Graphs: Analysis, Refinement, and Domain Applications

62 Khaleel Asyraaf Mat Sanusi (OU), Augmenting a learning model within immersive learning environments for psychomotor skills

63 Rashid Zaman (TU/e), Online Conformance Checking on Degraded Data

64 Jens d'Hondt (TU/e), Effective and Efficient Multivariate Similarity Search

65 Aswin Balasubramaniam (UT), Disentangling Runner Drone Interaction Potentialities

---

2026 01 Pei-Yu Chen (TUD), Human-Agent Alignment Dialogues: Eliciting User Information at Runtime for Personalized Behavior Support

02 Hezha Hassan Mohammedkhan (TiU), Estimating Body Measurements of Children from 2D Images: Towards the Automatic Detection of Malnutrition

03 Kyriakos Psarakis (TUD), Democratizing Scalable Cloud Applications: Transactional Stateful Functions on Streaming Dataflows

04 Boyu Xu (UU), Exploring Indirect Relations Between Topics in Neuroscience Literature Using Augmented Reality to Inform Experimental Design

05 Koen Minartz (TU/e), Stochastic Simulation with Geometric Deep Generative Models

06 Azim Afroozeh (CWI, VUA), FastLanes: A Next-Gen File Format

07 Inès Blin (VUA), Narrative Understanding with Knowledge Graphs

08 Paul van Vulpen (UU), Debating Digital Dominance: Decentralized Technology Governance For Strategic Autonomy

09 Afrizal Doewes (TU/e), Rethinking Automated Essay Scoring: Agreement, Fairness, and Feedback

10 Nikolaos Delapaschos Kondylidis (VUA), Establishing Task-Oriented Understanding between Agents

11 Işıl Baysal Erez (UT), Handling Missing Data with Meta-Learning and Large Language Models

12 Xue Li (UvA), From Fine-tuning to Prompting: A Paradigm Shift in Knowledge Graph Construction

13 Isaac da Silva Torres (VUA), Guidelines To Flux Between Conceptual Models: Understanding Complex Digital Business Ecosystems

14 Philip Lippmann (TUD), Synthetic Data for Robust Language Modelling

15 Rashmi Khazanchi (OU), Artificial Intelligence in Education: Impact of AI-Based Systems on Mathematics Achievement

16 Carolina Ferreira Gomes Centeio Jorge (TUD), Modelling Artificial Trust for Effective Human-AI Teamwork

17 Maria Tsfasman (TUD), Towards Predicting Memory in Multimodal Group Interactions