Towards the ISO 24617-2-compliant Typology of Metacognitive Events

Volha Petukhova and **Hafiza Erum Manzoor**
Spoken Language Systems Group, Saarland Informatics Campus
Saarland University, Saarbrücken, Germany
{v.petukhova,hemanzoor}@lsv.uni-saaarland.de

Abstract

The paper presents ongoing efforts in design of a typology of metacognitive events observed in a multimodal dialogue. The typology will serve as a tool to identify relations between participants' dispositions, dialogue actions and metacognitive indicators. It will be used to support an assessment of metacognitive knowledge, experiences and strategies of dialogue participants. Based on the multidimensional dialogue model defined within the framework of Dynamic Interpretation Theory and ISO 24617-2 annotation standard, the proposed approach provides a systematic analysis of metacognitive events in terms of dialogue acts, i.e. concepts that dialogue research community is used to operate on in dialogue modelling and system design tasks.

1 Introduction

Daily life is replete with determinations about the reliability of our own thoughts and feelings as well as attributions about the thoughts and feelings of others. These metacognitive capacities underlie cognitive and social adaptation, influence decision-making, can enhance self-efficacy. Metacognition enables cognitive control needed for people to anticipate the future task demands, improves learning and performance on complex memory tasks, knowledge transfer and task switching (Taatgen, 2013). Cognitive models of metacognitive processes, when integrated into human-computer dialogue, transform the dialogue system from a reactive dialogue participant into a proactive learner, accomplished multi-tasking planner and adaptive decision maker (Malchanau et al., 2018).

Metacognitive capabilities of existing interactive systems, even of complex smart learning environments (Spector, 2014), are still rather limited so as metacognitive strategies used. To exploit the full potential of efficient metacognitive support in a dialogue system, big multimodal data samples are required to reliably identify metacognitive states accounting for a complexity of multidimensional contingencies between tasks, performed actions and participants' cognitive and emotional dispositions. An elaborate computational model of (meta)cognitive states calls, in the first place, for a typology of *metacognitive events* – reflexive activities that express the sender's mindful awareness of own and others cognitive processes, e.g. checking out and verification of attention, recognition, understanding, evaluation and regulation of content, thought processes, attitudes, preferences, assumptions and emotions. Metacognitive events should be computable/learned from a range of low level multi-sensory and psycho-physiological indicators (markers). Methods are required to transform multimodal data in a meaningful way to enable appropriate measurements of metacognition, adaptive decision-making and efficient coordination of multiple dialogue tasks. The main goal of the presented study is to provide a theoretical framework, methodological insights and experimental design to model relevant metacognitive processes, and specify a set of recognizable and measurable indicators to assess metacognition in dialogue.

The paper is structured as follows. Section 2 reviews methods to assess metacognition in interactive setting. In Section 3, we specify the model of metacognitive processes within the framework of Dynamic Interpretation Theory (DIT). We adapt the established metacognition assessment instruments in order to discover potential correlations between dialogue acts and metacognitive events. Section 4 presents experimental design featuring data collection, processing and ISO 24617-2 compliant annotation protocols. We wrap up the paper by outlining expected project outcomes.

2 Metacognition Assessment Instruments

Assessment of metacognition traditionally involves **self-reported** measurements. The most widely used Metacognition Questionnaire (MCQ, Cartwright-Hatton and Wells (1997)) evaluates fac-

tors related to positive and negative metacognitive beliefs, metacognitive monitoring and judgements of cognitive confidence. Questionnaires are however of limited value since they are subjective and not always accurate (Schraw, 2009).

There are two online methods proposed to assess metacognition: *thinking aloud* and *reflection when prompting*. Participants speak about their own cognitive states or processes and their understanding of partner's states and processes, or are prompted to reflect on the reasons why they chose specific actions – **verbalized metacognition**. The methods enable assessment of three elements of metacognition - experiences (e.g. confidence, confusion), knowledge (e.g. gaps), and strategies (e.g. actions). Think-aloud and prompting protocols provide rich information about the metacognitive processes when performing a task and are powerful predictors of test performance (Bannert and Mengelkamp, 2008). Verbalization methods are proven valid, but time consuming. Moreover, elicitation of explicit monitoring, reflection and regulation moments may disrupt or even break down the interaction process, distort its naturalness, trigger attention theft, increase cognitive load and impact negatively participants' engagement.

There is research performed on the **psychophysiological** measurement of metacognition. Physiological measures make use of EEG electroencephalography (Wokke et al., 2020), heart rate (Meessen et al., 2018), and pupil dilation (Lempert et al., 2015), but require rather complex and often expensive hard- and software set ups. Other methods exploit information about interlocutor's behaviour via **log files** and efficiently combine it with questionnaire data (Linek et al., 2008).

Recently, increasing computational power and technological advances opened up new data-driven assessment scenarios. A huge diversity of inexpensive tracking and sensing devices enable rather exhaustive **real-time monitoring** and immediate assessment of affective cognitive states, including metacognitive aspects (Gašević et al., 2015). Significant progress has been booked in automatic affective cognitive state recognition from speech and visual signals (Kapoor and Picard, 2005; DMello et al., 2008). Large amounts of multimodal data is used to train deep learning algorithms to recognize facial expressions related to emotions and cognitive states in large variety of scenarios.

The definition and detection of metacognitive multimodal indicators requires transforming the raw multi-sensory data collected in a meaningful way so that it allows taking decisions, provide indicators of interlocutor's performance, efficiency and preferences (Greene and Azevedo, 2010). This has been done for interaction logs, the records of sequential actions users performed in an interface. Such actions are interpreted as any communicative action, i.e. having certain communicative functions. A set of dialogue acts has been proposed for screen events by translating the human-human communication mechanisms into human-computer interactions as functions of GUI (van Dam, 2006).

Coherence and interaction analysis is applied to analyse think-aloud interviews and prompting interaction transcripts (Ericsson and Simon, 1984); modern natural language processing techniques are used (Bosch et al., 2021). In multimodal interactions that involve speech, taking notes, nonverbal communication and graphical user interface actions, metacognitive strategies are observable via interaction logs, metacognitive experiences - via recorded and tracked behaviour, and metacognitive knowledge - via speech and typed transcripts. The interaction-based approach to measure metacognition that we propose will enable real-time and non-intrusive assessment of all metacognitive aspects – experiences, knowledge and strategies.

3 Modelling Metacognitive Processes in Dialogue Interaction

Metacognitive regulation refers to adjustments individuals make to their processes to help control their task performance, learning and interaction. Metacognitive processes underlie *awareness*, *monitoring*, *reflection* and *regulation* activities (Brown, 1987). Metacognition has *implicit* and *explicit* forms,[1] and is applied to *own* (sender's) and *others* (addressee's) cognitive processes. In human dialogue, metacognitive processes concern reasoning about interlocutors' intentions and knowledge, and are often modelled as parts of *shared* or *mutual* beliefs forming a *common ground* (Traum, 1994; Bunt, 2000). Common ground is not directly accessible. An access to self and others cognitive processes through questionnaires and think-aloud protocols is very limited; reports on own and others' intentions can be inaccurate. (Meta)cognitive processes underlying establishing and updating common ground (grounding), on the other hand, may

[1] Explicit metacognition is considered a uniquely human ability (Frith, 2012).

become accessible through or inferred from observable dialogue behaviour. For instance, gaze (re-)direction deliver information about the interlocutor attention by means of frequency and duration of gaze fixation on the Areas of Interest (AoI), but also provides an evidence about the positive versus negative emotional reaction on the fixated object. In face-to-face conversation, participants may present evidence of grounding through verbal and vocal signals, body movements and facial expressions; in interaction with graphical user interfaces, typing behaviour, mouse movements and clicks may signal changes in (meta)cognitive and motivational functioning. A metacognitive event is characterised through evidence of reflexive activities indicating any level of sender's mindful awareness about own (sender's) and others (partner's) cognitive process(-es):

- **Level 0:** ignore or offer false continuation;
- **Level 1:** pay and secure attention (mutual eye contact);
- **Level 2:** recognise, record change and respond with minimal signals (gaze (re-)direction, head nods, 'mmhmm', 'uhu'), check out and verify recognition;
- **Level 3:** interpret, check out and verify understanding, and respond to content and feeling ('I see what your mean...', 'I am confused...');
- **Level 4:** evaluate content and feeling, inspect/compare past experiences and verify hypotheses ('I am as worried as you are...');
- **Level 5:** regulate and align, correct/adjust, imitate, anticipate consequences, plan the ongoing procedure (content, sequences, timing,...).

At all these levels, positive and negative beliefs concern sender's awareness about: (i) his/her own thoughts (zero-order theory of mind abilities, Premack and Woodruff (1978)), (ii) about another person's thoughts (first-order theory of mind), and (iii) what another person thinks about sender's thoughts (second-order theory of mind). Consider the following example[2]:

(1) du1. A: The next train is at 11:02.
du2. B: At 11:02.
du3. A: That's correct.
du4. B: Okay thanks

In 1, A in order to continue the dialogue should know that B understands his utterance $du1$ and believes its content p. B's utterance $du2$ can be considered as such evidence where B is verifying its recognition or even on a higher level – its understanding. So after $du2$, A believes that B believes that p, and that B believes that A believes that p. However, A cannot be certain that B indeed believes that p, since in $du2$ he also seems to offer that belief for confirmation. A's response $du3$ gives that confirmation. At this point A does not yet know whether his utterance has reached B and was well understood. B's next contribution $du4$ provides evidence for that; upon understanding $du4$, A has accumulated the following beliefs:

(2) A believes that p
A believes that B believes that p
A believes that B believes that A believes that p
A believes that B believes that A believes that B believes that p
A believes that B believes that A believes that B believes that A believes that p

or represented as *mutual beliefs* equal to:

(3) A believes that it is mutually believed that p

To classify and model implicit and explicit metacognitive events (acts), the framework of the Dynamic Interpretation Theory (DIT, Bunt (1999)) and the ISO 24617-2 dialogue act annotation standard (ISO, 2012) will be used. DIT has emerged from the study of multimodal human-human dialogues uncovering fundamental principles observed in such interactions. DIT and its subset ISO 24617-2 are open multidimensional dialogue act taxonomies[3]. They are proven to provide theoretically grounded and empirically tested inventory of dialogue acts with fine-grained semantic distinctions presenting the semantic framework for the systematic analysis and computational modelling of multimodal dialogue behaviour in many interactive settings.

Special attention will be paid to *feedback* acts which we assume are crucial for successful recognition of metacognitive events: positive and negative feedback about speaker's own (*auto*-feedback) and the partner's processing (*allo*-feedback) at the five processing levels: attention, perception, interpretation, evaluation and execution (Bunt, 2000). Speaker's *repairs, (self-)corrections, partner completions* and *hesitations* (silent and filled pauses) are assumed to strongly correlate with moments of reflection and may reveal speaker's cognitive confidence. *Managing* allocation of *time*, *turn*, *struc-*

[2] Adapted from (Bunt et al., 2007).

[3] DIT, Release 5.2 and ISO 24617-2, Second Edition are available on https://dit.uvt.nl/

Metacognitive Activity	MCQ dimension	Dialogue Act			Indicators (example)
		Dimension	Function	Qualifier	
Awareness	cognitive (self-)conciseness	Auto-/Allo-Feedback	pos. attention pos. perception neg. attention neg. perception	responsiveness (dis)engagement	nonverbal: gaze, head orientation verbal: backchannels nonverbal: gaze aversion GUI: no activity
		Contact Man.	check indication		vocal: throat clearing nonverbal: leaning forward
Monitoring	cognitive confidence	Auto-/Allo-Feedback	pos./neg. interpretation	interest confusion	nonverbal: eye contact nonverbal: puzzled look
		Time Management	stalling	(un)certainty	verbal: filled pauses speech/GUI: slowing down
		Own Communication Management	retraction		verbal/speech: editing expressions GUI: back to initial position speech: disfluencies all: false/re- starts
Reflection	pos./neg. evaluation beliefs	Auto-/Allo-Feedback	pos./neg. evaluation elicitation	empathy worry respect surprise appraisal	nonverbal: thinking face, gaze up verbal: check out understanding verbal: paraphrases, summarization nonverbal: longer gesture strokes verbal: chunking/sorting content nonverbal: raise eyebrow, jerk verbal: make sense, right
Regulation	cognitive need for control	Auto-/Allo-Feedback	pos./neg. execution	irritation cooperation	nonverbal: thinking face, gaze up all: entrainment/alignment
		Own Communication Management	self-correction	frustration excitement	verbal/speech: replacement GUI: cancel
		Partner Communication Management	correct misspeaking completion		verbal: replacement verbal: completion hypothesis
		Discourse Structuring	topic shift		verbal: introduce another topic
		Turn Management	take, keep, release, grab		verbal: start, keep, stop speaking

Table 1: Tentative mapping between metacognitive actions, associated MCQ dimensions and DIT/ISO24617-2 dialogue acts illustrated with examples of possible multimodal metacognitive indicators.[4]

turing discourse and *control* over issues under *discussion* concern with planning aspects. Analysing socio-emotional aspects will enable modelling of metacognitive activities related to positive, negative thoughts, feelings of uncontrollability and danger, and engagement related emotions such as boredom, enjoyment and frustration. Table 1 provides a preliminary view on associations/correlations between metacognitive activities, MCQ dimensions and DIT/ISO dialogue acts illustrated with multimodal behaviour examples. The typology will be experimentally tested and extended as described in the next Section.

4 Experimental Design

Use case The importance of metacognition has been empirically proven for negotiations (Galluccio and Safran, 2015). High self- and others- monitors are more concerned that their negotiations go well, flexibly modify their actions to better adapt to the changing dynamics of the situation, typically by using other people's behaviour as a guide to their own. High self-monitors and -assessors are more likely to engage in argumentation and are better able to accomplish their goals.

As the use case, we will focus on patient-physician negotiations for shared decisions. Medical students and professionals tend to overestimate the value of medical knowledge and are known as poor self-monitors and self-assessors (Eichbaum, 2014). Therapy planning scenarios of varied complexity will be defined reflecting different participant's dispositions. Interaction concerns multi-issue bargaining where each issue involves multiple negotiation *options* with preferences representing parties negotiation positions. Preferences are weighted in order of importance (strength) and defined as the participant's beliefs about *attitudes* towards certain behaviour and *abilities* to perform this behaviour. The goal of each partner is to find out preferences of each other and to search for the best possible mutual agreement. The human participant - doctor - negotiates either with a human or artificial patient who will have different preferences and instructed (programmed) to apply several negotiation and decision-making strategies (Petukhova et al., 2019).[4]

Data Collection will be performed via a) human-human role-playing (small-scaled) and b) human-agent interactive simulations (large collections). Role-playing method is often used to collect

[4]The list of multimodal indicators is not complete, for more examples see (Petukhova, 2005).

interactive data in a controlled setting and underpins simulations of many real-life communicative situations (Brône and Oben, 2015). Here, one participant will be randomly assigned the role of a doctor, the other participant - a patient. Each participant will receive instructions and preference profile, and asked to negotiate a mutual agreement with the highest possible value. Procedures will be specified for the settings where both participants: (1) observe others' actions and flag problems or gaps; (2) verbalise their cognitive processes and their understanding of partner's states and processes; (3) explain his/her choices; and (4) are involved in free flow negotiation. The former three settings will be used as reference for the analysis of metacognition in the unconstrained close to authentic interactions.

Simulations of communicative situations with human and artificial Simulated Patients (SPs) will be arranged. Regular medical communication practice often takes place in a patient-simulated setting, where Simulated Patients (SPs) are involved to portray a particular set of symptoms or roles (Kaplonyi et al., 2017). Simulations with humans provide high fidelity training, but are costly, difficult to reproduce and access. AI agents as SPs can be used to create specific situations in which physicians metacognitive processes can be activated and assessed (Petukhova et al., 2019). Moreover, simulations will impose certain restrictions in order to investigate a controlled set of communicative (metacognitive) activities and related phenomena without having to deal with unrelated details. Multimodal data will be recorded. The quality of recordings will be adapted to the application conditions, i.e. a fairly good but not perfect acoustic and visual quality will be targeted. Prior to recordings, participants will complete the short MCQ-30 questionnaire. We will account for gender and role differences.

Data Recording and Processing Participants' speech will be transcribed by running the Kaldi-based[5] Automatic Speech Recognizer re-trained on the medical in-domain data and correcting the output manually. Since substantial deviations in patient and physician vocabularies are assumed, language models will be adapted to both groups. OpenSMILE tool[6] will be used to extract spectral, prosodic and voice quality features. OpenFace tool[7] will be used to extract 2D/3D facial landmark points for eyes, eyebrows, nose, mouth, jawline and head, and to compute 18 Facial Action Units (AUs). OpenFace enables real-time online/off-line feature extraction from a webcam input and videos, thus no expensive sensors and tracking devices are required. To record GUI interactions, a graphical utility `Atbswp` in Python3 will be used to record the mouse and keyboard actions.

Multi-sensory data will be synchronised and stored in the standard `tei` format, and exported to ELAN[8] to perform the ISO 24617-2 compliant annotations.

5 Expected Outcomes

The proposed project will contribute to a better understanding of metacognitive processes underlying dialogue participants decision-making and interactive performance progressing towards a computational cognitive model of social metacognition. An interaction-based method for metacognition assessment will be worked out providing an ISO-compliant typology of metacognitive events, a set of multimodal feature extraction and classification models as well as new tools for multidimensional dialogue analysis. Finally, substantial amount of multimodal data annotated with the ISO 24617-2 dialogue acts will be provided to the research community via DialogBank release. [9]

References

Maria Bannert and Christoph Mengelkamp. 2008. Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. does the verbalisation method affect learning? *Metacognition and Learning*, 3(1):39–58.

Nigel Bosch, Yingbin Zhangm, Luc Paquette, Ryan Baker, Jaclyn Ocumpaugh, and Gautam Biswas. 2021. Students verbalized metacognition during computerized learning. In *ACM SIGCHI: Computer-Human Interaction*.

Geert Brône and Bert Oben. 2015. Insight interaction: a multimodal and multifocal dialogue corpus. *Language resources and evaluation*, 49(1):195–214.

Ann Brown. 1987. Metacognition, executive control, self-regulation, and other more mysterious mechanisms. *Metacognition, motivation, and understanding*.

Harry Bunt. 1999. Dynamic interpretation and dialogue theory. *The structure of multimodal dialogue*, 2:139–166.

[5] https://kaldi-asr.org/

[6] https://www.audeering.com/opensmile/

[7] https://github.com/TadasBaltrusaitis/OpenFace

[8] https://archive.mpi.nl/tla/elan

[9] https://dialogbank.uvt.nl/

Harry Bunt. 2000. Dialogue pragmatics and context specification. *Abduction, Belief and Context in Dialogue. Studies in Computational Pragmatics. Amsterdam: Benjamins*, pages 81–150.

Harry Bunt, Roser Morante, and Simon Keizer. 2007. An empirically based computational model of grounding in dialogue. In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, pages 283–290.

Sam Cartwright-Hatton and Adrian Wells. 1997. Beliefs about worry and intrusions: The metacognitions questionnaire and its correlates. *Journal of anxiety disorders*, 11(3):279–296.

Hans van Dam. 2006. *Dialogue acts in GUIs*. Ph.D. thesis, Technische Universiteit Eindhoven, Department of Industrial Design.

Sidney DMello, Tanner Jackson, Scotty Craig, Brent Morgan, P Chipman, Holly White, Natalie Person, Barry Kort, R El Kaliouby, Rosalind Picard, et al. 2008. Autotutor detects and responds to learners affective and cognitive states. In *Workshop on emotional and cognitive issues at the international conference on intelligent tutoring systems*, pages 306–308.

Quentin G Eichbaum. 2014. Thinking about thinking and emotion: the metacognitive approach to the medical humanities that integrates the humanities with the basic and clinical sciences. *The Permanente Journal*, 18(4):64.

K Anders Ericsson and Herbert A Simon. 1984. *Protocol analysis: Verbal reports as data.* the MIT Press.

Chris D Frith. 2012. The role of metacognition in human social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1599):2213–2223.

Mauro Galluccio and Jeremy D Safran. 2015. Mindfulness-based training for negotiators: Fostering resilience in the face of stress. In *Handbook of International Negotiation*, pages 209–226. Springer.

Dragan Gašević, Shane Dawson, and George Siemens. 2015. Lets not forget: Learning analytics are about learning. *TechTrends*, 59(1):64–71.

Jeffrey A Greene and Roger Azevedo. 2010. The measurement of learners self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational psychologist*, 45(4):203–209.

ISO. 2012. *Language resource management – Semantic annotation framework – Part 2: Dialogue acts. ISO 24617-2*. ISO Central Secretariat, Geneva.

Jessica Kaplonyi, Kelly-Ann Bowles, Debra Nestel, Debra Kiegaldie, Stephen Maloney, Terry Haines, and Cylie Williams. 2017. Understanding the impact of simulated patients on health care learners communication skills: a systematic review. *Medical education*, 51(12):1209–1219.

Ashish Kapoor and Rosalind W Picard. 2005. Multimodal affect recognition in learning environments. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 677–682.

Karolina M Lempert, Yu Lin Chen, and Stephen M Fleming. 2015. Relating pupil dilation and metacognitive confidence during auditory decision-making. *PLoS One*, 10(5):e0126588.

Stephanie B Linek, Birgit Marte, and Dietrich Albert. 2008. The differential use and effective combination of questionnaires and logfiles. In *Computer-based Knowledge & Skill Assessment and Feedback in Learning settings (CAF), Proceedings of the International Conference on Interactive Computer Aided Learning (ICL), 24th to 26th September*.

Andrei Malchanau, Volha Petukhova, and Harry Bunt. 2018. Towards integration of cognitive models in dialogue management: designing the virtual negotiation coach application. *Dialogue & Discourse*, 9(2):35–79.

Judith Meessen, Stefan Sütterlin, Siegfried Gauggel, and Thomas Forkmann. 2018. Learning by heartthe relationship between resting vagal tone and metacognitive judgments: a pilot study. *Cognitive processing*, 19(4):557–561.

Volha Petukhova. 2005. *Multidimensional interaction of multimodal dialogue acts in meetings*. Ph.D. thesis, MA thesis, Tilburg University.

Volha Petukhova, Firuza Sharifullaeva, and Dietrich Klakow. 2019. Modelling shared decision making in medical negotiations: Interactive training with cognitive agents. In *International Conference on Principles and Practice of Multi-Agent Systems*, pages 251–270. Springer.

David Premack and Guy Woodruff. 1978. Does the chimpanzee have a theory of mind? *Behavioral and Brain sciences*, 1(04):515–526.

Gregory Schraw. 2009. A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and learning*, 4(1):33–45.

Jonathan Michael Spector. 2014. Conceptualizing the emerging field of smart learning environments. *Smart learning environments*, 1(1):1–10.

Niels A Taatgen. 2013. The nature and transfer of cognitive skills. *Psychological review*, 120(3):439.

David R Traum. 1994. A computational theory of grounding in natural language conversation. Technical report, Rochester Univ NY Dept of Computer Science.

Martijn E Wokke, Dalila Achoui, and Axel Cleeremans. 2020. Action information contributes to metacognitive decision-making. *Scientific reports*, 10(1):1–15.