What Is Going through Your Mind?
Metacognitive Events Classification in Human-Agent Interactions

Hafiza Erum Manzoor, Volha Petukhova

Spoken Language Systems Group, Saarland Informatics Campus
Saarland University, Saarbrücken, Germany
{hemanzoor,v.petukhova}@lsv.uni-saarland.de

Abstract

For an agent, either human or artificial, to show intelligent interactive behaviour implies assessments of the reliability of own and others' thoughts, feelings and beliefs. Agents capable of these robust evaluations are able to adequately interpret their own and others' cognitive and emotional processes, anticipate future actions, and improve their decision-making and interactive performances across domains and contexts. Reliable instruments to assess interlocutors' mindful capacities for monitoring and regulation - metacognition - in human-agent interaction in real-time and continuously are of crucial importance however challenging to design. The presented study reports Concurrent Think Aloud (CTA) experiments in order to access and evaluate metacognitive dispositions and attitudes of participants in human-agent interactions. A typology of metacognitive events related to the 'verbalized' monitoring, interpretation, reflection and regulation activities observed in a multimodal dialogue has been designed, and serves as a valid tool to identify relation between participants' behaviour analysed in terms of ISO 24617-2 compliant dialogue acts and the corresponding metacognitive indicators.

Keywords: metacognitive events, dialogue acts, metacognition assessment in human-agent interaction

1. Introduction

Human interactions are regulated by the participants' abilities to attribute mental and emotional states to self and others. These metacognitive abilities allude to higher order thinking that involves active control over the cognitive processes (Livingston, 2003). Metacognition helps people identify gaps in their knowledge and create strategies to fill those gaps (Dunning, 2011). Metacognition governs decision-making processes (Yeung and Summerfield, 2012) and plays an important role in guiding and regulating human intelligent behaviour and social functioning (Frith, 2012).

With tremendous increase in computational power and significant advances in sensing technologies, many multimodal dialogue systems claim to offer the mode of interaction that is more intuitive and natural for their users. Behaviour of such systems, however, most of the time presents only a rough approximation of what would be considered as intelligent. The incorporation of metacognitive processes into the dialogue model has a potential to make dialogue systems genuinely more intelligent: enable proactive cognitive control, anticipate future task demands and actions, improve knowledge transfer and task switching, enhance interactivity, and enable social and cognitive adaptation in behaviour and decision making (Malchanau et al., 2018).

In order to bring metacognition into the dialogue system design and to exploit the full potential of efficient regulation and control strategies, it is crucial to have appropriate real-time continuous measurement of metacognition. An access to own and others' cognitive processes through offline prospective or retrospective *self-reports* is not always accurate (Schraw, 2009). Introspective online methods include *verbalization* and

reflection when prompting, and provide rich information about the metacognitive processes when performing a task. They are powerful predictors of task performance (Bannert and Mengelkamp, 2008), but also disclose current sensations, emotions, focus of attention, plans, intentions (Ericsson and Simon, 1980). However, elicitation of explicit monitoring, reflection and regulation moments may disrupt or even break down the interaction process and cannot be used as real-time continuous assessment tool. *Observational* approaches, where authentic multimodal interaction and social processes are recorded, have advantages over self-report and think-aloud methods (Whitebread et al., 2009). Observations may be more ecologically valid than the other methods, because they are independent of interlocutor's verbal ability and working memory capacity. Nevertheless, metacognition is not directly accessible through observations. Multimodal data has to be transformed in a meaningful way to understand the relationships between components of metacognition. Thus, a multi-method approach is required. This study investigates what verbalised data adds to the understanding of metacognitive processes and corresponding behaviours enabling a real-time continuous assessment of metacognition by the system. For this, a series of human-agent interactions experiments have been performed that involved concurrent probing and thinking aloud. The collected multimodal data was annotated with dialogue act information as having a certain communicative function and semantic content with explicitly defined metacognitive components. The semantic framework of Dynamic Interpretation Theory (DIT, (Bunt, 1999)) and the ISO 24617-2 dialogue act annotation standard (2. Edition, (ISO, 2020)) was applied
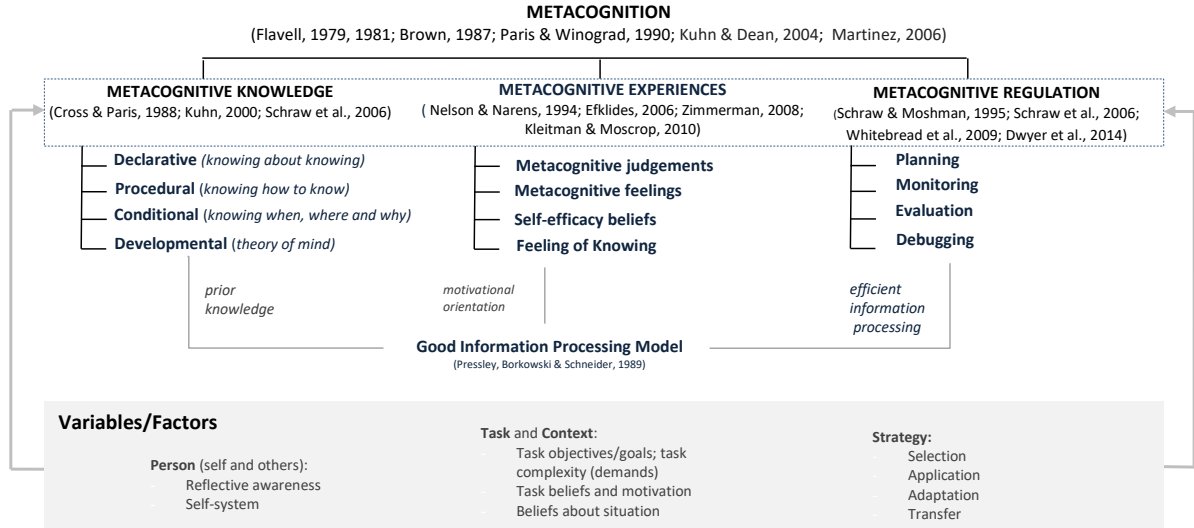
Figure 1: Conceptual overview of metacognition.

and extended to model the metacognitive events. We identified a number of associated metacognitive indicators that we will use in the future task of metacognitive states classification.

The paper is organized as follows. Section 2 discusses the key concepts of metacognition and presents the DIT/ISO 24617-2 based taxonomy of metacognitive acts by specifying their functions and contents. Section 3 outlines the experimental design. We discuss the think-aloud protocols for human-agent negotiation scenarios. The multimodal data collection, processing, dialogue act annotations and the obtained results are presented. Section 5 summarises important findings and limitations, draws conclusions and outlines future research.

2. Metacognition: Conceptual Overview and Events

There are various definitions and models proposed featuring fuzziness, expansiveness and complexity of the construct of metacognition (Tarricone, 2011). The general concept of metacognition as cognition about cognition (the 2nd order cognition: thoughts about thoughts, knowledge about knowledge) has been gradually broadened to include anything psychological, rather than just anything cognitive: knowledge of one's knowledge, processes, cognitive and affective states, the ability to consciously and deliberately monitor and regulate one's knowledge, processes, cognitive and affective states (Flavell, 1979; Flavell, 1981; Efklides, 2006; Zimmerman, 2008). Metacognition is at best described as a multidimensional construct comprising metacognitive knowledge, metacognitive experiences and the respective regulation. *Metacognitive knowledge* consists of beliefs of one's cognitive abilities, of particular tasks and of different strategies that are available and appropriate to the task. *Metacognitive experiences* are concerned with self-efficacy beliefs,

confidence, feelings and accuracy as the degree of correspondence between the subjective judgement and the actual performance. *Metacognitive regulation* involves conscious control and strategies such as planning, progress monitoring, effort allocation, strategy use and regulation of cognition, i.e. the ability to recognize and reflect on one's own and others mental states, as well as the ability to use metacognitive knowledge to tackle the difficulties in social interactions (Petty et al., 2007; Brune et al., 2011). Three major factors or variables interact to affect the course and outcome of cognitive enterprises: *person*, *task* and *strategy*. Discrete acts of metacognition, i.e. cognitive acts of monitoring and the respective behavioural regulation are complemented by synthetic forms of metacognition in which an array of intentions, thoughts, feelings, and connections between events are integrated into larger complex representations of self and others developed over time (Brune et al., 2011). Figure 1 depicts the conceptual overview of metacognition with references to the most cited research work in the area.

2.1. Metacognitive Events: Functions and Contents

In (Petukhova and Manzoor, 2021), *metacognitive events* are defined as reflexive activities that express any level of the sender's mindful awareness of own and others cognitive processes: pay attention, monitor, interpret and verify understanding, evaluate contents and feelings, and regulate and plan actions. These functional aspects of information processing concern *metacognitive functions* - abilities to recognize the various elements of one's mental state, and the ability to comprehend other's behaviour in terms of intentionality, variations and changes in intentional states.

Metacognitive events can be annotated using the Metacognition Assessment Scale (MAS) protocols designed to analyse interview transcripts with psy-

Metacognition Assessment Scale (MAS)		DIT/ISO24617-2 Dialogue Act	
Domain	**Function**	**Dimension**	**Communicative function**
Understanding One's <u>Own</u> Mind	Basic requirements	Auto-Feedback	pos./neg. attention
	Identification		pos./neg. recognition
	Relating variables		pos./neg. interpretation
	Differentiation		
	Integration		
	Decentration		
Understanding <u>Other's</u> Mind	Basic requirements	Allo-Feedback	pos./neg. attention
	Identification		pos./neg. recognition
	Relating variables		pos./neg. interpretation
	Differentiation		
	Integration		
	Decentration		
Master	Basic requirements	Auto-Feedback	pos./neg. evaluation
	1st level strategy	Auto-Feedback	pos./neg. execution
		Own Communication	error signal, retraction, self-correction
		Time Management	stalling
	2nd level strategy	Task& Task Management	various + (implied) Auto-Execution
		Interaction Management	various+ (implied) Auto-Execution
	3rd level strategies	Allo-Feedback	pos./neg. execution
			feedback elicitation (higher levels)
		Task& Task Management	various + (implied) Auto-& Allo-Execution
		Interaction Management	various+ (implied) Auto-& Allo-Execution

Table 1: Mapping between MAS domains and functions and DIT/ISO24617-2 dialogue acts.

chotherapy patients (Semerari et al., 2003). To the best of our knowlegde, this is the only taxonomy that can serve as a basis for our annotation efforts as it is suitable for our interactive setting. We compared concepts defined in the DIT^{++} (Bunt, 2006) and its subset, the ISO 24617-2 dialogue act annotation standard (ISO, 2020)[1], and the MAS metacognitive domains and functions. Table 1 provides an overview of our mapping efforts.

In both taxonomies, a special attention is paid to *feedback* acts which are crucial for the successful metacognitive functioning. DIT/ISO 24617-2 defines positive and negative feedback about sender's own (*auto*-feedback) and the partner's processing (*allo*-feedback). They correspond to the two MAS domains: (1) mental operations which involve knowing one's own mental states (*understanding one's own mind*), and (2) knowing mental states of the others (*understanding other's mind*), with six functions:

1. *basic requirements* refer to the ability to acknowledge own mental functions and existence of those in others, and to represent self and others as persons with autonomous thoughts and feelings;
2. *identification* – the ability to recognize one's own and other's cognitive and emotional states;
3. *relating variables* – the ability to establish relations among the separate components of one's own and other's mental state and between the components of mental states and behaviour;
4. *differentiation* – the ability to distinguish one's own and other individuals' mental states;
5. *integration* – the ability to work out coherent descriptions of one's own and other's mental states;

6. *decentration* – the ability to produce interpretations independent of other people's knowledge.

The first two MAS functions concern *monitoring* activities and involve gaining *attention* and setting *recognition* of each other's behaviour. The last three MAS functions concern *interpretation* processes. Higher processing levels involve *mastery* of regulation and control activities based on the ability to assess one's representations and mental states (*evaluation* as a *mastery basic requirement*, MBR), and the ability to implement effective action strategies to accomplish cognitive tasks or to cope with problematic mental states (*task execution*). Levels of mastery regulation and control strategies reflect the complexity of the metacognitive operations needed to be executed. For example, if an interlocutor acts directly on her/his own understanding and evaluation of the current dialogue state slightly modifying her/his mental states, this concerns the *1st level strategies*. Interlocutor may **report** (in-)consistencies in her/his mental state, successes or failures in execution of certain action. *Own communication* and *time management* acts are typically concerned with this level strategies. When an interlocutor performs a certain type of behaviour actively **modify**ing one's own level of attention, concentration, interpretation and evaluation, voluntarily thinking or not thinking about a problem, and adjusting her/his previous mental state, these acts require the *2nd level strategies*. The *3rd level strategies* involve **adopt**ing a rational and critical attitude to the beliefs that are behind a problematic state using one's knowledge about others' mental states, e.g. regulate interpersonal problems accepting one's personal limits and errors, or influencing events. These three levels strategies imply various reflection and regulation efforts.

Figure 2: Agent Graphical User Interface. From up to down: dialogue window with an emoji depicting the agent's satisfaction with the partner's strategies; generated preference profile for a human participant, with brighter orange colours for increasingly negative options and brighter blue - for positive ones; set of actions in response to agent's offers, to request agent's preferences, to propose (counter-)offers, concessions, final deals or terminate negotiation.

Metacognitive contents are the ideas and beliefs linked to beliefs about beliefs and are task dependent. They are concerned with beliefs about the current state of the world including partners' states (what do I know, what do I know about others, what am I asked, what can I do, what has happened before), and an action to be taken in that situation (e.g. give information, run tests, examine something, reason about others, change attitude). In our think aloud experiments, it is assumed that participants will verbalize metacognitive contents making them accessible for further modelling and assessments.

3. Experimental Design

In dialogue, to directly access participants' metacognitive knowledge, experiences and regulation strategies, and to understand partner's cognitive processes, Concurrent Think-Aloud (CTA) protocols were designed (Ericsson and Simon, 1980). Participants were encouraged to verbalize their experiences, thoughts, actions, and feelings whilst interacting with an artificial agent through its graphical interface. This method provides direct "real-time" insight into the cognitive processes employed by the participants. Participants were asked/prompted to not only verbalize their decisions and what they think about the decisions of their partners but also explain why they think these decision lead to certain outcomes or try to look ahead and predict how certain decisions will influence the outcomes.

3.1. Use Case and Scenario

Negotiations are chosen as the use case, a domain where the importance of metacognition has been empirically proven to significantly influence decision-making processes (Galluccio and Safran, 2015). In the designed scenario, the human participant - Doctor - negotiates with an Interactive Cognitive Agent - Simulated Patient (SP)- various plans for treatment of diabetes of Type 2. SPs hold different preferences and are trained to take actions and make decisions that people would take and make in real-life scenarios (Petukhova et al., 2019). The patient-doctor negotiation scenario is based on the recommendations of the International Diabetes Federation (IDF, 2017) addressing four issues: (1) medication, (2) diet, (3) activity and (4) exercise recommendations. 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 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.

3.2. Set Up and Experimental Protocols

A negotiation *session* consists of six think-aloud *rounds* featuring scenarios of various complexity. Human participant is assigned the role of a doctor and receives the background story and an automatically generated preference profile as depicted in Fig. 2. The task is to negotiate an agreement with an SP (agent) - select exactly one value for each issue, exchange and elicit offers concerning the agent's options. No further rules on the negotiation process, order of discussion of issues, or time constraints are imposed. Negotiators are allowed to withdraw or re-negotiate previously made agreements within a round, or terminate a negotiation. Participants were asked to verbalize everything what is on their mind: explain what action they are going to perform and why, and what they think about the agent's actions and the rationales/strategies behind.

Prior to experiments, participants are educated about the purpose and the course of the study. Subsequently, a declaration of informed consent is signed. The anonymity of the participants is guaranteed followed by European laws on personal data protection (GDPR, 2018 https://gdpr-info.eu/). During the experiment, the participant sits in a comfortable chair in front of a laptop monitor. Participants are briefed on how to interact with the agent using the GUI.

10 subjects (aged between 19 and 25 years, 4 male and 6 female, proficient but non-native speakers of English) participated in the experiments, each involved in one negotiation session. Interactions were recorded with LENOVO THINKPAD E570 (Core i7) equipped with the webcam (720p) and dual microphones. The lighting and sound conditions were close to the conditions
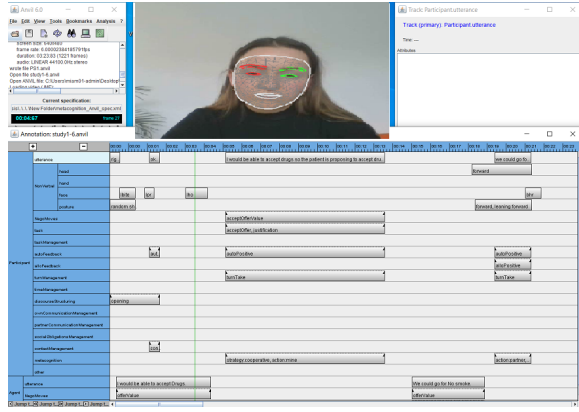
Figure 3: Example of Anvil annotations with video featuring OpenFace facial landmark detection.

of the intended interactive application entailing a fairly good but not perfect acoustic and video quality. Video background was white. The goal of the **technical set-up** was to let participants interact as free as possible keeping them away from any distraction and enabling stable continuous recording and logging of multimodal data.

3.3. Data Collection and Processing

Participants behaviour was video and audio recorded. The mouse and keyboard GUI actions were captured automatically and time aligned using the python library `atbswp`[2]. Video recordings (640x480, 30fps) were done with `ffmpeg`[3] tool applying the MPEG codec; audio signals were stored in `wav` format having frame per buffer (fpb) 1024 at rate of 44100. The video quality was sufficient to be further processed using `OpenFace`[4] and `MediaPipe`[5] libraries to extract features and facial landmarks reliably, see Fig. 3.

Participants nonverbal behaviour, mainly gaze re-direction, facial expressions, head gestures, posture shifts were coded using the scheme proposed in (Petukhova and Bunt, 2012) which supports a rather detailed characterization of movements in terms of low-level behavioural features, such as changes in muscular activity and types of these changes, direction, trajectory, speed, intensity and periodicity of movements. A moderate inter-coder agreement coding type of visible movement was observed (standard Cohen's kappa of 0.62). We measured the coding RealTime Factor (RTF), the amount of time spent on transcriptions and coding, as being RTF 19 on average and meaning that a coder spent 19 minutes annotating 1 minute of video. The detected mouse and 'on-screen' (touch screen) actions were categorized as 'mouse up', 'scroll', 'move to', 'mouse down' and 'sleep'. Timing, duration, speed

and direction of movements were computed automatically. The coding scheme is recently finalised, quality assessment as well as neural network based classification experiments are in progress.

Video, GUI logs and audio signals were synchronised and mixed using `FFmpeg`[6]. Participants' speech was transcribed in `PRAAT`[7]. Annotations were performed in `Anvil`[8].

3.4. Annotations

The collected multimodal data was annotated with multidimensional DIT/ISO 24617-2 tagset taking into account MAS categories specified in Section 2.1. Ten DIT/ISO 24617-2 dimensions address the information about a negotiation (*Task*); the sender's processing of dialogue contributions, awareness of her/his mental states and knowledge of the basic regulation strategies (*Auto Feedback*) or similar cognitive processing by the partner (*Allo Feedback*); the management of difficulties in the sender's contributions (*Own Communication Management*), or similar difficulties of the partner (*Partner Communication Management*); maintaining contact (*Contact Management*); need for time (*Time Management*); the allocation of the sender role (*Turn Management*); the *Structuring* of the *Dialogue*; and the *Management* of *Social Obligations*. A recently added ISO 24617-2 dimension (2. Edition) deals with the management of the negotiation, but also with the management of decision-making processes and metacognitive tasks (*Task Management*). In Task Management utterances, the sender verbalizes beliefs concerning her/his understanding and evaluation of the current negotiation state, procedures and strategies, anticipated (un-)favourable actions, explains why certain own and partner's decisions are or should be made. Dialogue act annotations were linked to different type of primary data: to verbal and (non)verbal behaviour and logged GUI actions. The inter-annotator agreement was measured and ranges from moderate to almost perfect for specific dimensions, see Table 2.

To annotate metacognitive contents, we enriched functional aspects of dialogue act with specifications of semantic content related to *reflection* and *decision*. These two categories were empirically observed when analysing think aloud dialogue transcripts. In their (meta)cognitive thinking verbalizations, participants mostly refer to their understanding of own and partner's mental states as the basis for their past and future decisions which concern (i) the sender's beliefs about her/his own and partner's past and future *actions*, (ii) the interpretation of the participants' *preferences*, (iii) own and partners' applied or to apply negotiation *strategies*, and (iv) *consequences* or *conditions* under which certain intermediate and final outcomes can be reached.
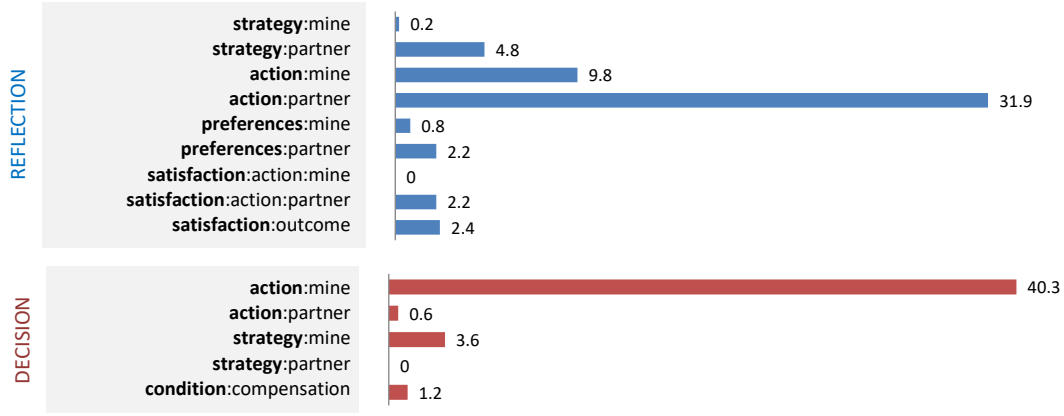

Figure 4: Dialogue acts with metacognitive content annotated in think-aloud interactions.

To sum up, dialogue acts were annotated across eleven DIT/ISO 24617-2^{++} dimensions as having a *communicative function* which specifies the sender's intention expressed in a dialogue utterance, and *semantic content* which indicates what the utterance is about. DIT^{++} tagset of possibly *qualified* communicative functions was applied. We distinguished two types of semantic content: (1) negotiation specific content which concerns negotiation moves such as offer, counter-offer, compromise as defined in (Petukhova et al., 2017); and (2) metacognitive content which concerns the sender's reflection efforts and decision related beliefs. The `Anvil` multi-tier annotations functionality was exploited to establish temporal, structural and semantic dependence relations of various types: (i) between (parts of) two or more segments in primary data; (ii) between functional components of two or more dialogue acts; (iii) between functional and semantic components of two or more dialogue acts; and (iv) between domain related and metacognitive components of two or more dialogue acts.

4.   Experimental Results

The analysis of the annotated data shows that think-aloud interactions are mostly concerned with task-related exchanges, i.e. task and task management acts prevail constituting from 16.6% to 27% of all dialogue acts performed (see Table2). The more task management acts were performed the higher success rate in negotiations was achieved, only 4% of all interactions resulted in negotiation termination. This confirms previous findings that explicit metacognitive thinking positively effects negotiation task performance and outcomes. The other important aspect of interactions concerns the understanding and communication of interlocutor's intentions related to one's own and partner's preferences, strategies and actions, as well as the assessment of the ongoing (meta)cognitive processes. Auto- and Allo-Feedback acts were therefore frequently observed in think aloud experiments, 19.2% and 9.6% respectively. Task management and feedback acts are directly concerned with metacognitive processes, with reflection efforts and regulation strategies, as illustrated in examples in (1):

(1) (i) Human: I will try to meet in the middle [preferences]
     (ii) Human: I would be stubborn too
     (iii) Human: I think now the patient wants to ...

We represent metacognitive contents as a set of attribute-value pairs corresponding to one's beliefs encoding reflections and decisions about preferences, strategies, actions and conditions, and one's beliefs about other's beliefs of the same kind. Following the ISO 24617-2 guidelines, a simple plug-in for representing semantic content as a list of attribute-value pairs is used. In the example in (2), the `<avContent>` element specifies this semantic information in terms of attribute-value pairs of an utterance in (1i).
The example in 2 illustrates this.

(2)

```
<dialogueAct xml:id="dap2TSKM74"
    sender="#p2" addressee="#p1"
    dimension="taskManagement"
    communicativeFunction="inform"
    target="#fsp2TSKMCV74">
    <avContent xml:id="av74"
    target="#fsp2TSKMCV74"
    attribute="decision:strategy:mine"
    value="middle"/>
    <contentLink dialAct="#dap2TSKM74"
    content="#c74"/>
</dialogueAct>
```

We were able to identify a range of feedback-related metacognitive indicators which often concerned various head movements, e.g. head nodding, shaking and waggles accompanied by a noticeable smile, lip pout or compression, raising or lowering eyebrows, conjugated lateral eye movements (CLEMs), and posture shifts, e.g. leaning forward, backward or aside, shifting one's weight in the chair.
Turn and time management acts were frequently observed (9.5% to 17.9%) to co-occur with and often precede metacognitive events concerned with sender's reflection about the partner's action as in (3):

DIT/ISO 24617-2 dimension	Relative frequency (in %)	Inter-annotator agreement (Cohne's kappa)
Task	16.6	0.71
Task Management	19.2	0.59
Auto-Feedback − >	9.6	0.72
. . . *attention*	12.3	
. . . *recognition*	7.1	
. . . *interpretation*	16.8	
. . . *evaluation*	18.7	
. . . *execution*	45.1	
Allo-Feedback − >	13.7	0.86
. . . *attention*	0.3	
. . . *recognition*	2.1	
. . . *interpretation*	44.1	
. . . *evaluation*	52.9	
. . . *execution*	0.6	
Turn Management	17.9	0.71
Time Management	12.6	0.96
Contact Management	1.7	0.96
Discourse Structuring	5.9	0.77
Own Communication Man.	2.5	0.97
Partner Communication Man.	0.0	0.91
Social Obligation Management	0.3	1.0

Table 2: Distribution and inter-annotator agreement of the annotated dialogue acts across ISO 24617-2^{++} dimensions in think-aloud interactions.

(3) (i) Human: He wants ... *uhm* something bad for him
(ii) Human: ... *uhm* $_{smile}$ ... I don't agree with this

Our experiments showed that the content of participants' metacognitive events was largely concerned with the **reflection** beliefs (54.3%) compared to the verbalised **decisions** (45.7%). The interpretation and evaluation of *partner's actions* (31.9%) predominantly occupied the sender's mind. Participants often reflected on their *own actions* (9.8%), partner's *strategies* (4.8%), interlocutors' *preferences* (3.0%), and on how *valuable* the performed actions were for the sender and for the overall negotiation outcome (4.6%). The metacognitive events concerning reflection beliefs were annotated as feedback acts and were used by the sender as a basis to decide what *action* to perform next (40.3%), and what strategy to follow (3.6%) based on the accumulated utility and values of alternative options to compensate for possible own or partner's loses (1.2%). Figure4 depicts distribution of metacognitive content expressed in dialogue acts.

The analysis of mouse movements and clicking behaviour largely shows the following pattern: slowing down and pausing in mouse movements indicated reflection moments, revealed sender's lower confidence and often happened around decision points; changes in the direction of mouse movements after pausing indicated retractions, while continuation in the same direction were often interpreted as gain in confidence; and sudden termination of any movement or GUI activity meant either technical problems or unexpected situations experienced by the sender. The more in-depth analysis of GUI and verbalised actions is required and will be performed in the nearest future for which the initial technical set up will be extended, more data collection and machine learning experiments carried out.

5.   Conclusion

The paper provides methodological insights and experimental design to assess metacognitive processes and contents relevant for human-agent interaction. We reviewed existing models of metacognition and available metacognition assessment instruments. Metacognitive events were defined by mapping and using the systematic analysis of concepts related to metacognitive activities, metacognitive domains and functions, and open assessment protocols. The identified metacognitive concepts were mapped to DIT/ISO 24617-2 dialogue acts - the concepts that the dialogue research community is used to operate on in dialogue modelling and system design. Subsequently, metacognitive functions and contents of dialogue acts were explicitly defined, plugged into the DIT/ISO 24617-2^{++} representations and annotated in multimodal interactive data collection experiments.

Experiments comprised behavioural observations, pragmatic and semantic analysis in think-aloud human-agent interactions. Doing this, we aimed at establishing relations between overall and task-specific metacognitive thinking, the complexity of metacognitive processes activated in social interactive setting and their multimodal behavioural indicators. We aimed at understanding to what extend and how dialogue participants use their metacognitive knowledge and regulation strategies in dialogue. Participants were asked to verbalise their metacognitive thinking to make metacognitive processes directly accessible and assessible through the use of any metacognitive indicators associated with certain type of dialogue acts in human-agent interactions. The identified indicators can serve in the future as a basis for data-driven metacognitive states classification needed for real-time monitoring and continuous metacognitive assessment pursuing two purposes: (1) to enable the system's behaviour which is intelligent, human-like and adaptive to their users, and (2) to encourage and support users to behave in the same way.

There are certain limitations of this study and much room for improvement and further research. Participants demography needs to be more diverse in age group. In the follow up experiments, we aim to replicate gender and age differences in a human-agent setting, e.g. to manipulate the agent's respective characteristics. A severe limitation of think-aloud experiments was that participants were non-native English speakers and sometimes not that verbally fluent, however metacognition may be confounded with verbal ability.

Our future research efforts will focus on the automatic detection and classification of nonverbal and GUI metacognitive indicators using modern machine learning algorithms and deep neural networks. We will refine our OpenFace based models and perform an in-depth quantitative and qualitative analysis of the logged GUI data.

6. References

Bannert, M. and Mengelkamp, C. (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.

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

Brune, M., Dimaggio, G., and H Lysaker, P. (2011). Metacognition and social functioning in schizophrenia: Evidence, mechanisms of influence and treatment implications. *Current Psychiatry Reviews*, 7(3):239–247.

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

Bunt, H. (2006). Dimensions in dialogue act annotation. In *LREC*, pages 919–924.

Cross, D. R. and Paris, S. G. (1988). Developmental and instructional analyses of children's metacognition and reading comprehension. *Journal of educational psychology*, 80(2):131.

Dunning, D. (2011). The dunning–kruger effect: On being ignorant of one's own ignorance. In *Advances in experimental social psychology*, volume 44, pages 247–296. Elsevier.

Dwyer, C. P., Hogan, M. J., and Stewart, I. (2014). An integrated critical thinking framework for the 21st century. *Thinking skills and Creativity*, 12:43–52.

Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational research review*, 1(1):3–14.

Ericsson, K. A. and Simon, H. A. (1980). Verbal reports as data. *Psychological review*, 87(3):215.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, 34(10):906.

Flavell, J. H. (1981). Cognitive monitoring. *Children's oral communication skills*.

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

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

ISO. (2020). *Language resource management – Semantic annotation framework – Part 2: Dialogue acts. ISO 24617-2, Second Edition*. ISO Central Secretariat, Geneva.

Kleitman, S. and Moscrop, T. (2010). Self-confidence and academic achievements in primary-school children: Their relationships and links to parental bonds, intelligence, age, and gender. In *Trends and prospects in metacognition research*, pages 293–326. Springer.

Kuhn, D. and Dean, Jr, D. (2004). Metacognition: A bridge between cognitive psychology and educational practice. *Theory into practice*, 43(4):268–273.

Kuhn, D. (2000). Metacognitive development. *Current directions in psychological science*, 9(5):178–181.

Livingston, J. A. (2003). Metacognition: An overview.

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

Martinez, M. E. (2006). What is metacognition? *Phi delta kappan*, 87(9):696–699.

Nelson, T. O. and Narens, L. (1994). Why investigate metacognition. *Metacognition: Knowing about knowing*, 13:1–25.

Paris, S. G., Winograd, P., et al. (1990). How metacognition can promote academic learning and instruction. *Dimensions of thinking and cognitive instruction*, 1:15–51.

Petty, R. E., Briñol, P., Tormala, Z. L., and Wegener, D. T. (2007). *The role of metacognition in social judgment.* The Guilford Press.

Petukhova, V. and Bunt, H. (2012). The coding and annotation of multimodal dialogue acts. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 1293–1300.

Petukhova, V. and Manzoor, H. E. (2021). Towards the iso 24617-2-compliant typology of metacognitive events. In *Proceedings of the 17th Joint ACL-ISO ISA Workshop*, pages 14–19.

Petukhova, V., Bunt, H., and Malchanau, A. (2017). Computing negotiation update semantics in multi-issue bargaining dialogues. In *Proceedings of the SemDial*, Saarbrücken, Germany.

Petukhova, V., Sharifullaeva, F., and Klakow, D. (2019). Modelling shared decision making in medical negotiations: Interactive training with cognitive agents. In *Proceeding PRIMA Conference*, pages 251–270. Springer.

Pressley, M., Borkwski, J. G., and Schneider, W. (1989). Good information processing: What it is and how education can promote it. *International Journal of Educational Research*, 13(8):857–867.

Schraw, G. and Moshman, D. (1995). Metacognitive theories. *Educational psychology review*, 7(4):351–371.

Schraw, G., Crippen, K. J., and Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in science education*, 36(1):111–139.

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

Semerari, A., Carcione, A., Dimaggio, G., Falcone, M., Nicolo, G., Procacci, M., and Alleva, G. (2003). How to evaluate metacognitive functioning in psychotherapy? the metacognition assessment scale and its applications. *Clinical Psychology & Psychotherapy*, 10(4):238–261.

Tarricone, P. (2011). *The taxonomy of metacognition*. Psychology Press.

Whitebread, D., Coltman, P., Jameson, H., and Lander, R. (2009). Play, cognition and self-regulation: What exactly are children learning when they learn through play? *Educational and Child Psychology*, 26(2):40.

Yeung, N. and Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. *Philosophical Transactions of the Royal Society*, 367(1594):1310–1321.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American educational research journal*, 45(1):166–183.