An explanation-oriented inquiry dialogue game for expert collaborative recommendations

Qurat-ul-ain Shaheen [a,c,*], Katarzyna Budzynska [b] and Carles Sierra [a]

[a] *IIIA, CSIC, Barcelona, Spain*
E-mails: qurat@iiia.csic.es, sierra@iiia.csic.es
[b] *Laboratory of The New Ethos, Faculty of Administration and Social Sciences, Warsaw University of Technology, Warsaw, Poland*
E-mail: Katarzyna.Budzynska@pw.edu.pl
[c] *Software Competence Centre Hagenberg, Hagenberg, Austria*
E-mail: qurat-ul-ain.shaheen@scch.at

Abstract. This work presents a requirement analysis for collaborative dialogues among medical experts and an inquiry dialogue game based on this analysis for incorporating explainability into multiagent system design. The game allows experts with different knowledge bases to collaboratively make recommendations while generating rich traces of the reasoning process through combining explanation-based illocutionary forces in an inquiry dialogue. The dialogue game was implemented as a prototype web-application and evaluated against the specification through a formative user study. The user study confirms that the dialogue game meets the needs for collaboration among medical experts. It also provides insights on the real-life value of dialogue-based communication tools for the medical community.

Keywords: Inquiry dialogue game, collaborative decisions, expert decisions, explainable artificial intelligence, human-centred computing

1. Introduction

As the human society has become more digitally connected, it has developed an increased appreciation for interdisciplinary collaboration [23]. Healthcare is one such domain which has a long tradition of interdisciplinary collaboration amongst different medical experts [67]. Imagine a distributed health recommendation system where different experts come together to find the best possible diagnosis for a patient. These experts could be human agents, artificial agents or a combination of human and artificial agents. The goal of the collaboration would be to integrate multiple perspectives through knowledge transfer and conflict resolution in order to recommend the best possible diagnosis. Additionally, the system should offer explanations for its recommendation in order to build trust between the users and the system [44].

---

*Corresponding author: Qurat-ul-ain Shaheen, Software Comptence Centre Hagenberg, Hagenberg, Austria. E-mail: qurat-ul-ain.shaheen@scch.at.

Table 1

Example dialogue between three medical experts α, β, γ

Id	Dialogue
1	α: Patient A is a 48 years old female. Her symptoms are fatigue, constipation, weight gain, drowsiness, and dry skin. It looks like a case of depression. I recommend lifestyle changes such as healthy eating and walking. I note that she has a family history of autoimmune diseases.
2	β: Can you justify your diagnosis of depression?
3	γ: I would say that it looks like a typical case of hypothyroidism.
4	α: Because her symptoms are typical of depression.
5	γ: I note that she does not have headache or back pain, which are common in case of depression.
6	β: Can you justify why you diagnose hypothyroidism?
7	γ: Because she has dry skin and a family history of auto immune diseases, which are typical for hypothyroidism.
8	α: You might be right. I recommend that we test her TSH, T4 and T3 levels.
9	β: I agree. I also recommend doing a blood complete count to rule out other similar conditions like anaemia.
10	γ: I agree to testing TSH, T4 and blood complete count. I think we can close the discussion now.
11	β: I agree.
12	α: I disagree. What about testing for T3?
13	γ: Why do you want to test for T3?
14	α: Because I want to rule out Hyperthyroidism.
15	β: Yes, it makes sense.
16	γ: I don't think it is necessary to test T3 at this stage since she is not asymptomatic. And her symptoms are closer to Hypothyroidism.
17	α: Okay. I think we can close the discussion now.
18	β: I agree.
19	γ: I agree.

As an example scenario, consider the dialogue given in Table 1 between three medical experts represented by α, β and γ. The first column of Table 1 shows the identifier for the statement while the second column indicates the expert name followed by the statement they are making. The experts are participating in a semi-structured formal discussion similar to the ones that take place in oncology [43]. They are already aware of the objectives and format of the meeting before the dialogue starts so there is no build up on the objectives of the discussion as is the case in oncology meetings. We cover this in detail in Section 3. In this example, the meeting is not about a cancer diagnosis, but rather a more general diagnosis for a patient denoted by *A*. This is a contrived example, aimed at illustrating the conversation flow among healthcare experts who are participating in a semi-structured discussion to collaboratively make a recommendation for a patient. The goal of this example is not to provide a comprehensive medical discussion on the patient's diagnosis as this is not the objective of this work. Moreover, note that the dialogue only shows the knowledge and reasoning of the experts to the extent that they choose to reveal through their conversation. It does not show the complete knowledge base or the reasoning process of the expert agents (who can be human or artificial).

In this example, we assume that agent α loosely represents a clinical psychologist, agent β a general practitioner and agent γ an endocrinologist (specialist in ear, nose and throat). Agent α first presents the facts of the case and offers his own diagnosis (depression). This is challenged by agent β. α then justifies their stance which is rejected by γ. γ then proposes their own diagnosis. β asks γ to explain their diagnosis. Subsequently, both α and β accept γ's explanation. Then α proposes some tests to which β agrees and propose an additional test. γ agrees and proposes yet another test. However, their suggestion is ignored by the other two as they propose to close the discussion. However, γ does not consent and
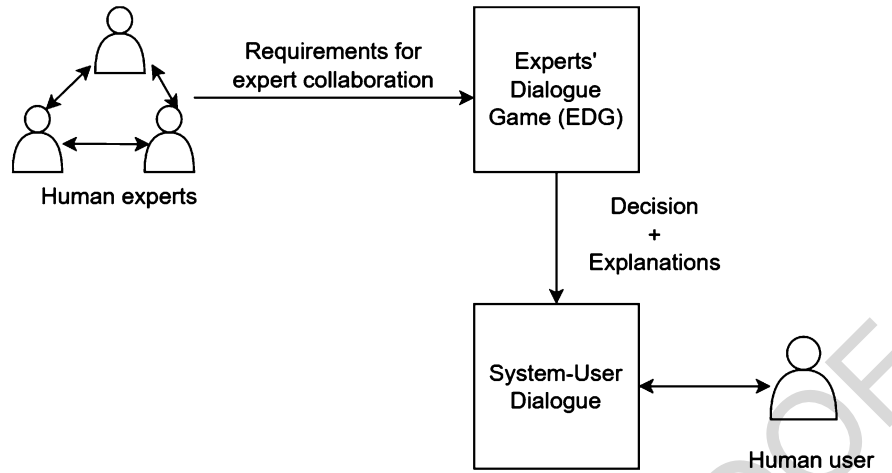
Fig. 1. Workflow diagram of an explainable multiagent recommendation system employing EDG.

asks the other agents to respond to their suggestion. They do respond and subsequently γ agrees to end the discussion. Throughout the discussion, the agents try to collaborate in a cooperative manner. We use this dialogue as a running example throughout the rest of the paper to illustrate our approach.

As a first step towards realising such a hybrid human-artificial multiagent system capable of such a dialogue, we propose a novel interaction protocol between experts agents (whether human or artificial). We call this protocol as *Experts' Dialogue Game (EDG)*. Figure 1 visualises how EDG would fit into the pipeline for building such a system. As a first step, real-life consultations among medical experts are formalised as a requirement specification. Building on this specification, EDG is defined as a dialogue game among the participants in a hybrid human-artificial multiagent system (MAS). The output of EDG is a recommendation from the system along with a sequence of explanations justifying the recommendation. These explanations can then be plugged into a system-user dialogue to justify the system's recommendation to the user. The user in this case could be the patient himself or a physician in charge of the patient. While most of the current literature [3,17,36,65,76] focuses on explaining the working of a system to a human user through a system-user dialogue, EDG investigates how an explanation dialogue can be used within a multiagent system as part of agent reasoning.

Dialogue games are *dialectical systems* [22] in the tradition of informal logic [77] and are formally defined as verbal interactions between two or more players according to some pre-defined rules for the dialogue [39]. Each interaction is specified with the use of *locutions* which represent speech acts permitted in a given dialogue [62]. Dialogue games require that each participant maintains consistency across its statements, also called *commitments*, at any point in the dialogue [77]. Characteristics of these verbal interactions are typically defined in multiagent communication according to the popular typology introduced by Walton and Krabbe [77]. Dialogue games have a long tradition of being used to solve formal problems as well as to model natural language communication in real-life settings. In the first case, they have been employed to search for formal logical proofs in the tradition of Lorenzen [1] and leading to the field of *dialogical logic*, as well as in the *prescriptive* approach such as Hamblin's system [22], to disallow logical fallacies during natural language argumentation. In the second case, they are used to study the communication dynamics in real life settings, giving rise to the *descriptive* approach to dialogue games [11,27,36,40,53,64]. This approach can also be used to inform computational models of interactions between agents in a multiagent system.

The goal of EDG is for the participants to collaboratively find the best recommendation through exchange of knowledge and mutual agreement. In Walton–Krabbe's typology, this scenario fits an *inquiry* type of dialogue in which an initial situation is a need to have proof (i.e. in our case – to find the best recommendation), the participants' goal is to find and verify an evidence (i.e. to consult an observation), and the goal of the dialogue is to prove or disprove a hypothesis (i.e. to argue for or against a recommendation). This is in contrast to a *deliberation dialogue* where the initial situation is a need for action, the participants' goal is to influence the outcome and the goal of the dialogue is to reach a decision on the best possible action [77]. In order to generate explanations of the recommendations, we define agents' communicative behaviour in the dialogue through *explanation-based* illocutionary forces which can then be traced back and retrieved in response to a query and presented to the user as an explanation of the recommendation: (1) *explanation requests* such as *wh-explain*(p) when the speaker knows that p is the case, but does not understand why it is the case; *wh-justify*(p) when the speaker does not agree that p is the case and asks the hearer for the justification of p; and *wh-clarify*(p) when the speaker does not understand a term in p and asks for the clarification of this term; and (2) *explanation replies* such as *explain*(p) when the speaker provides an explanation for p; *justify*(p) when they provide a justification of p; and *clarify*(p) when they provide a clarification of the term in p.

The contribution of this work is threefold. First, we present a requirement specification for collaboration between experts. While the requirement specification is grounded in the medical domain, it focuses on the general communication issues that can come up during expert collaborations. Hence, it can be abstracted to consultations among experts in general. Next, inspired by the tradition of dialogue embedding [77], we combine explanation-based illocutionary forces in an inquiry dialogue to generate richer traces for the inquiry process than what is possible with the *assert* locution, typically used in inquiry dialogues. Moreover, the dialogue game is unique in that sense that it meets the requirement specification from the domain experts. While the combination of different dialogue types is not ground breaking formally, the proposal makes an important methodological step into the applications of such formal dialogue systems into real-life domains which is empirically grounded in users' requirements. Furthermore, to the best of our knowledge, no other *descriptive* dialogue game has focused on interactions among experts. The rich game traces allow the system to be transparent and can be used to explain the recommendations of the system to a human user through human-machine interfaces such as verbal and visual interaction. Finally, we evaluate the dialogue game against the requirement specification and verify the evaluation through a formative user study. Thus, we introduce the methodology of user-centred software engineering practice to dialogue games. While other works [11,27,36,40,53,64] in *descriptive* dialogue games have used insights from their domain of interest to inform the dialogue games, none of them have provided a requirement specification as far as we are aware.

The rest of this work is structured as follows. Section 2 summarises related work. Section 3 presents the requirement specification for expert collaborations. Section 4 formally presents the Experts' Dialogue Game (EDG) while Section 5 provides implementation details of a web-based platform implementing the EDG. In Section 6, EDG is evaluated against the requirement specification presented in Section 3. Section 7 highlights user perspectives on EDG and the platform described in Section 5. Finally, Section 8 concludes the paper and provides directions for future work.

2. Related work

This section summarises related work on explanatory and inquiry dialogues, dialogue games in multi-agent systems for healthcare, argumentation in healthcare and communication tools for multidisciplinary collaborations in healthcare.

Amgoud et al. [2] present an argumentation system for resolving inconsistencies in an agent's knowledge base. Subsequently, they show how dialogue game theory can be applied on top of this to realise the different dialogue types in Walton and Krabbe's typology [77]. A minimal framework for an explanatory dialogue system is presented in Walton [76]. The goal of the dialogue is for an *explainer* (an entity that explains, usually the system) to fill in the gaps in the knowledge base of the *explainee* (the target of the explanation, usually a human user) by informing them why something happened. This is considered as a transfer of understanding from the explainer to the explainee. Building on Walton's minimal framework, Arioua et al. [3] combine an explanatory dialogue with argumentative illocutions. The explanatory illocutionary force is used by the system to explain the behaviour of some phenomena to the user while the argumentative force helps to resolve inconsistency in the knowledge bases of participants. In addition to commitment stores for each participant, they introduce an *understanding store* for the user which stores the missing links in understanding rather than what is currently understood. The discharging of all issues in the understanding store confirms a successful transfer of understanding.

Madumal et al. [36] analyse human-human and human-agent explanatory dialogues from various domains and propose an explanatory dialogue protocol based on induction. Their protocol also combines an explanatory dialogue with argumentative faculty. The goal of the dialogue is the same as proposed by Walton [76]. They use double acknowledgement for confirming understanding, that is, acknowledgement by the explainee on being satisfied with the explanation and explainer's return acknowledgement. Dennis et al. present an explanatory dialogue game for explaining the behaviour of a Belief–Desire–Intention System [17]. Their protocol, like EDG, is grounded in dialogue game theory from informal logic using argumentation schemes and critical questions rather than argumentation theoretic dialogue [75]. The goal of the dialogue in this case is to understand system behaviour through comparing traces of the system from different participants. Similar to these works, we use explanation-based illocutionary forces to fill in missing gaps in the knowledge bases of participants.

Ilia et al. [65] extend an information seeking dialogue with explanatory illocutionary forces. The goal of the dialogue is to offer factual and counter factual explanations of a classifier's output to a human user. They also evaluate the protocol through a user study and process analytics. Prakken and Ratsma [54] propose a *case-based explanation dialogue* to explain the outcome of a linear binary classifier. The goal of the dialogue is to provide a model-agnostic local explanation. The explanations in this case are not trying to explain the system reasoning but rather trying to come up with reasons to justify the system result. The dialogue game starts with the proponent of the dialogue presenting a similar example from the training set with the same outcome as the current instance. The opponent can argue against this using two strategies. The first is to use counter examples from the training set. The second is to highlight the differences between the current instance and the instance being presented as a *justification* by the proponent. A successful explanation amounts to a winning strategy for the proponent. However, unlike these approaches, we incorporate the explanatory illocutionary forces in the reasoning process itself rather than to only explain why the system behaved in a certain way. While these works target a transfer of understanding from the system to a human user, EDG involves explanations amongst the reasoning agents.

Although not as common as persuasion dialogues, a few other works have explored inquiry dialogues as stand alone dialogues or in combination with other dialogue types. Bex et al. [7] combine an inquiry dialogue with a persuasion dialogue for discussions between criminal investigators. The goal of the dialogue is to come up with the most robust explanation. They assume an adversarial setting in which each agent advocates for its own preferred explanation. Unlike this work, EDG assumes a cooperative setting where the main goal is to come up not only with the most robust explanations but decisions as well. Black and Atkinson propose a framework that embeds an inquiry dialogue over beliefs with a persuasion dialogue over actions. The inquiry dialogue allows the participants to collaboratively decide what to believe whereas the goal of the persuasion dialogue is to collaboratively decide what is the best action to do in order to reach the proponent's goal. Once all the arguments have been given, it is upto to the proponent of the dialogue to make the final decision based his personal preferences [8].

Black and Hunter [9,10] propose two types of inquiry dialogues, which they call as *argument inquiry* and *warrant inquiry*. The goal of *argument inquiry dialogue* [9] is for two agents to jointly construct an argument for a claim by sharing relevant beliefs. The protocol has three moves and allows nesting of argument inquiry dialogues. In addition to a conventional commitment store, they also introduce a *question store* which keep tracks of the premises that need to be proven in order to prove the claim representing the dialogue topic. They prove soundness and completeness for their protocol. The goal for a *warrant inquiry dialogue* [10] is for two agents to share arguments to jointly construct a *dialectical tree* in order to determine the acceptability of a particular argument. The main difference between the two types of dialogues is that argument inquiry is not concerned about the acceptability of the argument constructed while the latter is. Warrant inquiry dialogue allows embedding argument inquiry dialogue and also involves a question store like the former. They also provide a strategy for selecting the next move for a participant for both types of dialogues. However, to the best of our knowledge, none of these works combine an explanatory dialogue with an inquiry dialogue to make agent reasoning explainable.

Other works have incorporated argumentation and dialogue games in multiagent systems to provide clinical decision support in a distributed environment such as cancer diagnosis and management. Huang, Jennings and Fox [24,25] present a multiagent architecture for medical decision support in an interdisciplinary setting. The architecture has four components; a three layered knowledge base, a centralised working memory, a communications manager and a human-computer interface. The architecture also covers decision making under uncertainty, task management and agent cooperation. The communication manager works with communication *primitives* or locutions and a communication protocol. However, the locutions in this case are geared towards managing the tasks in a distributed environment rather than a discussion amongst different experts. For example, the locutions *request, accept, reject* and *alter* are used in the task allocation stage. The locution *inform* is used to report on the allocated task while the locution *propose* is used to recommend a treatment plan in response to a *query*. In contrast, EDG is focused on facilitating the discussion amongst the experts rather than a distributed management of responsibilities.

Beveridge and Fox [6] use a dialogue game as an interface between the underlying task structure and ontological knowledge and the spoken dialogue generation system. They implement their approach to provide clinical decision support for breast cancer diagnosis. They use several locutions as *initiating* locutions. For example, *inform* is used to present new information, *instruct* is used to request an action from the user, *query-yn* to ask a question with a yes or no answer and *query-w* to elicit a value from the user as part of data entry. They treat the dialogue started from each initiating locution as a sub-game. In contrast, the query locutions in EDG are targeting towards incorporating different types of explanations

into the multidisciplinary discussion amongst experts. This is because the goal for EDG is to support collaborative discussion among experts rather than supporting an individual user with decision support.

Vasileiou et al. [73] present an argumentation-based justification dialogue between two participants. The explainee is a human user who wants to understand the explainer's (artificial intelligent agent) reasoning. The dialogue game has four locutions, two of which are reserved for each of the participants. The game only allows a single locution per turn. They provide evaluation of their dialogue game through a user study and discuss its properties. Rago et al. [56] present the notion of explanatory dialogue between two participants as an *Argument eXchange (AX)*. They discuss desirable properties of AX for agents equipped with quantitative bipolar argumentation frameworks and gradual semantics. [29,59,60] propose an interactive clinical decision support system, called CONSULT, for multimorbidity patients to self-manage their treatment. The system integrates four types of data sources; the patient's electronic health record, data from sensors monitoring the patient's symptoms, the clinician's input and finally treatment guidelines. It uses computational argumentation to aggregate the data and resolve inconsistencies in the data sources. It also provides an argumentation-based dialogue interface for system-patient interaction to interactively deliver the recommendations to the user. The system-patient interaction uses templates for its text-based natural language interface. It is based on three different argumentation schemes and their associated critical questions. These cover deliberation, persuasion and explanation dialogues. Castagna et al. [13] propose an explanation dialogue between two participants that also interfaces with the CONSULT system through a chatbot. They propose an argument scheme based on practical reasoning and use it for the explanatory dialogue. Shaheen et al. [63] propose an explanatory dialogue between two participants to explain the recommended treatment plan for multimorbid patients. All of these works focus on explanation dialogues between two participants, mainly a system as an explainer and a human as an explainee. In contrast to these works, EDG aims to use different types of explanatory dialogue forces to generate richer traces during inter-agent reasoning processes.

Pancho et al. [41,71] propose an argumentation-based *deliberation* dialogue between two agents to discuss the viability of transplant organs. The dialogue is implemented as part of *Carrel+*, a health information system to manage organ transplants in Spain. The dialogue model, called *ProCLAIM*, [70] is based on argument schemes and case-based reasoning. ProCLAIM employs a mediator agent to guide the participant agents on their legal moves, decide the validity of submitted arguments and finalise the recommendation regarding the viability of the proposed transplant organ. The mediator agent uses argument schemes, existing guidelines, case-based reasoning and a component manager to manage the strengths of submitted arguments. It applies abstract argumentation semantics [18] to decide the winning argument.

Xiao et al. [81] present a group decision description language and a consensus protocol for a multiagent system. However, the protocol is not based on speech act theory [62], but is an agent communication protocol that uses functions like averaging and intersection to generate consensus values unlike the work presented here. Patkar et al. [49] developed a clinical decision support tool, called *MATE (Multidisciplinary meeting Assistant and Treatment sElector)*, to support multidisciplinary cancer conferences. The tool is responsible for information management and providing treatment recommendations after processing the data. It does not present a dialogue game to support multidisciplinary discussion amongst experts as is presented here. A comprehensive survey describing the use computational argumentation for explainable artificial intelligence can be found in Vassiliades et al. [74]. Some other works have proposed computational argumentation systems for clinical decision support [15,20]. [] present a negotiation protocol for agents in a Belief Desire Intention (BDI) architecture. However, the protocol is grounded in agent communication language rather than dialogue game theory.

The term Health Information Technology (HIT) refers to the application of information technology to facilitate healthcare. HIT systems fall on a wide spectrum ranging from administrative support, patient information management and retrieval, communication and decision support [14]. Carayon et al. [12] point out that most of the existing HIT systems are focused towards individual tasks rather than teams, even as team-based care is becoming a popular paradigm. Here we review representative HIT applications targeting multidisciplinary communication and collaboration support.

Care Connector [46,68,69] is a communication and collaboration platform implemented in a community hospital in Canada. It is a web application that integrates into the HIT of the hospital to retrieve and update electronic records. The application covers both care planning and monitoring modules in addition to a messaging module between multidisciplinary care providers. The patient information is stored as part of a *Care Planner* module. The messaging modules provides asynchronous communication of non-urgent messages using the information in the Care Planner as a shared knowledge base. The messaging module allows linking each conversation to a patient. It informs participants to post messages following the Situation, Background, Assessment and Recommendation (SBAR) framework which is used in healthcare communication. However, it does not force the participants to frame their messages according to this framework.

Kurahashi et al. [30] present another communication and collaboration tool called *Loop*. It allows multidisciplinary collaboration between teams. The teams can include healthcare professionals, caregivers as well as the patient. Each conversation loop is centred around a patient. The application includes a card with patient information on the left hand side while the right hand side has the messaging thread. The information exchange is secure and sharing information between different subgroups is allowed. For example, a professional only message exchange or with all the participants in the loop. It also allows tagging messages with user-defined labels to facilitate search later on. The labels represent different themes or *issues* described in the message.

[32] implemented a platform, called *one-stop platform*, for multidisciplinary collaboration among healthcare professionals in a Taiwanese hospital. The platform integrates into existing HIT system of the hospital. It covers administrative and planning aspects in addition to a messaging module. The messaging modules allows transparency and accountability for message posting and viewing. It supports exchanging text, audio and video messages. However, there is no specification on the format of the content that is exchanged.

Shared Care Platform (SCP) [38] is yet another collaboration tool implemented in a hospital in Spain that builds on social networking and open source tools. The tool is targeted towards facilitating healthcare professionals to manage multimorbidity patients. It has two components; a social networking component, called the *Clinical Wall*, and a decision support component. The Clinical Wall provides social networking like collaboration and communication support amongst healthcare professionals. It is integrated into the electronic health records of the patient. The record has an assessment section, a discussion section and a conclusion section. The assessment section includes information on patient history and assessments. The conversation starts with one clinicians posing a question to others. In the discussion section the clinicians exchange messages to arrive at an agreement with regards to the question. In the conclusion section all participants need to sign off on the agreed decision. During the discussion any clinician can be added to the conversation to invite their feedback. The decision support component uses the Clinical Wall and provides clinical guidelines in the form of rules. Other works [42,45,80,82] implement mobile applications to support care and communication amongst healthcare professionals, caregivers and patients. These applications mainly support administrative and information management tasks with simple messaging support for communication.

All these platforms and applications provide secure messaging amongst participants and well integrated interfaces for the existing HIT systems in place. In contrast, the prototype implementation provided for EDG does not provide any of these features since the goal in this work was to evaluate the underlying protocol rather than present a full-fledged web application. However, none of the existing applications provide support for framing the content and type of messages with an underlying dialogue protocol as is proposed in this work. So the platform provides a novel idea of supporting collaborative communications amongst healthcare experts based on an underlying dialogue protocol.

3. Requirements for expert collaboration

In this section, we propose a requirement specification for successful consultations among experts. Consultations among experts are common in the professional world, especially when critical decisions are concerned such as in medicine, aviation and engineering. We focus on consultations among medical experts as our domain of choice in order to develop a dialogue protocol for consultations among expert agents. This is because it is easier to abstract away from domain specific terminology in this case in order to understand the interaction dynamics. However, the requirement specification we present is abstract enough to be applied to consultations among experts in general since it avoids domain specific scenarios and terminology.

In order to understand collaboration scenarios between experts, we held informal discussion with some medical experts (specifically a gynaecologist, a radiologist, a general physician and a dentist). We identified two main scenarios, informal consultations such as during hand-off of a patient from an emergency room to a general or specialist ward, and formal consultations which usually take the form of *multidisciplinary cancer conferences* or *case conferences* for short. These conferences are structured discussions between different specialists to finalise diagnosis and treatment options for cancer patients [78]. The conference is attended by multiple specialists such as surgery, oncology and pathology. It starts off with the specialist in charge presenting each case history to the panel of experts. This is followed by a discussion amongst experts as to the best possible diagnosis and treatment options for each patient [21]. During the discussion, knowledge transfer between experts takes place. This happens through the *explanatory, inquisitive* and *cooperative* tone of the dialogue. Because of its explanatory value, many specialists consider it to have educational value for trainees [21]. We chose the case conference as our main use case to inform the protocol because of its formal and structured format. Moreover, some of the general communication issues [33,66,72] during informal hand-off also come up in case conferences. Subsequently, we reviewed medical literature on communication in case conferences to identify possible issues. We also included some works on general communication issues during informal collaboration between medical experts. These works were included since they were general enough to be understood by non-medical audiences.

Sutcliffe et al. [66] identify two types of communication failures in the medical profession: *systematic* and *individual*. While systematic failures result from a lack of sufficient organisation, individual failures have complex roots such as hierarchical and power dynamics and excessive workload. They suggest establishing communication guidelines to minimise both types of failures. In order to mitigate against these failures, we develop an interaction protocol grounded in the interaction dynamics during case conferences [16] as well as general communication dynamics [66] that can come up during informal collaboration between medical experts as a result of organisational subculture [66]. Formally, we used Scopus, PubMed and GoogleScholar to look for papers from the medical community that identify

Table 2

Requirements for effective communication between experts according to medical literature

Id	Requirement	#Papers (total: 14)	References
	Agent Oriented Requirement, *RA*		
RA1	To minimise the effects of individual constraints in communication.	7	[16,28,31,33,43,66,72]
	Cooperation Oriented Requirements, *RC*		
RC1	To enable and promote cooperation.	7	[16,28,31,33,34,57,66]
RC2	To provide quality control for the recommendations.	9	[16,28,31,34,43,57,66,78,79]
RC3	To allow detailed and open discussion.	11	[16,28,31,33,43,55,57,61,66,78,79]
RC4	To allow knowledge transfer between experts.	12	[16,28,31,33,34,55,57,61,66,72,78,79]
	Protocol Oriented Requirements, *RP*		
RP1	To enable communication of patient history.	8	[28,31,35,57,61,66,72,78]
RP2	To allow communication of critical points.	9	[31,34,55,57,61,66,72,78,79]
RP3	To allow explanations and clarifications in the discussion.	6	[16,28,33,55,66,72]
RP4	To provide mechanism for resolving disagreements.	5	[16,28,31,33,66]
RP5	To promote equal participation from all participants.	6	[16,28,31,33,66,78]
RP6	To allow equal distribution of illocutionary force among participants.	4	[16,31,33,66]
	Implementation Oriented Requirements, *RI*		
RI1	To have a coordinator for the dialogue.	9	[31,33–35,43,55,72,78,79]
RI2	To record the dialogue history and conclusion.	6	[31,35,43,61,78,79]
RI3	To protect patient privacy.	3	[43,57,78]

communication issues in cancer conferences and in general. We used the keywords 'communication multidisciplinary cancer conference', 'multidisciplinary cancer conference' 'communication failure medical experts' and 'tumour board decision making'. All open-access, English language articles between 2001 and June 2021 related to medicine were considered. Amongst these, manual filtering was done to narrow down results to works involving reflections on communication issues amongst medical experts in cancer conferences and in general. The included articles were either reporting reflections from user studies [16,31,33,34,55,61,66,72], surveys [35,57,78,79] or best practices [43,55,78] followed by professionals. Works related to communication between patients and healthcare professionals were excluded. As were works that focus on the diagnostic recommendations for different medical conditions. The articles included in the study are given in Table 2. We stopped our search for articles when the same ideas started to recur in different articles and we felt confident that new articles were not adding any new perspectives.

We identified fourteen basic requirements for consultations among medical experts which are presented in Table 2. These cover both the *systematic* and *individual* needs for effective consultations between experts. A requirement was considered as inferred from a publication if it was explicitly or implicitly mentioned as a standard practice, a desired outcome or as a lack thereof. All the best practices, guidelines, reflections in the papers were taken into account, grouped together and summarised. This robust process of systematic and rigorous data collection from the domain literature is treated as providing validation of requirements which are then further evaluated in the user study on requirements embedded in the dialogue protocol (see Section 6). These requirements adhere to the structural guidelines and best practices for case conferences while at the same time addressing the communication issues that come up during informal consultations. They are abstract enough to be applicable in a formal collaboration setting between experts in different domains. They can be seen as sub-goals that can facilitate the collaboration in order for it to be productive.

The requirements were then categorised into four classes depending on the mechanism through which they can be satisfied. Table 2 lists the requirements according to their proposed categorisation. Each row presents the requirement id, description, number of papers in the literature that mentioned this requirement and references to the corresponding works. Each of the categories and their corresponding requirements are described next.

3.1. Agent oriented requirements

This category represents communication requirements that are directly related to the dialogue participant. This category has only one requirement, labelled as *RA*1. It reflects that communication among different experts fares better in cases where the participants show strong communication skills such as self-confidence, assertiveness, amiability and politeness. This is especially true where there is an organisational hierarchy amongst the participants. Consequently, a collaboration framework should ideally try to offset the communication weaknesses of the participants as much as possible.

3.2. Cooperation oriented requirements

The cooperation oriented requirements, with identifiers RC1–RC4, stress different aspects of cooperation during the collaborative dialogue. *RC*1 expresses cooperation as a general goal to be fulfilled during the dialogue. *RC*2 outlines the goal for the cooperation itself: to achieve quality control over the recommendations. Achieving this quality control through consensus requires open and frank discussion amongst the participants. This is expressed by *RC*3. Finally, a fundamental aspect of cooperation is the transfer of knowledge between the participants. This is captured by *RC*4.

3.3. Protocol oriented requirements

Protocol oriented requirements cover communication and logistic aspects that should be handled at the protocol definition level. Six such requirements were identified. These are given identifiers RP1–RP6. RP1 ensures that patient history (or observations pertaining to the issue at hand in case of non-medical domains) is explicitly stated during the dialogue so that any faulty assumptions can be countered. RP2 brings critical concerns of the participants to the forefront of the collaboration process. By doing this, it ensures that the participants reflect on these issues. Explanations and clarifications can be useful tools for transfer of understanding amongst the participants. This can promote cooperation and help to align the knowledge and thinking of the participants. Hence, RP3 formalises this need and makes it part of the dialogue. Similarly, RP4 ensures that the protocol design incorporates a conflict resolution mechanism. Finally, RP5 and RP6 mitigate against possible power dynamics resulting from the organisational structure that might influence the dialogue participants. RP5 ensures that the protocol design incorporates inclusiveness while RP6 incorporates equality into the design.

3.4. Implementation oriented requirements

Implementation oriented requirements express logistic concerns that can only be addressed at the implementation level. There are three such requirements which are given identifiers from RI1–RI3. Since a collaborative dialogue between more than two experts can entail administrative overhead, most studies [31,34,79] found that having a designated role to oversee this greatly improves the collaboration process. Hence, this is captured by RI1. RI2 captures the necessity of recording the dialogue history so

that it can be referred back to at a later time if required. Finally, since expert collaborations generally cover confidential topics and data, RI3 ensures that any confidential information exchanged during the collaboration is protected.

We consider all the requirements to form a core part of discussions although some seem to come up more frequently in literature as compared to others. For example, requirements RC4, RP2, RI1, RC36 and RC2 are more frequently mentioned while some others such as RP6 and RI3 are not. Nevertheless these represent fundamental aspects of these exchanges.

4. A formal dialogue system for expert collaboration

This section formally presents the Experts' Dialogue Game (EDG) which embeds explanation-based illocutionary forces in an inquiry dialogue in order to emulate the inquisitive, explanatory and cooperative aspects of real-life consultations. This is done so that the dialogue game can generate richer reasoning traces and meet the needs of successful collaboration amongst experts. An *Inquiry Dialogue* is defined in Walton and Krabbe's popular typology of dialogue types [77] as a collaborative discussion amongst participants to find out the answer to one or more questions when none of them is presumed to know the correct answer beforehand. Later, Walton [76] introduces an *Explanatory Dialogue* as a discussion between two or more participants in order to bring about a transfer of understanding from one to another. In this case, the participants already agree on the topic but differ in their understanding of it.

Definition 1. A *Dialogue Game (DG)* is a tuple $DG = (X, L, Loc, R)$ where X is the set of participating agents, L is a logical language which represents the dialogue content, *Loc* is the set of permitted locutions and $R = CM \cup CB \cup TM \cup CS \cup T \cup PL$ represents the sets of rules for *commencement*, *combination*, *termination*, *commitment*, *turn-taking* and *politeness* respectively.

Each of these elements is described next.

4.1. Participants, X

The game requires two or more participating agents, each representing an expert in some area, belonging to the set $X = 1, 2, \ldots n$ where n is the total number of agents in the system.

4.2. Content language, L

Each agent has its own private knowledge base, represented as Σ_i where $i \in X$. The knowledge base is expressed in the content language L. Table 3 shows the initial knowledge bases of the agents for the running example in the content language L. It is a first order logic language with the following components:

- Let $H = \{h_1, h_2, \ldots, h_p\}$ be the set of all possible observations recorded in a dataset where each h_i represents a feature. The value for each feature belongs to the set $V = \{v | v$ is a categorical or non-categorical value$\}$. For the running example, this would be patient history recorded in a dataset and represented as atomic predicates and terms in first order logic such as *age*(48), *gender*(*female*), *symptom*(*fatigue*), *symptom*(*constipation*), *increase*(*weight*), *increase*(*sleep*) and *skin*(*dry*).

Table 3

Initial knowledge bases of agents α, β, γ for the example dialogue

Σ_α	Σ_β	Σ_γ
$h_1 - h_7$		
$d_1 - d_4$	$d_1 - d_4$	$d_1 - d_4$
$e_1 - e_4$	$e_1 - e_4$	$e_1 - e_4$
k_1, k_2	k_1, k_2	k_1, k_2
c_1, r_1, r_2		
$f_1, f_5, f_6, f_8, f_{12}$	$f_3, f_5, f_6, f_8, f_{12}$	$f_2, f_4 - f_{12}$

Key to formulas in the knowledge base.

h_1 *age*(48)

h_2 *gender*(*female*)

h_3 *symptom*(*fatigue*)

h_4 *symptom*(*constipation*)

h_5 *increase*(*weight*)

h_6 *increase*(*sleep*)

h_7 *skin*(*dry*)

r_1 *walk*

r_2 *healthy_diet*

d_1 *diagnosis*(*depression*)

d_2 *diagnosis*(*anaemia*)

d_3 *diagnosis*(*hypothyroidism*)

d_4 *diagnosis*(*hyperthyroidism*)

e_1 *test*(*TSH*)

e_2 *test*(*T4*)

e_3 *test*(*T3*)

e_4 *test*(*blood_complete_count*)

k_1 *symptom*(*headache*)

k_2 *symptom*(*backpain*)

c_1 *family_history*(*diagnosis*(*autoimmune_disorder*))

f_1 *symptom*(*fatigue*) $\land$ *increase*(*weight*) $\land$ *increase*(*sleep*) $\rightarrow$ *diagnosis*(*depression*)

f_2 *symptom*(*headache*) $\land$ *symptom*(*backpain*) $\rightarrow$ *diagnosis*(*depression*)

f_3 *symptom*(*fatigue*) $\land$ *increase*(*weight*) $\land$ *increase*(*sleep*) $\land$ *symptom*(*cold_hands*) $\rightarrow$ *diagnosis*(*anaemia*)

f_4 *symptom*(*fatigue*) $\land$ *increase*(*weight*) $\land$ *increase*(*sleep*) $\land$ *symptom*(*constipation*) $\land$ *skin*(*dry*) $\rightarrow$ *diagnosis*(*hypothyroidism*)

f_5 *confirm*(*diagnosis*(*anaemia*)) $\rightarrow$ *test*(*blood_complete_picture*)

f_6 *confirm*(*diagnosis*(*hypothyroidism*)) $\rightarrow$ *test*(*TSH*) $\land$ *test*(*T4*)

f_7 *confirm*(*diagnosis*(*subclinical_hypothyroidism*)) $\rightarrow$ *test*(*TSH*) $\land$ *test*(*T4*) $\land$ *test*(*T3*)

f_8 *TSH*(*high*) $\land$ *T4*(*low*) $\rightarrow$ *diagnosis*(*hypothyroidism*)

f_9 *TSH*(*high*) $\land$ *T4*(*normal*) $\land$ *T3*(*normal*) $\rightarrow$ *diagnosis*(*subclinical_hypothyroidism*)

f_{10} *skin*(*dry*) $\land$ *family_history*(*diagnosis*(*autoimmune_disorder*)) $\rightarrow$ *diagnosis*(*hypothyroidism*)

f_{11} $\rightharpoondown$ *symptom*(*fatigue*) $\land$ $\rightharpoondown$ *increase*(*weight*) $\land$ $\rightharpoondown$ *increase*(*sleep*) $\land$ $\rightharpoondown$ *symptom*(*constipation*) $\land$ $\rightharpoondown$ *skin*(*dry*) $\rightarrow$ *diagnosis*(*subclinical_hypothyroidism*)

f_{12} *confirm*(*diagnosis*(*hyperthyroidism*)) $\rightarrow$ *test*(*TSH*) $\land$ *test*(*T4*) $\land$ *test*(*T3*)

- Let $D = \{d_i, \ldots, d_m\}$ be the set of all verdicts. For the running example, this would include *diagnosis*(*depression*), *diagnosis*(*anaemia*) and *diagnosis*(*hypothyroidism*).
- Let $E = \{e_i, \ldots, e_w\}$ be the set of all evaluative measures that can be recommended for a particular case. For the running example, these could be the tests identified by the medical experts to get more

data on the patient's condition such as blood complete picture *test*(*blood_CP*), *test*(*TSH*)., *test*(*T*4) and *test*(*T*3) levels.

- Let $T = \{t_1, \ldots, t_y\}$ be the set of all possible remedial measures that can be taken. For the running example, this could be the drugs prescribed or recommendations for behaviour change for the patient such as *prescribe*(*idoine*) and *walk_steps*(10000).
- Let $C = \{c_1, \ldots, c_x\}$ be the set of all concerns/critical points that can be raised by an agent for a particular case. For the running example, this could be points of concern that the medical experts identify for a particular patient such as *family_history*(*diagnosis*(*autoimmune_disorder*)). Then $A = E \cup T$ will be the set of all recommendations that can be made for a single case and $O = D \cup A \cup C$ will be set of all verdicts, recommendations and concerns that can be discussed during the dialogue game.
- Let $K = \{k_1, \ldots, k_z\}$ be the set of all atomic facts in the domain knowledge such that $\{H, A, D\} \subset K$. For the running example, these would be *symptom*(*headache*) and *symptom*(*backpain*).
- Let $F = \{f_1, \ldots, f_v\}$ be the set of all inferences from elements of H, K and O that make up the domain knowledge. For the running example, this could be *symptom*(*fatigue*) $\wedge$ *increase*(*weight*) $\wedge$ *increase*(*sleep*) $\wedge$ *symptom*(*constipation*) $\wedge$ *skin*(*dry*) $\rightarrow$ *diagnosis*(*hypothyroidism*).

4.3. Locutions, Loc

Each locution, represented by the letter τ, is of the form $\tau = loc_i(p)$ where *loc* defines the speech act performed, $i \in X$ is the agent uttering the locution and $p \in L$ represents the content of the locution except for the *prompt* locution where $p = \tau' \in Loc$ represents a previous locution. All participants are assumed to be the receivers of all the messages so the receiver id is not tracked. The set of permitted locutions is given in column 2 of Table 4. The locutions can be grouped into four disjoint subsets such that $Loc = L1 \cup L2 \cup L3 \cup L4$ which cover different aspects of the collaborative dialogue. Respective classes are indicated in column 1 of Table 4. The characteristics of each subset are discussed next.

L1 (informational locutions). There are five locutions in this subset: *observation, verdict, advise, concern* and *assert*. These are labelled from L1.1–L1.5 respectively. In the commencement phase, L1.1–L1.4 are used to set the context of the dialogue. In the progress stage, all five can be used to introduce new knowledge into the conversation. While L1.1 to L1.4 are used to introduce facts pertaining to a specified topic, L1.5 (assert) is used for introducing inference rules that relate the content of the first four locutions to each other. No distinction is made between strict and defeasible facts and rules.

L2 (requests). We use a simplified version of the typology for different explanation requests (and replies) presented by [11]. This is because while they meet the conversational needs for a specific scenario in the financial domain, our protocol targets a general consultation setting between experts without going into domain specific details. Consequently, three types of requests are included. A request for explanation, represented by *wh-explain*, when the claim is agreed upon but one participant requires the other to provide more formal details or give an informal opinion. A request for justification, represented by *wh-justify*, when one participant needs the other to back up their standpoint. Finally, a clarification request, represented by *wh-clarify* when the participants agree about the claim but one of them has missing links in the reasoning process and so asks for this missing information. All three types of requests are assumed to be as generic as possible and cover not only the *why* aspect but also the *what*. Hence they are framed as *wh-requests*. These locutions embed explanatory illocutionary forces in the main inquiry dialogue, allowing for the generation of richer traces of the inquiry process. This is critical for making the result of the inquiry dialogue explainable.

Table 4

Combination rules for consultation between two expert agents $\alpha, \beta \in X$

Id	Locution	Reply
L1.1	$observation_\alpha(H_i)$ where $H_i \subset H$	$agree_b(H_i)$, $observation_\beta(H_j)$, $assert_\beta(F_k)$, $wh\text{-}clarify_\beta(H_k)$ where $H_i \subset H$, $H_k \subset H_i$, $H_j \in H$ and $F_k \subset F$. H_i, H_k represent observations exchanged so far about the case while H_j represents any new facts related to the case and $O_m \subset O$ represent any inference rules that apply to the observation respectively.
L1.2	$verdict_\alpha(D_i)$ where $D_i \subset D$	$agree_\beta(D_k)$, $wh\text{-}explain_\beta(D_k)$, $wh\text{-}justify_\beta(D_k)$, $verdict_\beta(D_j)$ where $D_k \subset D_i$ and $D_i, D_j \subset D$.
L1.3	$advise_\alpha(A_i)$ where $A_i \subset A$	$agree_\beta(A_k)$, $wh\text{-}explain_\beta(A_k)$, $wh\text{-}justify_\beta(A_k)$, $wh\text{-}clarify_\beta(F_j)$, $advise_\beta(A_j)$ where $A_i, A_j \subset A$, $A_k \subset A_i$ and F_j is property of $a_j \in A_i$.
L1.4	$concern_\alpha(C_i)$ where $C_i \subset C$	$agree_\beta(C_k)$, $wh\text{-}justify_\beta(C_k)$, $wh\text{-}explain_\beta(C_k)$, $wh\text{-}clarify_\beta(f_i)$ where $C_i \subset A$, $C_k \subset C_i$ and f_i is property of $c_i \in C_i$.
L1.5	$assert_\alpha(F_i)$ where $F_i \subset F$	$agree_\beta(F_i)$, $assert_\beta(F_j)$ where $j \neq i$ and $F_i, F_k \subset F$.
L2.1	$wh\text{-}explain_\alpha(\theta)$ where $\theta \in H \cup O$	$explain_\beta(\phi)$ where $\phi \in K \cup F$.
L2.2	$wh\text{-}justify_\alpha(\theta)$ where $\theta \in O$	$justify_\beta(\phi)$, $retract_\beta(\theta)$ where $\theta \in O$ and $\phi \in F$.
L2.3	$wh\text{-}clarify_\alpha(\theta)$ where $\theta \in H \cup O$	$clarify_\beta(\phi)$ where $\theta \in H \cup O$ and $\phi \in H \cup F$.
L3.1	$explain_\alpha(\theta)$ where $\theta \in K \cup F$	$agree_\beta(\theta)$, $assert_\beta(\psi)$, $wh\text{-}clarify_\beta(\theta_i)$, $explain_\gamma(\theta_k)$ where $\theta, \theta_k \in K \cup F$, $\psi \in F$, $\theta_i \in H \cup O$ such that θ_i is related to θ and $c \neq b$.
L3.2	$justify_\alpha(\theta)$ where $\theta \in F$	$agree_\beta(\theta)$, $assert_\beta(\psi)$, $wh\text{-}explain_\beta(\theta_i)$, $wh\text{-}clarify_\beta(\theta_j)$, $justify_\gamma(\theta_k)$ where $\theta, \theta_k \psi \in F$ such that $\theta \neq \psi \neq \theta_k$ and $\theta_i, \theta_j \in H \cup O$ such that θ_i, θ_j are related to θ and $\theta_i \neq \theta_j$.
L3.3	$clarify_\alpha(\theta)$ where $\theta \in H \cup F$	$agree_\beta(\theta)$, $assert_\beta(\psi)$, $wh\text{-}explain_\beta(\theta_i)$, $wh\text{-}justify_\beta(\theta_j)$, $clarify_\gamma(\theta_k)$ where $\theta, \theta_k \in H \cup F$, $\theta \neq \theta_k$, $\psi \in F$, $\theta_i \in K \cup F$ and $\theta_j \in F$ such that θ_j, θ_j are related to θ, $\theta_i \neq \theta_j$.
L3.4	$agree_\alpha(\theta)$ where $\theta \in K \cup F \cup O \cup H$.	-
L3.5	$retract_\alpha(\theta)$ where $\theta \in O$.	-
L4.1	$prompt_\alpha(Loc_k)$ where $Loc_k \subset Loc$ is the set of locutions moved so far.	Any of the valid responses entailed by each element of Loc_k.
L4.2	end_α	end_β, $prompt_\beta(Loc_k)$ where $Loc_k \subset Loc$.
L4.3	$pass_\alpha$	-

L3 (replies). This subset has five locutions *explain, justify, clarify, agree* and *retract* which are given identifiers from L3.1–L3.5 respectively. The first three are locutions for answering the corresponding *wh-requests* from the L2 subset while the last two cover other possible answers such as *agree* and *retract*. The protocol assumes that the agents are always able to clarify and explain, but not always able to justify, in which case they retract.

L4 (management locutions). This subset defines a total of three locutions, which are given identifiers from L4.1–L4.3. These are *prompt, end* and *pass*. These manage the dialogue in different ways. *prompt* serves two purposes: it allows the speaker to indicate to the other participants that they are awaiting a response on a particular locution and it can also be used during the termination stage to justify why a participant has disagreed to end the dialogue. *end* indicates an acknowledgement by a participant that they are satisfied with the dialogue outcome, thus, giving their consent to end the dialogue. If they have an outstanding issue, they can refuse to give their consent to end the dialogue. In this case, they are invited to justify this by using *prompt* to let other participants know which of their statements have not received a response yet. Finally, since the dialogue game allows participants to make multiple moves, *pass* is used to manage turn-taking. Whenever a participant has finished whatever they wanted to say (they are allowed to use multiple locutions) in their turn, they signal the end of their turn by using *pass*.

Thus, in the case of more than two agents, the protocol allows everyone to participate in the explanatory dialogue since the dialogue game allows using multiple responses to each locution (a detailed description is provided in Section 4.4 when the dialogue rules are introduced). This means that in response to locutions from subset L2, other agents who were not directly addressed in the preceding *wh-request* can choose to participate in the information exchange by making an appropriate move.

4.4. Rules, R

The game has three stages: an opening stage governed by *Commencement Rules*, a progress stage governed by *Combination Rules* and a termination stage described by *Termination Rules* [52]. Each of these are described next, followed by commitment, turn-taking and politeness rules.

Commencement rules. The topic of the dialogue game can be one or more subsets of O. The game always starts with the initiator agent presenting the facts of the case (observations), its own conclusions (verdicts), corresponding recommendations (advice) and any critical points (concerns) it deems important. Thus, the first turn is composed of the first four locutions from the locution subset L1. A move, represented by μ, is a tuple $\mu = \langle \tau, \tau' \rangle$ where τ is the new locution being introduced in the current turn while τ' is a previously introduced locution. τ' is null for all moves made in the commencement phase. Formally, entries L1.1 to L1.4 in the *Locution* column of Table 4 formally present the four *Informational* locutions that are part of the first turn. Strict ordering is enforced on the four locutions that make up the first move and is given by the sequence L1.1 to L1.4 in the table.

Combination rules. The protocol allows participants to start new threads in the conversation at any time. This is achieved by allowing one or more locutions in the same turn where each locution corresponds to a move. For a move which uses the *prompt* locution, τ' is null. Repeating the same move is not allowed, however, repeating the same locution in the same turn with different content is allowed. Hence, the next participant can respond to any number of locutions from the previous turns. In doing so, they have to specify the locution they are responding to (τ') and pick any of the valid responses for that locution as defined in Table 4 where the *Reply* column indicates possible reactions to each locution from the *Locution* column. Subscripts α and β identify the agent playing the move.

Termination rules. The dialogue terminates when all the participants agree to end it. Any participant can start the process for getting consent from others to end the dialogue. They can do this by using the *end* locution. This signals the start of the termination stage. Since the participants are assumed to be assertive and cooperative, this means that anything that the participants do not explicitly challenge is taken for granted as an agreement. Hence, when the dialogue ends, all the participants are assumed to have agreed on all the elements of set O under discussion. However, each voting for termination may not always end in successful termination since any participant can refuse to give their consent. They are then invited to highlight any outstanding issues they have by using the *prompt* locution as explained in Section 4.3. If this happens, the dialogue moves back into the progress stage. Otherwise, they give their consent to end the dialogue (and to accept all the statements that went unchallenged by them) by using the *end* locution. A move which uses the *end* locution also has τ' set to null.

Commitment rules. Dialogue games generally require each participant have their commitments publicly available in the form of a *Commitment Store*. The commitments are created as a result of particular speech acts and they ensure accountability for the participants. This is useful for making the dialogue coherent and productive. We follow Hamblin's notion of commitment stores as done by [48] where

an agent's commitment store, $CS(i)$ for agent $i \in X$, corresponds to a publicly available subset of its original knowledge base Σ_i. Specifically the commitment store $CS^t(i)$ at any time interval t for agent $i \in X$ contains elements of $H_i \cup O_i \cup F_i$, where $H_i \subset H$, $O_i \subset O$ and $F_i \subset F$. These represent the observations, verdicts, advice and concerns known by each agent at any time during the game such that $CS^{t+1}(i) = CS^t(i) + \{c | c \text{ is a new commitment}\}$. In addition we introduce a multilateral agreement store for all agents, $AS(MAS) = \{c | \exists i, j \in X \text{ such that } c \in CS(i) \cap CS(j) \text{ and } c \in K \cup F \cup O \cup H\}$. That is, it contains the observations, verdicts, advice and concerns that have had a multilateral agreement at any time during the dialogue. As in the case of [48], the union of all individual commitment stores reflect the dialogue state any time whereas $AS(MAS)$ shows the global agreements rather than the information state of the dialogue. $AS(MAS)$ can be considered as a collective agreement store which provides a summary of the agreements during the dialogue to all participants. This is because of the assumption that the agents are assertive and therefore, commit to any statement that they do not explicitly challenge. Locution L3.4 (*agree*) helps to incrementally build up this summary, making it easier to synchronise the collective commitments of all agents. The mechanism of how this works is explained in the commitment rules that follow. Each rule describes the changes to $CS(i)$ for agent $i \in X$ and $AS(MAS)$ in reaction to each locution.

- *C1*: For locutions subset L1 and L3.1 to L3.3, the content of the locution is added only to the individual commitment store of the speaker, $CS(i)$.
- *C2*: For locution L3.4, the content of the locution is added to the individual commitment store of the speaker agent and also to $AS(MAS)$. If the content of L3.4 is already added to $AS(MAS)$ as a result of a previous application of C2, it is not added again.
- *C3*: For a location $loc(\text{arg}) \in L2$ where arg $\in L$, no changes are made to the individual commitment store of the speaker. However, if arg had been added to $AS(MAS)$ following C2 earlier in the dialogue, it is removed from $AS(MAS)$ since it only contains multilateral agreements rather than bilateral ones.
- *C4*: For all locutions belonging to L4, no changes are made to neither the individual commitment store of the speaker nor $AS(MAS)$.
- *C5*: For L3.5, which represents a *retract*, the content of the locution is removed from the commitment store of the speaker and from $AS(MAS)$ if it has been added following rule C2.
- *C6*: After the dialogue terminates, the union of all individual commitment stores minus the conflicts is added to $AS(MAS)$ following the notion of implicit agreement to all unchallenged statements as described in *Termination Rules*.

Table 5 exemplifies the commencement, combination, commitment and termination rules for the running example. The first column of Table 5 indicates the turn identifier, the second column lists the identifier of the agent making the move, the third column identifies the locutions moved, the fourth column shows the changes in the commitment store of the speaker and the last column indicates the changes in the multilateral agreement store. In T_1, the first agent sets the context of the dialogue in accordance with the commencement rules. In T_2, agent β asks α to justify their diagnosis of depression since its inference rule f_3 is in conflict with the diagnosis made by α. In T_4, α provides this justification. However, in T_3, γ provides its own diagnosis and challenges α's justification in T_5. Similarly, β then asks γ to justify their diagnosis of hypothyroidism in T_6 because it is in conflict with their inference rule f_3. γ's justification is accepted by the other two agents in T_8 and T_9 respectively. In T_13, γ asks α to explain why it recommends testing T3 levels. This is because it already has inference rule f_12 in its knowledge base but is confused with α's reasoning because of inference rules f_7 and f_{11}. Figure 2 shows how the dialogue

Table 5

Example of a dialogue game between three medical expert agents, $\alpha, \beta, \gamma \in X$[1]

Id	Speaker	Sequence of locutions	CS(*Speaker*)	AS(*MAS*)
T_1	α	*observation*($h_1 - h_7$) *verdict*(d_1) *advise*(r_1, r_2) *concern*(c_1) *pass*	$h_1 - h_7$ d_1 r_1, r_2 c_1	
T_2	β	$\langle$*wh-justify*(d_1), *verdict*(d_1)$\rangle$ *pass*		
T_3	γ	$\langle$*verdict*(d_3), *verdict*(d_1)$\rangle$ *pass*	d_3	
T_4	α	$\langle$*justify*(f_1), *wh-justify*(d_1)$\rangle$ *pass*	$CS^{T_1}(\alpha) \cup f_1$	
T_5	γ	$\langle$*assert*($\rightarrow f_2$), *justify*(f_1)$\rangle$ *pass*	$CS^{T_3}(\gamma) \cup \rightarrow f_2$	
T_6	β	$\langle$*wh-justify*(d_3), *verdict*(d_3)$\rangle$ *pass*		
T_7	γ	$\langle$*justify*(h_7, c_1, f_{10}), *wh-justify*(d_3)$\rangle$ *pass*	$CS^{T_5}(\gamma) \cup \{h_7, c_1, f_{10}\}$	
T_8	α	$\langle$*agree*(h_7, c_1, f_{10}), *justify*(h_7, c_1, f_{10})$\rangle$ $\langle$*advise*(e_1, e_2, e_3), *advise*(r_1, r_2)$\rangle$ *pass*	$CS^{T_4}(\alpha) \cup$ $\{h_7, c_1, f_{10}, e_1, e_2, e_3\}$	$\{h_7, c_1, f_{10}\}$
T_9	β	$\langle$*agree*(e_1, e_2, e_3), *advise*(e_1, e_2, e_3)$\rangle$ $\langle$*advise*(e_4), *advise*(e_1, e_2, e_3)$\rangle$ $\langle$*assert*(f_3), *justify*(h_7, c_1, f_{10})$\rangle$ *pass*	$\{e_1, e_2, e_3, e_4, f_3\}$	$AS^{T_8}(MAS) \cup \{e_1, e_2, e_3\}$
T_{10}	γ	$\langle$*agree*(e_1, e_2, e_4), *advise*(e_1, e_2, e_3)$\rangle$ *end* *pass*	$CS^{T_7}(\gamma) \cup \{e_1, e_2, e_4\}$	$AS^{T_9}(MAS) \cup e_4$
T_{11}	β	$\langle$*end*, *end*$\rangle$ *pass*	No change	
T_{12}	α	$\langle$*prompt*, *advise*(e_1, e_2, e_3)$\rangle$ *pass*	No change	
T_{13}	γ	$\langle$*wh-explain*(e_3), *advise*(e_1, e_2, e_3)$\rangle$ *pass*	No change	$AS^{T_{10}}(MAS) - e_3$
T_{14}	α	$\langle$*explain*(f_{12}), *wh-explain*(e_3)$\rangle$ *pass*	$CS^{T_8}(\alpha) \cup f_{12}$	
T_{15}	β	$\langle$*agree*(f_{12}), *explain*(f_{12})$\rangle$ *pass*	$CS^{T_9}(\beta) \cup f_{12}$	$AS^{T_{13}}(MAS) \cup f_{12}$
T_{16}	γ	$\langle$*assert*(f_4, f_7, f_{11}), *explain*(f_{12})$\rangle$ *pass*	$CS^{T_{10}}(\gamma) \cup \{f_7, f_{11}\}$	
T_{17}	α	$\langle$*agree*(f_4, f_7, f_{11}), *assert*(f_4, f_7, f_{11})$\rangle$ *end* *pass*	$CS^{T_{14}}(\alpha) \cup \{f_4, f_7, f_{11}\}$	$AS^{T_{15}}(MAS) \cup \{f_4, f_7, f_{11}\}$
T_{18}	β	$\langle$*end*, *end*$\rangle$ *pass*		
T_{19}	γ	$\langle$*end*, *end*$\rangle$ *pass*		$AS^{T_{17}}(MAS) \cup$ $\{h_1, h_2, h_3, h_4, h_5, h_6\} \cup$ $\{r_1, r_2, d_3\}$

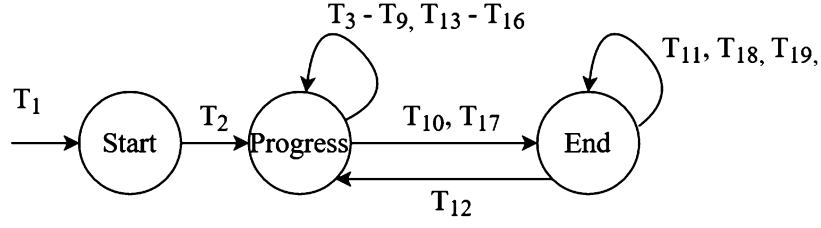[1] Subscripts of locutions are not mentioned for clarity.

Fig. 2. State transitions between commencement, progress and termination states for the example dialogue in table 5.

Table 6

Alternative ending for the running example. α, β, γ

Id	Dialogue
13$'$	γ: I don't think it is necessary to test T3 at this stage since she is not asymptomatic.
14$'$	α: Okay. I think we can close the discussion now.
15$'$	β: I agree.
16$'$	γ: I agree.

Table 7

Dialogue game between three medical expert agents for the alternative ending in Table 6[2]

Id	Speaker	Sequence of locutions	CS(Speaker)	AS(MAS)
$T_{13'}$	γ	$\langle assert(f_7, f_{11}), advise(e_1, e_2, e_3)\rangle$ *pass*	$CS^{T10}(\gamma) \cup \{f_7, f_{11}\}$	
$T_{14'}$	α	$\langle agree(f_7, f_{11}), assert(f_7, f_{11})\rangle$ *end* *pass*	$CS^{T8}(\alpha) \cup \{f_7, f_{11}\}$	$AS^{T10}(MAS) \cup \{f_7, f_{11}\}$
$T_{15'}$	β	$\langle end, end\rangle$ *pass*	No change	
$T_{16'}$	γ	$\langle end, end\rangle$ *pass*	No change	$AS^{T14}(MAS) \cup$ $\{h_1, h_2, h_3, h_4, h_5, h_6\} \cup$ $\{r_1, r_2, d_3\}$

[2] Subscripts of locutions are not mentioned for clarity.

switches between the commencement, progress and termination states as a result of the different turns. Each node in the diagram represents a state with the state label given in the centre. Each arc represents a transition with the arc labels corresponding to the turn Ids in Table 5.

EDG promotes making justifications, explanations and clarifications explicit in the discussion by not allowing *assert* in response to locutions $L1.2$–$L1.4$. Table 6 highlights how this can affect the dialogue. It shows how the running example in Table 1 would change after the twelfth move by α if *assert* was allowed in response to locution $L1.3$. The first column indicates the identifiers of the alternative statements made by the experts. The identifiers for the alternative scenario are appended with a $'$ symbol to indicate that this is the alternative scenario. The second column shows the expert id and the statement they are making. Table 7 shows the corresponding dialogue game. The columns are organised in the same way as for Table 5. In this case, γ does not ask α for an explanation, rather it provides it's own reasoning and the dialogue can close earlier. Hence, it is clear that EDG promotes justifications, explanations and clarifications at the expense of shorter discussion time.

The collective agreement store serves as the output of the multi-agent system. It allows the most relevant knowledge for the decision making to be pooled together in a systematic way which is more

computationally efficient than pooling all the knowledge bases of the agents. In the process, it also preserves the privacy of the agents since only publicly shared information is used. This approach allows for the building of a modular explainable multiagent system in which the multiagent decisions can be made independently of the human-machine interface. For example, it can be used to provide justified decisions made by expert agents to a user using another explanatory protocol for human-machine interaction such as the one proposed by Ilia et al. [65]. In this case, the collective agreement store can serve as the interface between the two modules of the explainable Artificial Intelligence (XAI) system.

EDG relies on the locutions *agree* and *retract* to incrementally synchronise agreements from the agents' individual commitment stores to the collective agreement store. However, if the agents fail to use these markers sufficiently, the burden of synchronising the agreement store will move to the end of the dialogue, stressing computational resources. Hence, the agents need to be made aware that explicit agreements will make the protocol more effective.

Promoting elicitation of justifications, explanations and clarifications allows EDG to keep track of collective agreements and resolve discrepancies in the agreement store. An example of this can be seen by comparing the example in Table 5 with the alternative scenario presented in Table 7. In the first case, the *wh-request* in turn T_{13} results in removing the disputed recommendation of e_3 from the agreement store. In contrast, since γ never makes their stance explicit in the scenario in Table 7, e_3 remains in the collective agreement store when the dialogue ends. So, the assumption that the agents are assertive is very important to ensure the success of EDG. An unassertive agent might end up being committed to beliefs that are not consistent with its knowledge base.

Turn-taking rules. EDG identifies two roles for the participants, *initiator* and *participant*. However, the initiator role ends after the first turn, whereby everyone becomes a participant. The initiator provides sufficient context for the dialogue through the locutions in the first turn. The protocol enforces turns but no particular turn-order is enforced. Each agent has to move at least one locution in response to the proponent's moved locutions. Since multiple locutions are allowed in each turn, each agent has to end his turn with the *pass* locution to mark that he is finished.

Politeness rules. Structurally, dialogue games can allow participants to respond only once to each move *(single-reply)* or offer several responses as well *(multi-reply)*, to use only one locution in each move *(single-move)* or more than one *(multi-move)* and to transfer the turn as soon as some objective condition is met *(immediate-reply)* or later *(non-immediate-reply)* [52]. Based on these definitions, we consider EDG to be *multi-reply*, *multi-move* and *non-immediate-reply*. A brief discussion justifying each of these properties follows next.

Multi-reply. The protocol achieves this in three ways. The first two enable this property for the respondent while the last one enables the speaker to proactively demand an additional response. For the respondent, it allows multiple arguments in one turn by not imposing any restrictions on the number of arguments included as content of each locution. For two, it allows respondents to come back to earlier choice points in the dialogue since it does not impose the restriction on addressing the preceding move. So they can move several arguments referring to different previous moves if desired. For the speaker, it enables them to direct the conversation back to issues that were not addressed to their satisfaction using locution L4.1.

Multi-move. The protocol does not limit the number of locutions that can be moved in one turn by each participant (see Section 4.4). Hence, it is by construction *multi-move*.

Non-immediate-reply. Since the protocol does not enforce an external condition to shift the turn, it allows each agent to complete its move uninterrupted and proactively transfer the turn, it is then *non-immediate-reply*.

All these properties make EDG very flexible and close to natural conversation. However, this flexibility can lead to dialogues that are incoherent or compromise the explanability and cooperative aspects of the dialogue. Hence, it calls for introducing the same mannerisms in place in natural conversations that act to counter these complications in real life conversation. So, EDG introduces two such mannerism into the dialogue as *politeness rules*. It identifies two such rules to ensure dialogue progression and conflict resolution. The first is related to *wh-request*. Since the protocol does not force a participant to respond to only the previous move, the participants can ignore explanation, clarification and justification requests, defeating the explanatory objectives of the dialogue. In order to mitigate this, the first politeness rule requires that all *Wh-Requests* must be responded to by the addressee first before they are allowed to respond to any other locution. Other participants who were not the direct addressee, can respond to a *wh-request* if they choose to do so. This allows the participants to collaboratively build explanations. The second rule concerns *prompt*. A *prompt* serves as a reminder to the participants that this particular agent is awaiting a response for the prompted locution. The second rule gives the receivers of the *prompt*, the flexibility to choose to respond to it immediately or in a later turn.

4.5. Semantics

We take a protocol-oriented view of Agent Communication Language (ACL) semantics [50,52]. In this view, the semantics and use of utterances should be defined at the dialogue level rather that at the level of individual locutions [52]. Pitt and Mamdani [50] distinguish between the *content* and *conversational* states of the dialogue. The former is dependent on the *information state* of the agent. *Information state* of an agent reflects its knowledge base. Semantics at this level define the change in the agent's information state. The latter is determined by the speech acts exchanged earlier in the dialogue, which are referred to as *the conversation state*. The conversation state can be described by the set of possible responses for each speech act. Consequently, the commitment rules described in Section 4.4 form the *content* level semantics for the protocol while the combination rules given in Table 4 define the *conversational* semantics for the dialogue. Since the protocol treats the commitment store as a subset of the agent's knowledge base, the commitment rules express *post-conditions* about the agent's information state as a result of the speech act. Next we describe the *pre-conditions* for making the move.

Pre-conditions for managing information state.

- *P1*. For locutions subsets L1 and L3.1 to L3.3, there are no constraints on the content except that it should be relevant to the dialogue topic.
- *P2*. For locution subsets L2, L3.4 and L4.1, the content of the locutions should already be part of the commitment stores of one of the agents.
- *P3*. For L3.5, the content of the locution should belong to the commitment store of the speaker.
- *P4*. For L4.2 and L4.3 no conditions apply as there is no content.

Pre-conditions for managing conversation state.

- *P5*. For locution subsets L1, L2, L3 and L4.1, those imposed by Table 4.
- *P6*. For L4.2, the agent finds no conflicts or objections in the information state of the dialogue as represented by its own commitment store.

- *P7*. For L4.3, the agent making this move must have used at least one other valid locution before this one.

While pre-conditions can relate to both constraints on the agent's information state [2,48] or to constraints on the conversational state of the dialogue [76], here we specify pre-conditions to manage the information and conversational states of the dialogue itself. The limits introduced on the content of locutions in Table 4 also form part of the pre-conditions for managing the information state of the dialogue. We require that the agents maintain dialogue history and do not repeat a locution with the same content.

5. A platform for expert collaboration

A prototype of EDG was implemented as a web application in order to evaluate it with human experts through a user study. The web application allows the participants to 'chat' while enforcing EDG protocol. However, the participants do not need to remember the protocol, the web application enforces it for them. For each 'message' in the chat, it shows the possible locutions from Table 4 that can be used in response as a drop down menu. The participant can select a locution to frame their response and type their text in the corresponding text field. This section provides details on the implementation and the user study design.

5.1. Implementation

EDG was implemented as a prototype full-stack web application that allows human participants to engage in discussion regarding the best diagnosis and treatment options for a patient. Figure 3 shows a screenshot of graphical user interface of the web application with a hypothetical example. The application was implemented using JavaScript frameworks for client and server. The dialogue history is recorded in an SQLite database on the server. The server keeps track of the number of participants in a game and rotates the turns in a cyclic manner in the order in which the participants join the game session. The application enforces politeness rules by using highlights. It alerts the user who was the target of a *wh-request* by highlighting the request in red and not allowing this user to play any other locution until all the *wh-requests* to them have been discharged. Similarly, if a user plays the *prompt* locution, the target locution is highlighted in blue on all participants' user interface to alert them on the request for response. However, they are not forced to respond to this alert. The web application was used to evaluate the usability of EDG through a user study.

5.2. User study

We conducted a user study to evaluate the *usability* of the platform and its underlying protocol. According to the International Organization for Standardization (9241–11:2018), *Usability* measures how *effectively*, *efficiently* and *satisfactorily* a system, product or service can be used by the specified users for achieving their specified goals [26]. We carried out *formative usability testing* with a total of six participants. *Formative* usability testing is done during the development process with a relatively small number of participants to identify potential issues [4]. While traditional usability testing requires 30 to 50 test subjects, a type of *formative* usability testing approach known as *discount usability testing* has been shown to uncover 85% of the issues for the task at hand with 5 test subjects, with no significant subsequent increase in the ratio of testing cost to benefits gained as the number of participants is increased [47]. Due to the highly specialised profile (i.e. medical experts) required for the test subjects in

Experts' Dialogue Game

Collaborative Decision Making. *Redefined.*

Welcome **Alpha!**

Participants

Alpha

Beta

Gamma

Discussion

Alpha. 12/08/2022, 20:38:22

Moveid: 1

Patient history is

Patient A is 20kg overweight and lactose intolerant. `Prompt`

> **Gamma.** 12/08/2022, 20:40:53
>
> Moveid: 8
>
> Patient history is
>
> low stamina. `Reply`

Alpha. 12/08/2022, 20:38:22

Moveid: 2

I diagnose

obesity. `Prompt`

> **Beta.** 12/08/2022, 20:38:47
>
> Moveid: 5
>
> I agree
>
> that A is obese.
>
> **Beta.** 12/08/2022, 20:39:09
>
> Moveid: 6
>
> I diagnose
>
> allergy to lactose. `Reply`
>
> > **Gamma.** 12/08/2022, 20:40:08
> >
> > Moveid: 7
> >
> > Can you justify
> >
> > why you diagnose lactose allergy? `Reply`
> >
> > > **Beta.** 12/08/2022, 20:48:21
> > >
> > > Moveid: 12
> > >
> > > Because
> > >
> > > Because an intolerance to lactose is a disorder and a case of lactose allergy. `Reply`

Fig. 3. Screenshot of the web application implementing EDG.

this user study, it was difficult to recruit test subjects. So discount testing was done with the minimum number of participants for the formative study. The goal of the study was to elicit user's perspectives on effectiveness, efficiency and satisfaction of the platform and the underlying protocol in meeting their professional communication needs.

Participants. The participants were final year medical students from a medical university in Barcelona, Spain. They were volunteers who responded to a call for participation after reading the advertised information sheet through their University's human resource department.

Design. For each session of the usability testing, participants were divided into groups of three. They were then tasked with collaboratively deciding on the best possible diagnosis and advice for an anonymous patient with a thyroid disorder. The patient data was taken from the publicly available thyroid dataset from the UCI ML repository (https://archive.ics.uci.edu/ml/datasets/Thyroid+Disease). Each participant was given an identical task description but the patient data was unequally distributed between the participants to see whether transfer of information would take place. All sessions were conducted in a computer lab at a university where the participants had access to computers. All the participants in one session were in the same room at the same time. This was done to facilitate the administration of the study. However, participants were not allowed verbal or non-verbal (gestures such as eye contact) communication with each other during the study in order to ensure that all communication took place through the web application. After the participants finished the task, semi-structured interviews were conducted to get their qualitative feedback on the application and the underlying protocol. The participants were given 30 minutes for the task and 11–20 minutes for the post session interviews. All the participants who were in the same session were interviewed together as a group. Figure 4 shows the sequence of steps participants performed during the user study. First, all participants registered with the web application. The
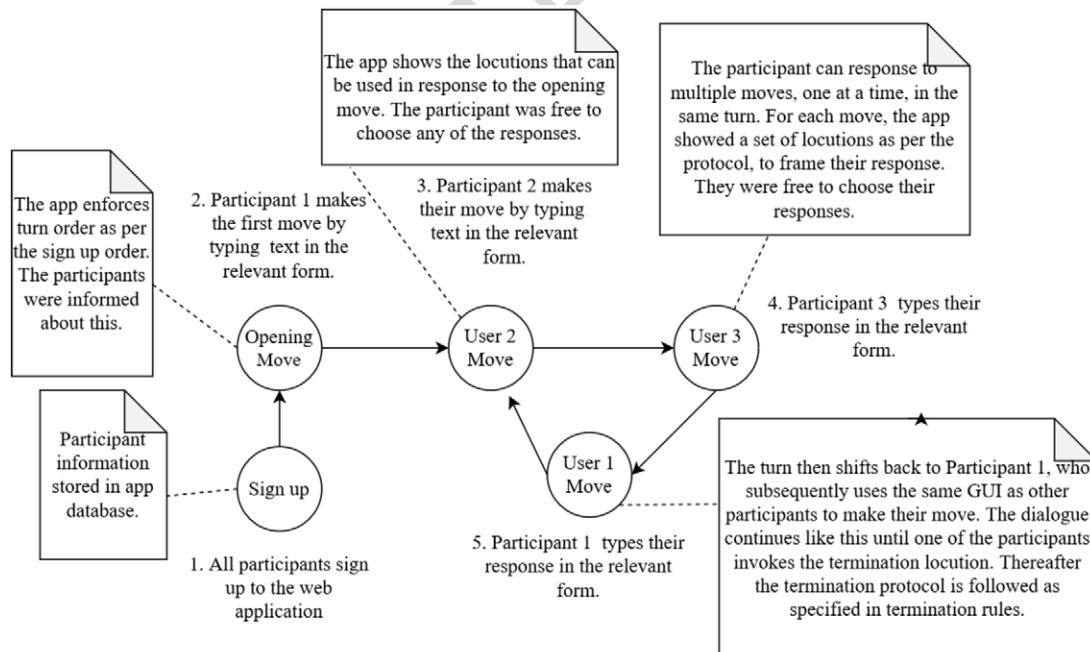

Fig. 4. Workflow diagram of the user study design.

web application enforced the sign up order as the turn taking order in accordance with the turn taking rules. The first participant then made the opening move using the Graphical User Interface (GUI) of the web application. The turn was then passed onto the next participant in line. Each participant was notified by the web application when it was their turn. Each participant was free to choose which locutions they wanted to respond to and also which locution to use as their response. The online dialogue proceeded like this until one participant invoked the termination protocol by using the *end* locution. Afterwards, the termination protocol was followed as described in termination rules in Section 4.4.

Goals.   The goal of the user study was to elicit perspectives of medical experts on aspects related to effectiveness, efficiency, engagement and ease of learning for the discussion platform and the underlying protocol. These aspects also tie in to the requirement specification from Section 3. However, they allow some additional usability considerations to be explicitly taken into account. Specifically, it aimed to answer the following questions for each of these aspects:

(1) **Effective**

    (a) Does the platform add value to professional discussions of medical experts?
    (b) Did knowledge transfer take place as a result of the discussion?
    (c) Are participants satisfied with the explanations provided?
    (d) Are participants satisfied with the final decision and its justification?

(2) **Efficient**

    (a) Do participants have the moves they need at each step to express themselves?
    (b) Does the application impede the participants in some way during the dialogue?

(3) **Engaging** Do participants rate the experience as enjoyable?
(4) **Easy to Learn**

    (a) Does the application and the protocol promote discussion?
    (b) Do participants find the classification of explanation requests useful or confusing?

The next section provides details on the result of the user study. Additionally it provides a detailed evaluation of EDG according to the requirement specification of Section 3.

6. Evaluation against requirements for expert collaboration

In this section we discuss how the protocol and its implementation measure up against the different types of requirements identified in Section 3. We introduce three evaluation criteria, referred to as *levels*, for this: *dialogue*, *system design* and *user study*. Each of these is introduced next.

Dialogue.   This level determines how well the dialogue rules presented in Section 4 contribute towards satisfying each requirement.

System design.   This level evaluates each requirement against the protocol implementation since some requirements can only be met at the implementation level.

User study.   This level validates the satisfaction of each requirement through the user study described in Section 5.2.

Table 8 presents a summary of the results. The first column shows the id of each requirement, columns second to fourth show whether the corresponding requirement is verified through the *dialogue, system*

Table 8

Summary of requirements evaluation

Id	Evaluated against			Participant quote
	Dialogue	System design	User study	
RA1	✓	△	✓	*When answering the question if it felt natural to ask and be asked for clarifications and explanations etc*: "Maybe sometimes there are people who will not tell you so directly (in real life) but the advantage of having this is that people will be more used to be asked for clarification and understand that it is the program that is predetermined." [Participant 5]
RC1	✓	△	✓	"I think that it (the app) is more useful for medicals (doctors) to be in a team and to be more active if you don't have 'I disagree' because (if) the other one doesn't give a correct explanation or contra answer and only puts 'I disagree', you might take it personally, it may demotivate you." [Participant 4]
RC2	✓			
RC3	✓	△	✓	*Answering the question if the app allowed them to have a productive discussion?* "Yeah, it is interesting because we can have a quick chat to discuss case, it's a good option to discuss." [Participant 3]
RC4	✓	△	✓	"A hundred per cent for me because I hadn't done the exam of Endocrinology so it was kind of fresh." [Participant 6]
RP1	✓	△		
RP2	✓	△		
RP3	✓	△		
RP4	✓	△		
RP5	✓	✓		
RP6	✓			
RI1		✓		
RI2		✓		
RI3		Partially		

design or *user study*. Finally, the last column shows the representative quote from a participant in case the requirement satisfaction is verified through the user study. The checkmark symbol in a table cell shows that the corresponding requirement is satisfied for the level indicated in the column header. The triangle symbol indicates that the corresponding requirement is satisfied indirectly by virtue of implementing the dialogue protocol. We distinguish it from the checkmark to show that the system design level does not make an active contribution to fulfilling the requirement for these cases. So, the evaluation for these requirements is superficial at this level. A blank value for the cell indicates that corresponding requirement is not satisfied by the level indicated in that column header. Following paragraphs give a detailed discussion on each row of Table 8.

6.1. Evaluation against agent oriented requirement

RA1 is the only requirement that is directly derived from the participant's characteristics. Hence, it can be evaluated through the user study as well as with dialogue rules. By virtue of the fact that the dialogue rules already satisfy *RA1*, the system design also satisfies the requirement inherently. We use the symbol △ to indicate this.

All locution groups **L1** to **L4** enable *RA1* because presenting the set of possible options to speakers avoids failures from their possible lack of attention and assertiveness. It also enables richer dialogues

by making the participants aware of the possible directions for branching out. This was also verified during the user study where the participants were of the opinion that the direct requests for explanations, clarifications and justifications allowed for more assertiveness and clearer communication.

6.2. Evaluation against cooperation requirements

This section describes how the system and the dialogue game measures up against each of the cooperation oriented requirements on the three levels.

At the dialogue level, *RC1* is enabled through two mechanisms. The first is absence of any explicit locution to express disagreement such as 'I disagree'. This allows the protocol to avoid deadlocks amongst participants. Secondly cooperation is enforced through implicit disagreements using locution subsets **L2** and **L3**. This was also verified during the user study where participants expressed the view that disallowing explicit disagreement was more useful in promoting cooperation and goodwill. *RC1* is also part of the system design inherently but marked as △ since it is superficially satisfied at the system design level.

RC2 is considered as a cooperation requirement since the dialogue protocol requires cooperation between the participants as a means for quality control. Hence, the protocol ensures *RC2* because it requires at least one explicit agreement by another participant in order for the recommendation to be added to the collective commitment store. Even at that stage, they are open to non-monotonic debate. Hence at the end of the dialogue, only decisions that have survived the critical discussion are recommended. The embedding of explanatory illocutionary force, represented by locution **L2.2**, **L2.1** and **L2.3**, allows the participants to not only probe into each other's decisions but explanations as well. As in the case of decisions, only explanations which survive the critical debate make it into the collective commitment store. Although the dialogue rules provide a mechanism to ensure quality, the effectiveness of these quality control mechanism can only be verified in a deployment environment. Hence, the *RC2* cannot be verified at the system design level so it is left as blank in the table. Similarly, since validating quality control in a deployment environment is a rigorous process and requires certified professionals, this requirement could not be validated within the scope of the current user study. Hence, it is left blank in the table.

Since *RC3* also requires a cooperative setting, it is listed as a cooperation requirement. The dialogue rules enforce *RC3* by not forcing the participants to reply only to the preceding move, rather the rules allow participants the flexibility to reply to any number of previous locutions at any time. This allows any participant to either introduce new knowledge into the conversation by providing facts using locution rule subsets **L1** and **L3** or by asking questions using subset **L2**. Moreover, subsets **L2** and **L3** of locution rules allow the participants to probe into the statements of other participants, making detailed discussion possible. The dialogue is open since each participant has access to the dialogue state at all times. This was also verified during the user study where the participants felt that the discussion platform allowed them to have a *productive* discussion. Since the system design also fulfils this requirement inherently by virtue of implementing the protocol, it is marked as △ for system design.

RC4 is also dependent on cooperation of participants since only cooperation can ensure knowledge transfer. The dialogue game protocol enables this through locution classes **L1, L2** and **L3**. Any member of these classes can be selected by the respondent to share their knowledge. This was verified during the user study because participants felt that they were able to gain new knowledge as a result of the information exchange. However, as before the system design satisfies this property inherently so it is marked as △ under the system design column.

6.3. Evaluation against protocol oriented requirements

Next we evaluate the dialogue game against each of the six protocol oriented requirements in detail as summarised in Table 8.

RP1 and *RP2* are topic-based and lay down specific requirements for the inclusion of these topics. Their satisfaction can be verified through locution rules **L1.1** and **L1.4** which allow any participant to introduce patient history and critical points into the dialogue at any time. Moreover, there is no limitation on the amount of information transfer that can be done because the locutions can be repeated as many times as needed. The system design inherently satisfied these requirements so these are marked as △ in Table 8. Verifying these requirements through the user study would be trivial, so these are marked as empty cells in Table 8.

The protocol incorporates *RP3* through locution rules **L2.1, L2.3, L3.1** and **L3.3** which cover explanation and clarification requests as well as the corresponding responses. As before, since *RP3* is inherently satisfied by system design, it is marked as △ in Table 8. Similarly, it can only be verified superficially through the user study, so the corresponding cell in Table 8 is left blank.

The dialogue game provides mechanism to resolve conflicts through the introduction of locution rules **L2** and **L3**, thus satisfying *RP4*. Through allowing for embedding of explanatory illocutionary forces within persuasive illocutionary forces and vice versa, disagreements are resolved indirectly by forcing the participants to spell out the nature of their disagreement rather than merely expressing it. For example, a disagreement due to need for evidence (represented by **L2.2**), as a result of missing link (**L2.3**) or a request for more information (**L2.1**). The participants can subsequently engage in a series of embedded explanation dialogues until the issue is resolved to their satisfaction. Since the system design inherently satisfies the requirement, it is marked as △ in Table 8. In the current user study, no conflicts appeared during the dialogues so this requirement could not be verified through the user study. Hence, it is left as blank in Table 8.

By enforcing turn-taking, the protocol ensures equal opportunity for getting input from all participants, satisfying *RP5*. Since it is trivial to verify this requirement through the user study, it is marked as empty cell in Table 8. However, since the system design enforces a turn-taking mechanism, it satisfies *RP5*.

Finally, the protocol incorporates *RP6* by giving the same validity to all moves by all participants. This requirement concerns the dialogue protocol definition so it does not make much sense to evaluate it through the user study or against system design. Hence, the corresponding cells are marked as empty in Table 8.

6.4. Evaluation against implementation oriented requirements

Next we discuss evaluation of each of the implementation oriented requirements according to the three evaluation criteria.

RI1 and *RI2* require that the dialogue game be coordinated and the dialogue should be recorded. Both of these are satisfied at the system design level because the system acts as coordinator of the dialogue and does administrative book keeping tasks such as regulating turns, recording the dialogue and informing participants of the moves available at each time in the dialogue. Although the protocol description requires that these two requirements be met, it does not specifically provide a mechanism to enforce this. *RI2* in particular, can only be enforced through system design. Hence, both these requirements are evaluated and verified at the system design level rather through dialogue rules or user study. So, the corresponding columns are marked as blank for the latter.

The dialogue rules do not provide any mechanism to protect patient privacy as specified by *RI3*. However, this requirement is partially satisfied through system design. This is because the platform limits the scope of information transfer to only the participants of the dialogue. Therefore, it protects patient privacy by design. Moreover, in the case of using anonymised data, it guarantees absolute privacy of the patient. However, the current prototype does not encrypt the information exchanged to secure it from malicious interference and leaves the anonymisation of the data exchanged to the participants' discretion. Hence, it only partially fulfils *RI3*. Since it does not make much sense to evaluate it through the user study, the corresponding cell is marked as blank in Table 8.

7. User perspectives on expert collaboration system

The post session interviews from the user study were recorded with the consent of participants, transcribed and the interview data was *thematically analysed* to discover key insights. Thematic analysis is concerned with identifying patterns in qualitative data. The analysis can identify themes at the surface meaning level, referred to as *semantic* level or go beyond what was said to discover underlying concepts, known as *latent level* [37]. Attention was paid to both semantic and latent meaning implied in the feedback. Specifically, the comments from the participants were organised into related concepts following the bottom-up organisation method of *affinity matching*. Affinity matching is a bottom-up analysis technique for analysing data from a user study. In this case, relevant findings are grouped and the category labels are inferred from these groupings [5]. The advantage as opposed to a top-down approach with pre-defined labels is that it keeps an open mind to what the data might reveal. We identified six themes from the participants' feedback on various aspects of the dialogue game. Table 9 shows the representative quotes form participants against each identified theme. Next, we describe the insights from users' perspectives for each theme.

Value in real-life use. The most common theme was the practical value of such a discussion system in the medical community. All participants in group 1 agreed that the platform would facilitate professional communication between medical experts, especially multidisciplinary communication and communication between public and private sectors. Participants agreed that the platform facilitates knowledge transfer and allows for a productive discussion. One participant was of the opinion that the platform helps clear communication by taking on the burden of *politeness*, in that doctors in general do not ask each other for clarifications and justifications directly as it could be considered rude but since this is part of the platform's functionality, it no longer seems like a personal affront, rather just the way the platform works. This is a really significant comment since providing a means of getting around personal issues that hamper communication was one of the basic requirements the platform and the underlying protocol aimed to fulfil.

Satisfaction with move options. One of the most important practical aspects of a dialogue game for expert discussion is whether it allows the participants to express themselves completely. This was evaluated through taking participants' feedback on whether they found the dialogue locution sufficient for their discussion. All participants in group 1 expressed overall satisfaction with the request-response options that are part of the underlying protocol. They thought that it presented the *ideal* scenario. Three out of six participants thought that the response *I agree* was not sufficient by itself because they would have liked to express partial agreement as well. For example, *I agree but*. One participant was of the view that not having any *I do not agree* option explicitly was good because doctors are in general always coming

Table 9

Representative quotes of participants for each emergent theme

S. no.	Theme	Participant quotes
1	Value in real-life use	"Yeah it is a good tool because we have to talk with this specialist living in Madrid and it's a good chance to realise the questions or try to discuss on a case." [Participant 3] "...I had different values from my mate (for) TSH so I could receive more information from her and I think this is really useful because sometimes you can use the analytical values from another speciality where he has made the analytics and do interdisciplinary (exchange) between different specialities without repeating (the) exams because what they do is, that they already have the exam and when they go to a different doctor from another, they repeat the exam." [Participant 5]
2	Satisfaction with move options	"I think it is like ideal, it would be a good real scenario. It doesn't happen but I think it would be much easier like that." [Participant 3] "...Sometimes I feel like we need an option that I am agree but like...it's not completely white or black." [Participant 2]
3	Utility of different explanation requests	"Maybe not so hard [as] justify or why you think about this or why you think about that. Not...not, we don't use that kind of hard expressions because justify is like WHY you are saying that." [Participant 2] *Answering why they don't consider justification request as rude?* "Because maybe she knows what she wants to say but the way she explained to us is not as clear as she liked so." [Participant 4]
4	Effect of turn-taking	"No, I think that it is good to give your turn for speaking because if not all the people will share information really quick and it will be difficult to extract important things...umm...and it gives you time to think what you are gonna say and correct it if you have any error but maybe...it will be good if you could ask for the turn to the program and you can answer in the order you have asked for your turn because maybe in one moment I didn't have an answer and it was my turn and the other person needed to wait for me but I think it's a really good...[illegible]." [Participant 5] "...The part in turns, is challenging because if somebody else is writing something and you see the message and you have an idea that you would like to comment (on) but you have to wait for another one, then it could, I don't know, slip your mind (or) whatever and then like (you would) not give your advice or recommendation that you should." [Participant 6]
5	User interface	"Maybe it will be easier if the values, number values of the diagnosis are shown at the right of the screen in a box so that you don't need to look for them in the dialogue and you don't lose information." [Participant 4] "I liked the discussion is divided by topics. So if you have the I diagnose part, then you can see like what everybody has said about diagnosis and it's not like one message about diagnosing, one message about exams, treatment all over, that get's like really messy. It was easier to read all about one part." [Participant 6]
6	User study design	"Yeah if it's in a clinical setting I think [Participant 1] would have had all that data and we would only advise on the data. It felt kind of weird." [Participant 2] "I am not a hundred per cent sure about giving fragments of patient history to each of us because is like even with software and stuff, for patients, you have already all the information." [Participant 6]

to a deadlock because of this so it was useful to ask for an explanation instead rather than expressing explicit disagreement.

Utility of different explanation requests. In order to evaluate whether the classification of different explanation requests is useful in a practical scenario, participants were asked whether they found the categorisation useful. Three out of six participants in group 1 felt that all explanation requests were useful and they would use them in their communication. However, the other half argued that a justification request sounded really aggressive and unnatural and they would never use it in a real world scenario.

They would prefer to use the explanation request instead. However, all participants agreed that they were not concerned about the subtle differences between explanation, clarification and justification requests as long as the intention of asking for an explanation was conveyed.

Effect of turn-taking.   Since traditional collaboration between experts is real-time, it was an open research question for protocol design whether it should allow synchronous or asynchronous communication. In order to meet requirements *RP5* and *RP6*, the protocol was designed to be synchronous and participants were asked to give their feedback based on their experience. Two different perspectives emerged on the turn-taking mechanism implemented in the platform and the underlying protocol. Two out of six participants in group 1 felt that the turns were good and helped to coordinate the discussion. One participant felt that it would be useful to be able to edit your response when it was not your turn because you might remember something after your turn had passed. One participant was of the view that ideas slipped your mind while waiting for your turn while another participant felt that the turn-taking put a lot of pressure on you to say something during your turn even when you did not think that you had anything to say. They suggested that a dynamic turn-taking mechanism would be better to preserve coordination of the discussion and to deal with the last two problems. They thought that an option to queue for the turn token would be best.

User interface.   Although the main aim for the user study was to evaluate the dialogue game protocol, it was anticipated that participants would end up evaluating the system at the user interface level. This was confirmed in the user study where the participants ended up offering several valuable suggestions for user interface improvements. All participants in group 1 thought that the clear separation of the discussion into history, diagnosis, advice and concerns was very useful and efficient. Three out of six participants seemed happy with the user interface. Three out of six participants agreed that having the history on a side panel would be more efficient because it would help reduce scrolling time to check on it. One participant also felt that displaying history in separate lines would also improve presentation. One participant felt that nesting the messages would make the presentation more clear and efficient.

User study design.   One of the themes that emerged was that the participants were surprised with how the task in the user study was designed. Specifically, they found the distribution of patient history amongst the different participants unnatural compared to their experiences in the professional scenarios. All participants in group 1 found the division of patient history between the participants unnatural compared to the professional real life scenarios in which all the history is presented upfront. While this was done to check if transfer of information took place, they argued that knowledge transfer and information took place irrespective of this small test. So for the second group all the patient history was provided upfront to participant 1.

8. Conclusion and future work

This work envisioned a human-artificial agent hybrid collaborative recommendation system. As a first step towards this end, it presented a requirement specification for collaborative interactions between experts and an inquiry dialogue game grounded in the specification. The dialogue game allows multiple expert agents to collaborative on the best recommendations for a user. The game combines explanatory illocutionary forces in an inquiry dialogue. The motivation for doing this is to make the inquiry process and consequently, the output of the multiagent system explainable by generating richer traces for the reasoning process itself. The game presents an approach towards incorporating explainability within

multiagent systems. This work also presented an evaluation of the dialogue game against the requirement specification through a user study. The user study was also significant in that it highlighted the real life utility for such dialogue platforms in the medical domain. Such platforms can enable clearer and systematic communication across multiple healthcare disciplines and sectors. This work proposes to import the methodology of software engineering into the area of formal dialogues. This methodology consists of the following steps: the collection of requirements for a dialogue game in a selected domain application; the design and implementation of a protocol according to these requirements; and the evaluation of a dialogue system against the requirements.

The next step would be to implement and evaluate EDG for a multiagent system. This would involve evaluating formal properties of the system such as deadlocks, livelocks and termination guarantees. One possible approach for investigating runtime termination guarantees for EDG is to explore multiagent frameworks that can provide this kind of guarantee through an appropriate moderator role. For example, the governor role in Electronic Institutions [19] can be extended to close termination property at runtime. Another research direction could be to close the protocol against disruption by non-cooperative and malicious agents. It might also be interesting to investigate whether the protocol can do away with turn-taking and synchronous communication since it may not scale well to a system with many participants.

While EDG presents which locutions can be used in response to others, it does not investigate how an agent can select the best locution in response to another based on its knowledge base. This is a very important aspect to implement the protocol in a multiagent system. Hence, an important future direction is to develop reasoning strategies for agents for participating in the EDG. One possible approach for doing this is to explore argumentation-based reasoning for EDG on the lines of [10]. Subsequently, the next step would be to adapt and implement the protocol in a human-artificial agent hybrid system. Other possible interesting directions to investigate include investigating how well the requirement specification presented in this work generalises to other domains such as engineering or aviation. On the flip side, it can also be interesting to evaluate how well existing dialogue game protocols [10,51,58] conform to the requirement specification through a user study. Finally, the prototype of the platform developed as part of this work can be refined and released as open source software for facilitating communication between experts in the healthcare domain.

The politeness rules introduced here show that in order to make dialogue games more human-centred, we need to introduce the same societal machinery being employed in real life conversation in order to safeguard the integrity of the dialogue in a computational context.

Acknowledgement

References

[1] J. Alama, A. Knoks and S.L. Uckelman, Dialogue games in classical logic, in: M. Giese, R. Kuznets, eds, *TABLEAUX 2011: Workshops, Tutorials, and Short Papers, Technical Report IAM-11-002*, Universität Bern, Bern, 2011, pp. 82–86, available from: https://doi.org/10.5167/uzh-190012.

[2] L. Amgoud, N. Maudet and S. Parsons, Modelling dialogues using argumentation, in: *Proceedings – 4th International Conference on MultiAgent Systems, ICMAS 2000*, Institute of Electrical and Electronics Engineers Inc., 2000, pp. 31–38.

[3] A. Arioua, P. Buche and M. Croitoru, Explanatory dialogues with argumentative faculties over inconsistent knowledge bases, *Expert Systems with Applications* **80** (2017), 244–262. doi:10.1016/j.eswa.2017.03.009.

[4] C.M. Barnum, Establishing the essentials, in: *Usability Testing Essentials*, C.M. Barnum, ed., 2nd edn, Morgan Kaufmann, 2021, pp. 9–33, available from: https://www.sciencedirect.com/science/article/pii/B9780128169421000010. doi:10.1016/B978-0-12-816942-1.00001-0.

[5] C.M. Barnum, 8, in: *Analyzing the Findings*, S. Merken, ed., Elsevier, 2021, pp. 287–319.

[6] M. Beveridge and J. Fox, Automatic generation of spoken dialogue from medical plans and ontologies, *Journal of Biomedical Informatic* **39**(5) (2006), 482–499. doi:10.1016/j.jbi.2005.12.008.

[7] F. Bex and H. Prakken, Investigating stories in a formal dialogue game, in: *Frontiers in Artificial Intelligence and Applications.*, Vol. 172, IOS Press, 2008, pp. 73–84.

[8] E. Black and K. Atkinson, Dialogues that account for different perspectives in collaborative argumentation, in: *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems*, Vol. 2, 2009, pp. 867–874, available from: papers2://publication/uuid/851E4870-2CCD-4C21-BBAD-819066F85D11.

[9] E. Black and A. Hunter, A generative inquiry dialogue system, in: *Proceedings of the International Conference on Autonomous Agents*, ACM Press, New York, USA, 2007, pp. 1014–1021, available from: http://portal.acm.org/citation.cfm?doid=1329125.1329417.

[10] E. Black and A. Hunter, An inquiry dialogue system, *Autonomous Agents and Multi-Agent Systems* **19**(2) (2009), 173–209, available from: www.cossac.org.

[11] K. Budzynska, A. Rocci and O. Yaskorska, Financial dialogue games: A protocol for earnings conference calls, in: *Frontiers in Artificial Intelligence and Applications.*, Vol. 266, IOS Press, 2014, pp. 19–30, available from: https://ebooks.iospress.nl/doi/10.3233/978-1-61499-436-7-19.

[12] P. Carayon and H.P. Human, Factors and usability for health information technology: Old and new challenges, *Yearbook of medical informatics* **28**(1) (2019), 71–77, available from: https://pubmed.ncbi.nlm.nih.gov/31419818/. doi:10.1055/s-0039-1677907.

[13] F. Castagna, A. Garton, P. McBurney, S. Parsons, I. Sassoon and E.I. Sklar, EQRbot: A chatbot delivering EQR argument-based explanations, *Frontiers in Artificial Intelligence* **6** (2023), available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10076765/.

[14] B. Chaudhry, J. Wang, S. Wu, M. Maglione, W. Mojica, E. Roth et al., Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care, *American College of Physicians* (2006).

[15] R. Craven, F. Toni, C. Cadar, A. Hadad and M. Williams, Efficient argumentation for medical decision-making, in: *Proceedings of the Thirteenth International Conference on Principles of Knowledge Representation and Reasoning. KR12*, AAAI Press, 2012, pp. 598–602.

[16] G. Delaney, S. Jacob, R. Iedema, M. Winters and M. Barton, Comparison of face-to-face and videoconferenced multi-disciplinary clinical meetings, *Australasian Radiology* **48** (2004), 487–492, Wiley Press, available from: https://pubmed.ncbi.nlm.nih.gov/15601329/.

[17] L.A. Dennis and N. Oren, Explaining BDI agent behaviour through dialogue, in: *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, Vol. 1, International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS), 2021, pp. 429–437.

[18] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, *Artificial Intelligence* **77**(2) (1995), 321–357, available from: http://linkinghub.elsevier.com/retrieve/pii/000437029400041X.

[19] M. Esteva, B. Rosell, J.A. Rodríguez-Aguilar and J.L. Arcos, AMELI: An agent-based middleware for electronic institutions, in: *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS 2004*, Vol. 1, 2004, pp. 236–243.

[20] X. Fan, R. Craven, R. Singer, F. Toni and M. Williams, Assumption-based argumentation for decision-making with preferences: A medical case study, in: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, LNAI, Vol. 8143, Springer, Berlin, Heidelberg, 2013, pp. 374–390, available from: https://link.springer.com/chapter/10.1007/978-3-642-40624-9_23.

[21] G.E. Gross, The role of the tumor board in a community hospital, *CA: A Cancer Journal for Clinicians* **37**(2) (1987), 88–92.

[22] C.L. Hamblin, *Fallacies*, 1st edn, Methuen, London, UK, 1970.

[23] C.H. Hennessy and A. Walker, Promoting multi-disciplinary and inter-disciplinary ageing research in the United Kingdom, *Ageing and Society* **31**(1) (2011), 52–69, available from: https://www.cambridge.org/core/journals/ageing-and-society/article/abs/promoting-multidisciplinary-and-interdisciplinary-ageing-research-in-the-united-kingdom/044A8FFB174A3BF742EBA7F7DB3A4BD1. doi:10.1017/S0144686X1000067X.

[24] J. Huang, N.R. Jennings and J. Fox, Cooperation in distributed medical care, in: *2nd Int. Conf. on Cooperative Information Systems (CoopIS-94) (01/01/94)*, 1994, pp. 255–263, available from: https://eprints.soton.ac.uk/252139/.

[25] J. Huang, N.R. Jennings and J. Fox, An agent architecture for distributed medical care, in: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 890, Springer Verlag, 1995, pp. 219–232, available from: https://link.springer.com/chapter/10.1007/3-540-58855-8_14.

[26] ISO, ISO 9241-11:2018 – ergonomics of human–system interaction – part 11: Usability: Definitions and concepts, 2018, available from: https://www.iso.org/standard/63500.html.

[27] M. Janier, M. Snaith, K. Budzynska, J. Lawrence and C. Reed, A system for dispute mediation: The mediation dialogue game, *Frontiers in Artificial Intelligence and Application* **287** (2016), 351–358, available from: https://doi.org/doi:10.3233/978-1-61499-686-6-351.

[28] A.R. Kagan, The multidisciplinary clinic, *International Journal of Radiation Oncology Biology Physics* **61**(4) (2005), 967–968. doi:10.1016/j.ijrobp.2004.10.040.

[29] N. Kokciyan, I. Sassoon, E. Sklar, S. Modgil and S. Parsons, Applying metalevel argumentation frameworks to support medical decision making, *IEEE Intelligent Systems* **36**(2) (2021), 64–71. doi:10.1109/MIS.2021.3051420.

[30] A.M. Kurahashi, J.N. Stinson, M. van Wyk, S. Luca, T. Jamieson, P. Weinstein et al., The perceived ease of use and usefulness of loop: Evaluation and content analysis of a web-based clinical collaboration system, *JMIR Human Factors* **5**(1) (2018), available from: https://pubmed.ncbi.nlm.nih.gov/29317386/.

[31] B.W. Lamb, N. Sevdalis, S. Arora, A. Pinto, C. Vincent and J.S.A. Green, Teamwork and team decision-making at multidisciplinary cancer conferences: Barriers, facilitators, and opportunities for improvement, *World Journal of Surgery* **35**(9) (2011), 1970–1976, available from: https://pubmed.ncbi.nlm.nih.gov/21604049/. doi:10.1007/s00268-011-1152-1.

[32] H.J. Lin, Y.L. Ko, C.F. Liu, C.J. Chen and J.J. Lin, Developing and evaluating a one-stop patient-centered interprofessional collaboration platform in Taiwan, *Healthcare (Switzerland)* **8**(3) (2020), available from: https://pubmed.ncbi.nlm.nih.gov/32751264/.

[33] P. Liu, A. Lyndon, J.L. Holl, J. Johnson, K.Y. Bilimoria and A.M. Stey, Barriers and facilitators to interdisciplinary communication during consultations: A qualitative study, *BMJ Open* **11**(9) (2021), available from: https://pubmed.ncbi.nlm.nih.gov/34475150/.

[34] N.J. Look Hong, A.R. Gagliardi, S.E. Bronskill, L.F. Paszat and F.C. Wright, Multidisciplinary cancer conferences: Exploring obstacles and facilitators to their implementation, *Journal of oncology practice* **6**(2) (2010), 61–68, available from: http://www.ncbi.nlm.nih.gov/pubmed/20592777 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2835483.

[35] E.J. Macaskill, S. Thrush, E.M. Walker and J.M. Dixon, Surgeons' views on multi-disciplinary breast meetings, *European Journal of Cancer* **42**(7) (2006), 905–908. doi:10.1016/j.ejca.2005.12.014.

[36] P. Madumal, T. Miller, L. Sonenberg and F. Vetere, A grounded interaction protocol for explainable artificial intelligence, in: *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, Vol. 2, 2019, pp. 1033–1041, available from: http://arxiv.org/abs/1903.02409.

[37] M. Maguire and B. Delahunt, Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars, *All Ireland Journal of Higher Education* **9**(3) (2017), available from: https://ojs.aishe.org/index.php/aishe-j/article/view/335.

[38] A. Martínez-García, A. Moreno-Conde, F. Jódar-Sánchez, S. Leal and C. Parra, Sharing clinical decisions for multimorbidity case management using social network and open-source tools, *Journal of Biomedical Informatics* **46**(6) (2013), 977–984. doi:10.1016/j.jbi.2013.06.007.

[39] N. Maudet, S. Parsons and I. Rahwan, Argumentation in multi-agent systems: Context and recent developments, in: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, LNAI, Vol. 4766, Springer Verlag, 2007, pp. 1–16.

[40] P. McBurney, R.M. Van Eijk, S. Parsons and L. Amgoud, A dialogue game protocol for agent purchase negotiations, *Autonomous Agents and Multi-Agent Systems* **7**(3) (2003), 235–273. doi:10.1023/A:1024787301515.

[41] S. Modgil, P. Tolchinsky and U. Cortés, Towards formalising agent argumentation over the viability of human organs for transplantation, in: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, LNAI, Vol. 3789, Springer, Berlin, Heidelberg, 2005, pp. 928–938, available from: https://link.springer.com/chapter/10.1007/11579427_95.

[42] R.S. Morse, K. Lambden, E. Quinn, T. Ngoma, B. Mushi, Y.X. Ho et al., A mobile app to improve symptom control and information exchange among specialists and local health workers treating Tanzanian cancer patients: Human-centered design approach, *JMIR Cancer* **7**(1) (2021), available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8088847/.

[43] A.J. Munro and S. Swartzman, What is a virtual multidisciplinary team (vMDT)?, *British Journal of Cancer* **108**(12) (2013), 2433–2441, available from: https://pubmed.ncbi.nlm.nih.gov/23756866/. doi:10.1038/bjc.2013.231.

[44] S. Naveed, T. Donkers and J. Ziegler, Argumentation-based explanations in recommender systems: Conceptual framework and empirical results, in: *UMAP 2018 – Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*, Association for Computing Machinery, Inc., New York, NY, USA, 2018, pp. 293–298, available from: https://dl.acm.org/doi/10.1145/3213586.3225240. doi:10.1145/3213586.3225240.

[45] V. Ngo, C.G. Matsumoto, J.G. Joseph, J.F. Bell, R.J. Bold, A. Davis et al., The personal health network mobile app for chemotherapy care coordination: Qualitative evaluation of a randomized clinical trial, *JMIR mHealth and uHealth* **8**(5) (2020), e16527, available from: https://mhealth.jmir.org/2020/5/e16527.

[46] J.X. Nie, C. Heidebrecht, A. Zettler, J. Pearce, R. Cunha, S. Quan et al., The perceived ease of use and perceived usefulness of a web-based interprofessional communication and collaboration platform in the hospital setting: Interview study with health care providers, *JMIR Human Factors* **10**(1) (2023), e39051, available from: https://humanfactors.jmir.org/2023/1/e39051.

[47] J. Nielsen, Why you only need to test with 5 users, 2000, available from: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/.

[48] S. Parsons, M. Wooldkidge and L. Amgoud, Properties and complexity of some formal inter-agent dialogues, *Journal of Logic and Computation* **13**(3) (2003), 347–376. doi:10.1093/logcom/13.3.347.

[49] V. Patkar, D. Acosta, T. Davidson, A. Jones, J. Fox and M. Keshtgar, Using computerised decision support to improve compliance of cancer multidisciplinary meetings with evidence-based guidance, *BMJ Open* (2012), e000439, available from: http://dx.doi.org/10.1136/bmjopen-2011-000439.

[50] J. Pitt and A. Mamdani, Communication protocols in multi-agent systems: A development method and reference architecture, in: *Issues in Agent Communication*, Springer, Berlin, Heidelberg, 2000, pp. 160–177, available from: https://link.springer.com/chapter/10.1007/10722777_11. doi:10.1007/10722777_11.

[51] H. Prakken, Relating protocols for dynamic dispute with logics for defeasible argumentation, *Synthese* **127**(1–2) (2001), 187–219. doi:10.1023/A:1010322504453.

[52] H. Prakken, Coherence and flexibility in dialogue games for argumentation, *Journal of Logic and Computation* **15**(6) (2005), 1009–1040, available from: http://academic.oup.com/logcom/article/15/6/1009/1086845/Coherence-and-Flexibility-in-Dialogue-Games-for. doi:10.1093/logcom/exi046.

[53] H. Prakken, A formal model of adjudication dialogues, *Artificial Intelligence and Law* **16**(3) (2008), 305–328, available from: https://link.springer.com/article/10.1007/s10506-008-9066-4. doi:10.1007/s10506-008-9066-4.

[54] H. Prakken and R. Ratsma, A top-level model of case-based argumentation for explanation: Formalisation and experiments, *Argument & Computation* **13** (2022), 159–194.

[55] M. Quinn, J. Forman, M. Harrod, S. Winter, K.E. Fowler, S.L. Krein et al., Electronic health records, communication, and data sharing: Challenges and opportunities for improving the diagnostic process, *Diagnosis* **6**(3) (2019), 241–248, available from: https://pubmed.ncbi.nlm.nih.gov/30485175/. doi:10.1515/dx-2018-0036.

[56] A. Rago, H. Li and F. Toni, Interactive explanations by conflict resolution via argumentative exchanges, 2023, arXiv:2303.15022.

[57] R.B. Rajasekaran, D. Whitwell, T.D.A. Cosker, C.L.M.H. Gibbons and A. Carr, Will virtual multidisciplinary team meetings become the norm for musculoskeletal oncology care following the COVID-19 pandemic? – Experience from a tertiary sarcoma centre, *BMC Musculoskeletal Disorders* **22**(1) (2021), 18, available from: https://pubmed.ncbi.nlm.nih.gov/33402136/.

[58] L. Riley, K. Atkinson, T. Payne and E. Black, An implemented dialogue system for inquiry and persuasion, in: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, LNAI, Vol. 7132, Springer, Berlin, Heidelberg, 2012, pp. 67–84, available from: https://link.springer.com/chapter/10.1007/978-3-642-29184-5_5.

[59] I. Sassoon, N. Kökciyan, S. Modgil and S. Parsons, Argumentation schemes for clinical decision support, *Argument and Computation* **12**(3) (2021), 329–355. doi:10.3233/AAC-200550.

[60] I. Sassoon, N. Kökciyan, E. Sklar and S. Parsons, Explainable argumentation for wellness consultation, in: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, LNAI, Vol. 11763, Springer Verlag, 2019, pp. 186–202, available from: https://doi.org/10.1007/978-3-030-30391-4_11.

[61] B. Schellenberger, A. Diekmann, C. Heuser, N. Gambashidze, N. Ernstmann and L. Ansmann, Decision-making in multidisciplinary tumor boards in breast cancer care – an observational study, *Journal of Multidisciplinary Healthcare* **14** (2021), 1275–1284, available from: /pmc/articles/PMC8179814/ /pmc/articles/PMC8179814/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8179814/. doi:10.2147/JMDH.S300061.

[62] J.R. Searle, *Speech Acts*, Cambridge University Press, 1969, available from: https://www.cambridge.org/core/product/identifier/9781139173438/type/book.

[63] Q. Shaheen, A. Toniolo and J.K.F. Bowles, Argumentation-based explanations of multimorbidity treatment plans, in: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, LNAI, Vol. 12568, Springer Science and Business Media Deutschland GmbH, 2021, pp. 394–402, available from: https://doi.org/10.1007/978-3-030-69322-0_29.

[64] M. Snaith, D. De Franco, T. Beinema, H. Op Den Akker and A. Pease, A dialogue game for multi-party goal-setting in health coaching, in: *Frontiers in Artificial Intelligence and Applications.*, Vol. 305, IOS Press, 2018, pp. 337–344, available from: https://ebooks.iospress.nl/doi/10.3233/978-1-61499-906-5-337.

[65] I. Stepin, K. Budzynska, A. Catala, M. Pereira-Fariña and J.M. Alonso-Moral, Information-seeking dialogue for explainable artificial intelligence: Modelling and analytics, *Argument & Computation* (2023), 1–59, preprint.

[66] K.M. Sutcliffe, E. Lewton and M.M. Rosenthal, Communication failures: An insidious contributor to medical mishaps, *Academic Medicine* **79**(2) (2004), 186–194, available from: https://pubmed.ncbi.nlm.nih.gov/14744724/. doi:10.1097/00001888-200402000-00019.

[67] M. Taberna, F.G. Moncayo, E. Jané-Salas, M. Antonio, L. Arribas, E. Vilajosana et al., *The Multidisciplinary Team (MDT) Approach and Quality of Care*, Frontiers Media S.A., 2020, available from: /pmc/articles/PMC7100151/ /pmc/articles/PMC7100151/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7100151/.

[68] T. Tang, C. Heidebrecht, A. Coburn, E. Mansfield, E. Roberto, E. Lucez et al., Using an electronic tool to improve teamwork and interprofessional communication to meet the needs of complex hospitalized patients: A mixed methods study, *International Journal of Medical Informatics* **127** (2019), 35–42. doi:10.1016/j.ijmedinf.2019.04.010.

[69] T. Tang, M.E. Lim, E. Mansfield, A. McLachlan and S.D. Quan, Clinician user involvement in the real world: Designing an electronic tool to improve interprofessional communication and collaboration in a hospital setting, *International Journal of Medical Informatics* **110** (2018), 90–97. doi:10.1016/j.ijmedinf.2017.11.011.

[70] P. Tolchinsky, K. Atkinson, P. Mcburney, S. Modgil and U. Cortés, Agents deliberating over action proposals using the ProCLAIM model, in: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, LNAI, Vol. 4696, Springer Verlag, 2007, pp. 32–41, available from: https://link.springer.com/chapter/10.1007/978-3-540-75254-7_4.

[71] P. Tolchinsky, U. Cortés, S. Modgil, F. Caballero and A. López-Navidad, Increasing human-organ transplant availability: Argumentation-based agent deliberation, *Institute of Electrical and Electronics Engineers Inc.* (2006).

[72] I. Vargas, I. Garcia-Subirats, A.S. Mogollón-Pérez, M. Ferreira-De-Medeiros-Mendes, P. Eguiguren, A.I. Cisneros et al., Understanding communication breakdown in the outpatient referral process in Latin America: A cross-sectional study on the use of clinical correspondence in public healthcare networks of six countries, *Health Policy and Planning* **33**(4) (2018), 494–504, available from: https://pubmed.ncbi.nlm.nih.gov/29452401/. doi:10.1093/heapol/czy016.

[73] S.L. Vasileiou, A. Kumar, W. Yeoh, T.C. Son and F. Toni, DR-HAI: Argumentation-based dialectical reconciliation in human–AI interactions, 2023, arXiv:2306.14694.

[74] A. Vassiliades, N. Bassiliades and T. Patkos, *Argumentation and Explainable Artificial Intelligence: A Survey*, Cambridge University Press, 2021, available from: https://doi.org/10.1017/S0269888921000011.

[75] D. Walton, Argumentation theory: A very short introduction, in: *Argumentation in Artificial Intelligence*, Springer US, Boston, MA, 2009, pp. 1–22, available from: http://link.springer.com/10.1007/978-0-387-98197-0_1.

[76] D. Walton, A dialogue system specification for explanation, *Synthese* **182**(3) (2011), 349–374, available from: https://link.springer.com/article/10.1007/s11229-010-9745-z.

[77] D. Walton and E.C.W. Krabbe, *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*, State University of New York Press, Albany NY, 1995, available from: https://books.google.com.pk/books?id=6nU8TpVmW08C.

[78] F.C. Wright, C. De Vito, B. Langer and A. Hunter, Multidisciplinary cancer conferences: A systematic review and development of practice standards, *European Journal of Cancer* **43**(6) (2007), 1002–1010. doi:10.1016/j.ejca.2007.01.025.

[79] F.C. Wright, N. Look Hong, D. Urbach, D. Davis, R.S. McLeod and A.R. Gagliardi, Multidisciplinary cancer conferences: Identifying opportunities to promote implementation, *Annals of Surgical Oncology* **16**(10) (2009), 2731–2737, available from: https://link.springer.com/article/10.1245/s10434-009-0639-6.

[80] R. Wu, P. Rossos, S. Quan, S. Reeves, V. Lo, B. Wong et al., An evaluation of the use of smartphones to communicate between clinicians: A mixed-methods study, *Journal of Medical Internet Research* **13**(3) (2011), available from: https://pubmed.ncbi.nlm.nih.gov/21875849/.

[81] L. Xiao, K. Hu and J. Fox, A group decision description language and its clinical application, in: *ACM International Conference Proceeding Series*, Association for Computing Machinery, 2021, pp. 160–167, available from: https://dl.acm.org/doi/10.1145/3488838.3488866.

[82] C. Zheng, X. Chen, L. Weng, L. Guo, H. Xu, M. Lin et al., *Benefits of Mobile Apps for Cancer Pain Management: Systematic Review*, JMIR Publications Inc., 2020, available from: https://mhealth.jmir.org/2020/1/e17055.