Explainable Argumentation for Wellness Consultation*

Isabel Sassoon[1], Nadin Kökciyan[1], Elizabeth Sklar[1], and Simon Parsons[1]

King's College London, Department of Informatics, WC2R 2LS, UK
{isabel.k.sassoon,elizabeth.sklar,nadin.kokciyan,simon.parsons}@kcl.ac.uk

Abstract. The work presented here explores the application of *computational argumentation* and *argumentation-based dialogue* to the domain of *clinical consultation*, with particular focus on self-management of chronic health conditions. Chronic conditions are characterised by regular and sometimes frequent monitoring of various biometric signs, along with a prescribed regimen of diet, exercise and medication. A variety of different types of exchanges can occur between a patient and their health care provider(s) when managing chronic conditions. In this paper, we propose the application of existing work on agent-based dialogue systems to support wellness consultation for a patient, as a supplement to typical periodic interactions with health care professionals. The system is designed to provide decision support and explanation for recommended actions. Explanations are supported by argumentation-based dialogues using a novel combination of argumentation schemes and explanation templates. The approach is illustrated through the use of a clinical scenario.

Keywords: Explainable AI · Computational Argumentation · Argumentation-based Dialogue

1 Introduction

1.1 Scope

In the work presented here, we explore the application of *computational argumentation* and *argumentation-based dialogue* to the domain of *clinical consultation*, particularly focusing on patient self-management of chronic health conditions. Such conditions are characterised by regular and sometimes frequent monitoring of various biometric signs, along with a prescribed regimen of diet, exercise and medication. A variety of different types of exchanges may occur between a patient and their health care provider(s). Here, we propose an agent-based system designed to support consultations that a patient may engage in as a supplement to typical periodic appointments with general practitioners and other professional health care providers that are required of patients with chronic

health conditions. We consider the types of interactions that a patient might have with an agent-based system that supports the patient's needs for such consultations, and we have identified three key functionalities that such a system could provide: (a) information, (b) recommendation, and (c) explanation.

1.2   Wellness consultations

A patient with chronic and/or multiple medical conditions will rely on periodical and regular interactions with their General Practitioner (GP), however there may be some questions or advice that is required that does not need to rely on a face to face meeting with their GP. These interactions would be complementary to regular contact with GPs but would offer the opportunities for some advice or recommendations to be obtained sooner. It is this type of asynchronous interaction through a dialogue that we define as a *wellness consultation* throughout this paper.

The first key function involves the patient querying the system and requesting *information* about their condition—for example, details about symptoms, treatment options and drug interactions—the kinds of information that appear in brochures provided by GPs, other health care professionals (HCPs) or pharmacists. This is characterised by a two-step interaction in which the patient initiates a query and the system retrieves the answer from a knowledge store and presents it to the patient.

The second function involves the patient asking the system to *recommend* a course of action, such as engaging in exercise activities tailored to the patient's condition, considering an over-the-counter medication such as *ibuprofen* to alleviate pain, or seeking further advice from an HCP. The distinction from the first function is an expectation that the patient will follow up their interaction with the system with some type of action; and so the system may later ask the patient if they have performed that action in order to help track their condition. While the initial query is initiated by the patient, the system could initiate the follow-up query to check later, e.g., if the patient has taken the recommended painkiller and is still experiencing pain or if the patient has taken the recommended exercise.

The third function involves the patient asking for clarification or *explanation* about a response given by the system in either of the first two cases. Where the first two functions are characterised by two-step interactions (one party initiates and the other responds), the third function involves an iterative process in which several queries and responses occur until the patient understands the information and/or recommendation offered by the system.

The first function is not an uncommon feature in today's IT[1]-rich society, with many people around the world being connected on mobile (and desktop) devices to internet sites that can offer health information. The second function is less common as a commercial product, but is certainly the focus of many IT research projects (including the one that led to this paper and those mentioned

---

[1] IT: Information Technology

human user	software agent	
	patient	*GP*
patient	self-care	consultation
GP	training	second opinion

Fig. 1. The different roles that a *human user* and *software agent* might play in an agent-based system designed to support self-management of chronic conditions.

in Section 5). The third function is even less common, both in commercial and research forms. The notion of *Explainable AI* has become a trend relatively recently (although many of the ideas and motivation behind Explainable AI are not so new, as pointed out in Section 5). The work presented here focuses on the third functionality—*explanation*—and proposes the use of *computational argumentation* and *argumentation-based dialogue* (see Section 2) as the means to implement the *explanation* functionality.

Consider a wellness consultation involving a conversation between a GP and a patient, where the patient is either being offered advice or is asking for some. Imagine an interactive agent-based system that could facilitate this conversation. Hypothetically, this "agent" could take either role: i.e., that of the patient or the GP, as illustrated by Figure 1. When the human user is a patient and the agent acts as a GP, then the interaction is an example of a *consultation* situation, as we have described above. When the human user is a patient and the agent also acts as a patient, then the interaction is akin to a conversation between peers in which the agent could offer reminders and encouragement regarding *self-care*. When the human user is a GP and the agent acts as a patient, then the interaction can be seen as a form of *training*, where the GP user could practice conversing with patients in order to confirm or practice the diagnostic process. When the human user is a GP and the agent also acts as a GP, then the interaction is similar to that of seeking a *second opinion*, or a case review. Note that these scenarios make assumptions about differing levels of expertise corresponding to the two possible user roles, assuming, with respect to medical knowledge, a relatively naïve patient and a more educated GP.

1.3 Contributions

In this paper, we explore the dialogue types relevant to wellness consultations as motivated by the CONSULT project[2], and propose dialogue templates for a set of types of interactions between the CONSULT system and the patient, through the use of a chatbot. In order to do this, we first explore the wider context of interactions between a human and an agent through dialogue, and use the framework introduced in [30] to map the possible interactions within Clinical Decision Support in general, and the CONSULT project in particular. Hence, in this initial approach we focus on the types of interactions between an agent acting

---

[2] `https://consult.kcl.ac.uk`

as the GP, and the patient as the human in the wellness consultation dialogue (This interaction is in the bottom left quadrant of the matrix in Figure 1).

The contributions of this paper are as follows: (1) we show how wellness consultations can be supported by existing types of human-agent argumentation-based dialogues; (2) we articulate how these argumentation-based dialogues can provide explanations, in particular by using a novel combination of argumentation schemes and explanation templates; and (3) we show the applicability of our approach on a clinical example from the CONSULT project. The paper is structured as follows: Section 2 provides background on argumentation, argumentation based dialogue and explainable AI. In Section 3, we outline our proposed approach. We then use an example from the CONSULT project in Section 4 to illustrate our approach. Section 5 briefly discusses related work and Section 6 summarizes the work and outlines our plans for future research.

2   Background

2.1   Computational Argumentation

Computational Argumentation [27] is a well-founded method which allows reasoning with incomplete and at times conflicting information or knowledge. An argument is structured so that it has a conclusion or a claim, and the support for the claim. When argumentation is employed in the context of decision support, its structure supports a human-like reasoning process where arguments' conclusions, their support, and the relationships between them, can be modelled. Argumentation has been extensively explored within the multi-agent community, and there are examples of its application to clinical decision support, which is the domain this work is focusing on. In Section 5 we discuss some of these applications.

An argument, $Arg = \langle S, c \rangle$, consists of a set of premises, S, defined in some language, $\mathcal{L}$, which support the conclusion, c. An *Argumentation Framework (AF)* [7] represents a set of Arguments $\mathcal{A}$, and the relationships between the members of the set. Formally an *AF* is a pair $\langle \mathcal{A}, \mathcal{R} \rangle$, where $\mathcal{A}$ is a set of arguments and $\mathcal{R}$ is binary relation representing *attack* relationships between arguments. For example if $Arg_1 = \langle p_1, c_1 \rangle$ and $Arg_2 = \langle p_2, \neg c_1 \rangle$ then an attack relation exists between Arg_1 and Arg_2 since these arguments have conflicting conclusions (i.e. rebuttal attack). Given a set of conflicting arguments, there are well-founded methods [7] for computing *extensions*, consistent sets of arguments which represent coherent opinions.

2.2   Argumentation-based Dialogue

In their seminal work, Walton and Krabbe [32] described six primary types of dialogue: *information seeking*, *inquiry*, *persuasion*, *negotiation*, *deliberation* and *eristic*. The distinguishing factors amongst these different types of dialogue are based on a participant's knowledge, their individual goals and goals that they

share with others. A logic-based formalism for modeling such dialogues as a formal *game* between agents was introduced in [16] and extended in [30]. This formalism supports the combination of multiple dialogues. [22] examines a subset of these dialogue types and shows how their properties depend on the behaviour of theagents engaging in the dialogue. The different types of dialogues include:

- *information seeking* [32], where one participant asks a question that she does not know the answer to and believes the other participant can answer;
- *inquiry* [15], where both participants seek an answer to a question that neither knows the answer to;
- *persuasion* [24], where one participant tries to alter the beliefs of another participant;
- *negotiation* [26], where participants bargain over the allocation of a scarce resource;
- *deliberation* [17], where participants decide together on taking a particular action;
- *eristic* [32], where participants quarrel verbally;
- command [9], where one participant tells another what to do;
- chance discovery [13], where a new idea arises out of the discussion between participants; and
- verification [5], where one participant asks a question that she already knows the answer to and she believes the other participant also knows the answer, so her aim is to verify her belief.

In the context of wellness consultation interactions it can be seen that not all dialogue forms will be relevant. The types of dialogues of most relevance are those related to deliberation and persuasion. An application of some of these dialogue types in the context of human robot interaction is outlined in [30].

2.3   Argumentation Schemes

An *argumentation scheme or argument scheme* (AS) is a semi-formal reasoning template that matches common reasoning patterns in real life. An *Argumentation Scheme* is a model for instantiating arguments within a specific context and is used to provide a formal basis for instantiating arguments and defining their internal structure. An AS consists of a set of support premises (S), which support the conclusion premise, c, necessary for this derivation [20,21,23,33].

Formally: An argument scheme $AS = \langle S, c, \mathcal{V} \rangle$ consists of the set of premises, S, which support a conclusion, c, and are instantiated with the set of variables, $\mathcal{V} = S.V \cup c.V$.

One of the key features of argumentation schemes is the list of associated critical questions (CQs). The claim or conclusion that the scheme supports is presumptive and the claim is withdrawn unless the CQs posed have been answered successfully. The instantiation of the appropriate argumentation scheme and its associated CQs is a way of generating a set of arguments that can then be reasoned with as an argumentation framework. This mechanism ensures that only arguments that have not been defeated by the CQs will be generated, furthermore CQs can also point to additional arguments to consider.

3   Our Approach

Our approach to supporting explanation functionality as part of a wellness consultation between a human and an agent involves three steps: (1) defining an *argumentation scheme* specific to the provision of health related treatments or actions; (2) identifying existing, and possibly defining new, *argumentation-based dialogues* that are appropriate for this domain; and (3) showing how these schemes in conjunction with the argumentation-based dialogues can be used by the agent to provide *explanation* to patients (human users). In this section, we describe each of these elements to our approach. Then following, Section 4, we provide concrete examples of how our approach is (will be) implemented.

Our approach relies on a method to generate recommendations about a suitable treatment or action required to attain a clinical goal. For this purpose we propose the use of a clinically specialized argumentation scheme in order to instantiate and reason with all the possible relevant options. Reasoning with set of generated arguments can be achieved using a variety of approaches such as [7], which will result in an extension containing arguments in support of a treatment or an action. These will then form the basis of the elements of the dialogues and will leverage a set of explanation templates to ensure they can be communicated clearly to the user. The dialogue protocols we propose dictate the order and options the dialogue can progress in.

3.1   Argumentation Scheme for Proposed Treatment (ASPT)

The domain of relevance to this approach involves recommending a course of action in the clinical context, we therefore need to employ an argumentation scheme that is specialized to this domain. We propose to use the *Argumentation Scheme for proposed treatment (ASPT)* [11], which is a specialization of *argumentation scheme from practical reasoning ASPR* [1]. There are undoubtedly additional argumentation schemes that can be used to support dialogues in the clinical context, in the future we will explore the use of additional schemes as part of dialogues related to wellness consultations.

Within our proposed approach the wellness consultation between a patient and the agent to deliberate about possible actions or treatments is underpinned by ASPT. ASPT is in Figure 2. The arguments instantiated by ASPT are in support of different treatments or actions.

ASPT
$p1$ - Given the patient Facts F
$p2$ - In order to realise the goal G
$p3$ - Treatment T promotes the goal G
therefore : Treatment T should be considered

Fig. 2. Argumentation scheme for a proposed treatment [11].

ASPT is subject to a set of critical questions (CQs), these can be the source of additional or counter-arguments to the arguments instantiated by this scheme. There is no general agreed method for structuring AS and their CQs as such when outlining ASPT and its CQs we made use of the applicable premises and CQs from ASPR as a starting point and specialized them to this clinical setting. We also made a distinction in separating the premises of the scheme and the critical questions by the nature of the information they require. The premises of the scheme rely on facts and information unlikely to change such as clinical guidelines or a specific patient's demographics, whilst critical questions look more at the patient's specific clinical history. The list of critical questions outlined herein is a subset of the CQs proposed for this Argument Scheme. We are using this subset of three critical questions to illustrate our approach, as the formalization of ASPT is not the focus of this paper. The subset of critical questions for the argument scheme ASPT are in Figure 3.

Critical Questions for ASPT
[CQ1] Has treatment T been unsuccessfully used on the patient in the past?
[CQ2] Has treatment T caused side effects on the patient?
[CQ3] Given patient facts F, are there any counter-indications to treatment T?

Fig. 3. Critical questions for ASPT [11].

Given a goal that needs to be realised, instantiating ASPT and its critical questions will result in an Argumentation Framework. Reasoning with this framework will generate extensions that contain treatment suggestions. Such an approach is outlined in [11] and [2]. When exploring options to explain any recommendations made, then all the relevant arguments and their critical questions will be possible sources of explanations.

The aim of the dialogue as part of a wellness consultation is for the agent to propose an action or treatment to the patient, and to allow the patient to query the rationale underlying the recommendation made, if they wish to do so. In the next section we model the possible situations that will provide a framework for the possible dialogues required. These rely on the beliefs that the human will have in respect to the action or treatment proposed by the agent.

3.2 Dialogue

In the context of a wellness consultation, we assume that when a dialogue takes place it will be regarding a specific action, for example one supported by an argument instantiated by ASPT. We assume that the goal that underlies the recommendation is mutually agreed upon by the human and the agent. In other words the dialogue is not about the goal, but about the actions to take to achieve the goal. This is a strong assumption because, in general, there will be many interactions in which it is necessary to discuss the goal — the agent may need to

elicit the goal from the human or persuade them to adopt a goal. However, we think it is a reasonable assumption in our current work since we are only focused on providing recommendations when the human specifically asks for one — it seems reasonable that a user who has asked for a recommendation for a course of action will already have committed to taking that course of action — and we defer considering more general scenarios to future work.

The agent has a known set of beliefs $Ag.\Sigma$, and the human may also have a set of beliefs. $Ag.\Gamma(H)$ represents the agent's beliefs about the human's belief. So far this uses a notation similar to [30]. b represents a belief about an action, this can be an action recommended by instantiating ASPT. For example in case of treating a condition with a specific drug then the action is $b = \mathbf{offer}(arb, hbp)$ (where arb is a treatment for *high blood pressure (hbp)*). $\neg b$ is a disbelief in that action, and $?b$ is a situation where there is no information about b. The possible dialogues that can occur in the context of a wellness consultation when the agent is the GP and the human is the patient are in Figure 4.

	$b \in Ag.\Gamma(H)$	$\neg b \in Ag.\Gamma(H)$	$?b \in Ag.\Gamma(H)$
$b \in Ag\Sigma$	case 1 agreement - (no dialogue)	case 2 disagreement - (persuasion)	case 3 agreement + explanation - (information seeking)
$\neg b \in Ag\Sigma$	case 4 disagreement - (persuasion → deliberation)	case 5 agreement - (deliberation)	case 6 (deliberation)
$?b \in Ag\Sigma$	case 7 (deliberation)	case 8 (deliberation)	case 9 (deliberation)

Fig. 4. The space of possible dialogues between agent and uer. The agent, Ag, assumes the role of GP and the user, H, is the patient.

The following situations are relevant in the context of an agent acting as the GP and the human being a patient:

- *Agreement*: Both parties' beliefs do not conflict. There is no need for a dialogue on the specific action in question. However if the agreement is against the action (i.e. both agree $\neg b$), and there is a need to consider an alternative action then this may lead to a deliberation dialogue (case 5).
- *Disagreement*: Beliefs are in conflict (such as in case 2 and 4), the type of dialogue to initiate depends on the user's expertise. If the user is an expert (another GP) or a layperson in medicine (a more plausible scenario) then the dialogue burden of proof and dynamics will vary. In this initial approach we assume that the patient's expertise is less trusted compared to the GP, therefore in this case the sgent (as the GP) will initiate a persuasion dialogue. However there may be different levels of disagreement such as the patient not just disagreeing with the proposed action, but proposing a different action instead.
- *Deliberation (referral)*: If the agent does not believe in the action being discussed or has no information on it, then this requires an alternative action

to be agreed upon. Hence a deliberation dialogue, which may require a nested persuasion dialogue [14].

The dialogue flow is made up of three basic building blocks, an overall *deliberation* dialogue, and depending on the locutions of the human, this can then nest either a *persuasion* dialogue or a *explanation (or follow up) dialogue* dialogue. The latter two are similar but their suitability to a given situation depends on whether the locution of the human was one of agreement/acceptance (in which case this is a *explanation (follow-up) dialogue*. Should the human locution be confrontational then this would be more suitably addressed by a *persuasion dialogue*. This flow is illustrated in Figure 5.
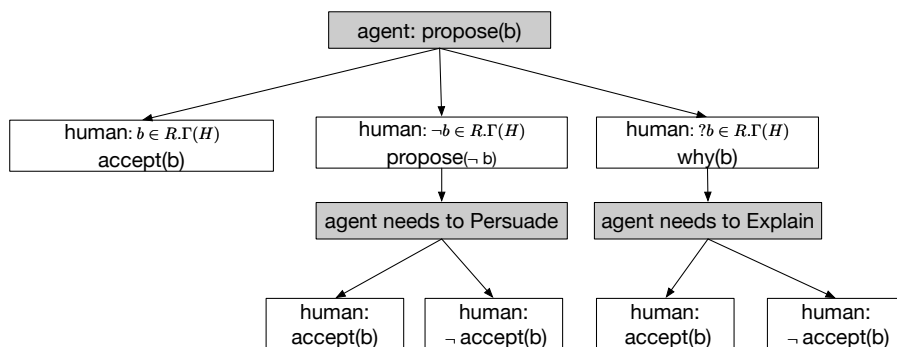

Fig. 5. The possible dialogue options matching the top row of Table 4

3.3 Explanation Templates

An important element of our approach is the *explanation* functionality as part of the *Argumentation-based Dialogue*. We propose the use of *explanation templates* which will be defined for the argument scheme as well as for its critical questions. These templates are specific to the reasoning and specialization of the scheme, and include placeholders for the actual instantiated variables specific to a given application of the scheme. A sample set of templates are given in Figure 6.

These templates work as follows. Given a goal G, patient facts F and a knowledge base of treatments, when a dialogue results in the instantiation of ASPT, the relevant critical questions are also instantiated, resulting in an argumentation framework that can be solved to establish a set of extensions. Each element in an extension is mapped to their source (the AS or CQ it was derived from) and their instantiated values. The templates in Figure 6 can then be used to create explanations — for example if a user queries why a particular treatment T is recommended, the explanation template for ASPT can be instantiated with the relevant values of G and F (see below).

AS or CQ	Explanation template
ASPT	Treatment T should be considered as it promotes goal G, given patient facts F
CQ1	Treatment T should not be considered as it was not effective for this patient in the past
CQ2	Treatment T should not be considered as it caused side effects for this patient in the past
CQ3	Treatment T should not be considered as patient fact $f_i \in F$ is a counter-indication to its use

Fig. 6. Mapping of argument schemes and critical questions to explanation templates.

4   Example from the Medical Domain

In order to illustrate the approach proposed in this paper we make use of an example interaction between the agent and the patient that arises in the context of the CONSULT project. The aim of the CONSULT project is to develop and test the feasibility of a collaborative mobile decision support system to help patients who suffer from chronic diseases and multiple co-morbidities to manage their treatment. The prototype system integrates data from wellness sensors, electronic health records and relevant guidelines to support data-backed argumentation based decision support. This is accessible to the patient via a mobile app that includes a dashboard and a chatbot. In this example we assume that the agent is acting the role of the GP, and the human is the patient (as outlined in Figure 1).
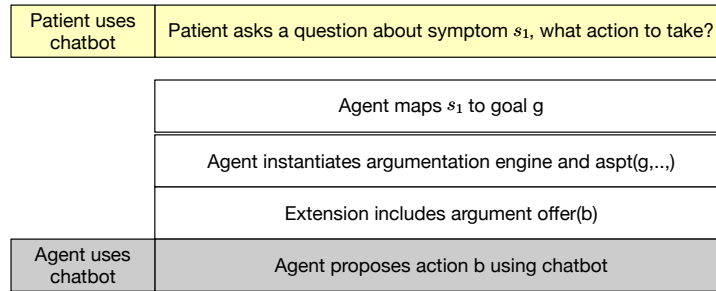
Patient uses chatbot	Patient asks a question about symptom s_1, what action to take?
	Agent maps s_1 to goal g
	Agent instantiates argumentation engine and aspt(g,..,)
	Extension includes argument offer(b)
Agent uses chatbot	Agent proposes action b using chatbot

Fig. 7. The steps to start the dialogue when a patient asks for advice

4.1   Patient Asks for Advice

This interaction will be one initiated by the patient when they ask for a recommendation, and we will assume we have a fictitious patient called Bob. In this scenario the patient (Bob) asks a question or seeks advice regarding a specific

symptom s_1 via a chatbot dialogue. The initial processes generated by this are in Figure 7. Figure 7 illustrates the steps taken before the dialogue commences. These are the mapping of the symptom $s1$ to a goal g, the instantiation of the Argumentation Scheme for Possible Treatment (ASPT) using the relevant variables (one of which is g). The mappings between symptoms and goals are assumed to be in the agent's knowledge base. Reasoning with all the arguments results in an extension including an action (for example b) will form the basis for the dialogue. A view of a dialogue options that are possible after a recommendation is made by the agent is depicted in Figure 8. Note that we assume the argumentation engine constructs an argument based on its knowledge base and the information it received from external sources, such as the patient's EHR, sensor data and relevant clinical guidelines.
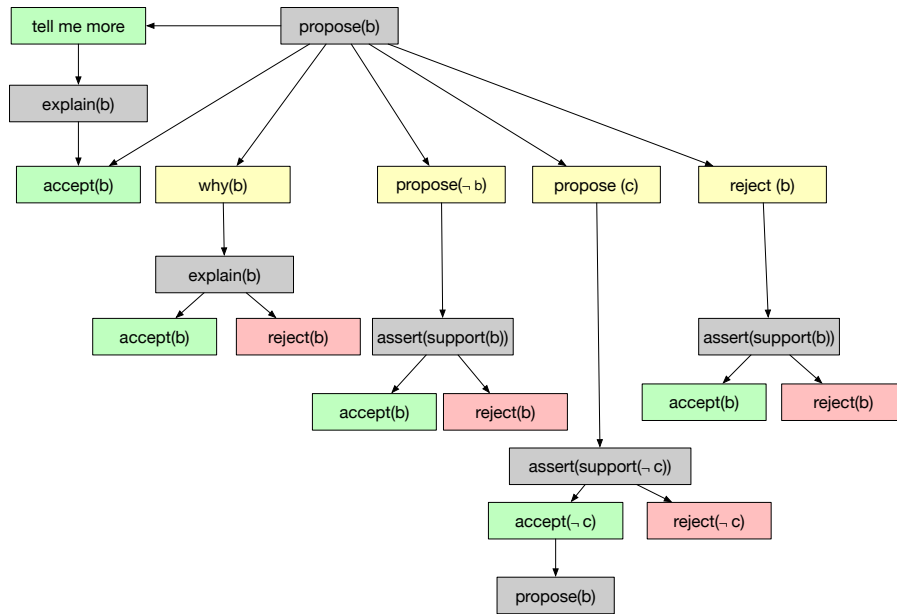

Fig. 8. A possible patient asks for advice dialogue tree

In this dialogue when Bob asks for a recommendation, Bob uses the agent to self manage pain by initiating a wellness consultation. The protocol for this dialogue, split into its possible branches, for this dialogue is illustrated in Figures 9, 10 and 11. This illustration of dialogue protocol is the same format used in [30]. These Figures include the different possible ways in which the dialogue can evolve and the elements of the arguments, critical questions and explanations used within these.

The instantiation of ASPT and its related CQs in this case results a recommendation for *ibuprofen* as a painkiller. The agent proposes the recommended action to Bob (Figure 9), and Bob can respond in a few ways. Bob can: simply accept the recommendation; ask for more information (Figure 10); reject the recommendation (Figure 10); or propose an alternative action (Figure 11).

Should Bob ask for more explanations (by asking why) then this would trigger an explanation dialogue that would initially outline the reasoning within the argument in support of the use of *ibuprofen* using the template in Figure 6. An example dialogue within the flow of Figure 10 where Bob asks for more information (The elements in parenthesis are not part of the actual dialogue):

- *agent*: It is recommended that you take Ibuprofen
- *bob*: Why should I take Ibuprofen?
- *agent*: Ibuprofen (Treatment T) should be considered as it promotes back pain relief (goal G) given your clinical history (Bob patient facts F).

    Here the agent is instantiating the explanation template from Figure 6 that matches ASPT.

Bob would then be able to accept the recommendation, or probe further. In the latter case the critical questions would be employed in turn to further provide rationale for this recommendation.

Another example of a possible dialogue should Bob propose an alternative pain killer as illustrated in Figure 11:

- *bob*: Should I take Codeine?
- *agent*: Codeine (Treatment T) should not be considered as it caused you side effects in the past.

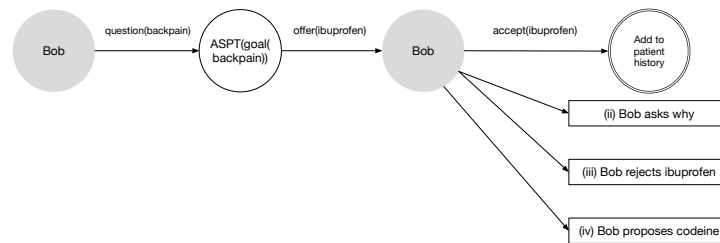   Here the agent is using the explanation template from Figure 6 that matches CQ2.

Fig. 9. Start of the Dialogue protocol using approach in [30]
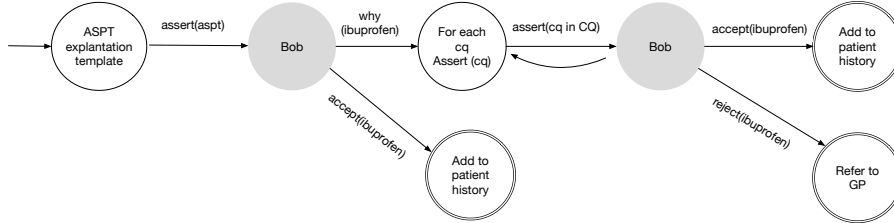
Fig. 10. (ii) Explanation Dialogue: Why branch of Dialogue protocol and (iii) Persuasion Dialogue branch of Dialogue protocol using approach in [30]
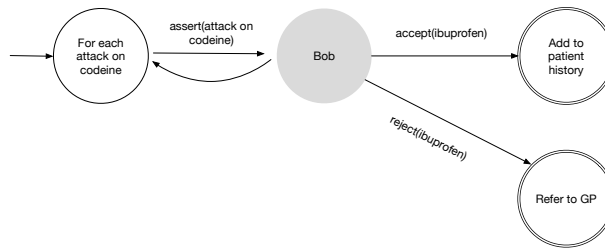

Fig. 11. (iv) Persuasion Dialogue on alternative action branch of Full Dialogue protocol using approach in [30]

5   Related Work

This section succinctly outlines some of the work that relates to the elements proposed in this paper. This includes work on argumentation based clinical decision support, argumentation-based dialogues and explainable AI.

While "explainable AI" is a recently coined term, work on explanation has deep roots within the field of AI. As Van Lent *et.al* argue, such work goes back at least to the 1970s through the requirement of expert systems, specifically clinical ones, to be able to supply a justification for the recommendations that they made [29]. Despite this long history, it is only recently that the AI community has identified explainability as a key challenge problem [10]. For example, suggesting that it will be a requirement for the adoption of AI-based systems (such as deep learning systems) that they are able to explain the reasoning process rather than just recommending some actions to take.

The topic of what constitutes an explanation is broadly addressed by Miller *et al.* [18,19] who describes several approaches taken in the literature such as *explanation selection* where a set of reasons are chosen to explain a decision. This overview suggests several areas for investigation including *explanation as conversation*. Such an approach is closely related to what we discuss here where we model human-agent interactions for conveying explanations to users.

Turning to the more specific topic of argumentation-based (specifically argumentation scheme-based) systems for clinical decision support, we find several

systems that are close to what we describe. In StAR [8,12] argumentation was used as a means to support reasoning about risk in the absence of numerical estimates of uncertainty. This was in the context of reasoning about the carcinogenic risk of different chemicals. The RAGs approach [6] leverages arguments as a way of visually representing the different reasons underpinning different possible treatment actions for a given patient. The DRAMA agent proposed in [2] uses argumentation schemes to generate arguments used to reason about a treatment, but this work does not touch on explanations or offer a dialogue based interface. Similarly the *Carrel+* project [31], where the objective of the argumentation based tool was to supervise and validate the deliberation process on organ transplant viability. In *Carrel+*, arguments were generated based on argumentation schemes and the goal was not explaining decisions but agreeing on a decision. The Argumentation based decision support approach proposed in the *CISpaces* project in [4] presents an explanation template, and there is also some groundwork for explainability from argumentation frameworks articulated in [25]. Finally, we note that [3] also points out that there is a close relationship between dialogue and explanation.

6   Conclusions

We have presented a novel approach that makes use of argumentation-based dialogues and explanation templates to implement the explanation functionality in the context of wellness consultations between an agent acting as a GP and the patient as the human. Our contributions include modeling the types of human-agent dialogues of relevance to wellness consultations, articulating an approach to providing explanations as part of an argumentation-based dialogue and illustrating the applicability of this approach on a clinical example related to patient self management.

In future work, we plan to extend this approach to additional types of interactions as sketched in Figure 1. Here we modelled interactions where a human plays the role of patient and a software agent plays the role of GP. Other interactions include those in which both agent and human play the role of patient, and so the two interact as peers. In this work we made some assumptions on the differing level of knowledge of the user as compared to the agent, and in peer to peer interactions this would not be the case in general. Thus this future work will explore the effect of the hierarchy of levels of knowledge on the dialogue models. In addition, our plans for future work recognise that reasoning about what possible treatments should be adopted in a clinical context involves more than one argumentation scheme. Thus we will study additional schemes. Indeed, we are currently in the process of mapping the general process of deciding on treatments for stroke to existing argumentation schemes, specializing these schemes where appropriate and highlighting gaps where existing argumentation schemes do not cover the necessary reasoning. This will provide an exhaustive set of schemes covering medical treatment relevant to stroke. This exhaustive

set of stroke treatment schemes will then drive the development of an associated set of explanation schemes and their associated critical questions.

These dialogues are being implemented in a chatbot, and we will evaluate the approach as part of a user study with stroke patients. This study will allow us to assess the feasibility of the approach and establish, qualitatively, how patients interact with such a system. We will be evaluating this consult app as part of a user study in the near future. The dialogue system is being developed in conjucntion with a GP, but it would be desirable to test this interaction with a wider group GPs to assess whether these protocols align with the conversations GPs have with patients.

Other future work will concentrate on critical questions. In this paper we have articulated the role of critical questions in the reasoning about recommended actions and the rationale for these actions. In future work we will consider the role of critical questions both when they attack an argument and when they are satisfied and therefore do not generate an attack. Explanations provided within the dialogues should be able to explain both outcomes for each critical question. This will facilitate situations where argumentation schemes are associated with different critical questions and the answer to the critical questions may change over time.

Finally, we will consider alternative ways to create arguments. Here we assumed that all the arguments supporting a recommended action or treatment were constructed using the clinical-specific argumentation scheme ASPT, and we have already discussed expanding the set of argument schemes to cover a wider set of situations. However, this is not the only way to create arguments. For example, arguments can be mined from data [25,28], and, in the long term, we would like to see the arguments used by a system like ours being mined from the medical literature.

References

1. Atkinson, K., Bench-Capon, T.: Practical reasoning as presumptive argumentation using action based alternating transition systems. Artificial Intelligence **171**(10-15), 855–874 (2007)
2. Atkinson, K., Bench-Capon, T.J.M., Modgil, S.: Argumentation for decision support. In: Database and Expert Systems Applications (DEXA). pp. 822–831 (2006)
3. Bex, F., Walton, D.: Combining explanation and argumentation in dialogue. Argument & Computation **7**(1), 55–68 (2016)
4. Cerutti, F., Norman, T.J., Toniolo, A., Middleton, S.E.: CISpaces.org: from fact extraction to report generation. Computational Models of Argument: Proceedings of COMMA 2018 pp. 269–280 (2018)
5. Cogan, E., Parsons, S., McBurney, P.: What kind of argument are we going to have today? In: Proceedings of the 4th International Joint Conference on Autonomous Agents and Multiagent Systems. pp. 544–551. ACM (2005)

6. Coulson, A., Glasspool, D., Fox, J., Emery, J.: RAGs: A novel approach to computerized genetic risk assessment and decision support from pedigrees. Methods Of Information in Medicine **40**(4), 315–322 (2001)
7. Dung, P.M.: On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games. Artificial Intelligence **77**(2) (1995)
8. Fox, J.: Will it happen? can it happen? a new approach to formal risk analysis. Risk Decision and Policy **4**(2), 117–128 (1999)
9. Girle, R.: Commands in Dialogue Logic. In: Gabbay, D.M., Ohlbach, H.J. (eds.) Practical Reasoning: Proceedings of the First International Conference on Formal and Applied Practical Reasoning (FAPR), vol. Lecture Notes in Artificial Intelligence 1085, pp. 246–260. Springer, Berlin, Germany (1996)
10. Gunning, D.: Explainable Artificial Intelligence (XAI). Defense Advanced Research Projects Agency (DARPA) (2017)
11. Kokciyan, N., Sassoon, I., Young, A.P., Chapman, M., Porat, T., Ashworth, M., Curcin, V., Modgil, S., Parsons, S., Sklar, E.: Towards an argumentation system for supporting patients in self-managing their chronic conditions. In: Proceedings of the AAAI Joint Workshop on Health Intelligence (2018)
12. Krause, P., Fox, J., Judson, P., Patel, M.: Qualitative risk assessment fulfills a need. In: Applications of Uncertainty Formalisms, pp. 138–156. Springer (1998)
13. McBurney, P., Parsons, S.: Chance discovery using dialectical argumentation. In: Proceedings of the Workshop on Chance Discovery, Fifteenth Annual Conference of the Japanese Society for Artificial Intell igence. Matsue, Japan (2001)
14. McBurney, P., Hitchcock, D., Parsons, S.: The eightfold way of deliberation dialogue. International Journal of Intelligent Systems **22**(1), 95–132 (2007)
15. McBurney, P., Parsons, S.: Representing epistemic uncertainty by means of dialectical argumentation. Annals of Mathematics and Artificial Intelligence **32**(1–4), 125–169 (2001)
16. McBurney, P., Parsons, S.: Games that agents play: A formal framework for dialogues between autonomous agents. Journal of Logic, Language and Information **11**(3), 315–334 (2002)
17. McBurney, P., Parsons, S.: A denotational semantics for deliberation dialogues. In: Proceedings of the 3rd International Conference on Autonomous Agents and Multi-Agent Systems. IEEE Press (2004)
18. Miller, T.: Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence (2018)
19. Miller, T., Howe, P., Sonenberg, L.: Explainable AI: Beware of inmates running the asylum. In: Proceedings of the IJACI Workshop on Explainable AI (2017)
20. Parsons, S., Atkinson, K., Li, Z., McBurney, P., Sklar, E.I., Singh, M., Haigh, K., Levitt, K., Rowe, J.: Argument schemes for reasoning about trust. Argument & Computation, Special Issue: Trust, Argumentation and Technology **5**(2–3) (2014)
21. Parsons, S., McBurney, P., Sklar, E., Wooldridge, M.: On the relevance of utterances in formal inter-agent dialogues. In: Proceedings of the 6th International Conference on Autonomous Agents and Multiagent Systems (2007)
22. Parsons, S., Wooldridge, M., Amgoud, L.: Properties and complexity of some formal inter-agent dialogues. Journal of Logic and Computation **13**(3), 347–376 (2003)
23. Parsons, S., Wooldridge, M., Amgoud, L.: On the outcomes of formal inter-agent dialogues. In: Proceedings of the 2nd International Conference on Autonomous Agents and Multi-Agent Systems. ACM Press, New York (2003)
24. Prakken, H.: Formal systems for persuasion dialogue. The Knowledge Engineering Review **21**(02), 163–188 (2006)

25. Rago, A., Cocarascu, O., Toni, F.: Argumentation-based recommendations: Fantastic explanations and how to find them. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence. pp. 1949–1955 (2018)
26. Rahwan, I., Ramchurn, S.D., Jennings, N.R., Mcburney, P., Parsons, S., Sonenberg, L.: Argumentation-based negotiation. The Knowledge Engineering Review **18**(4), 343–375 (2003)
27. Rahwan, I., Simari, G.R.: Argumentation in artificial intelligence, vol. 47. Springer (2009)
28. Rajendran, P.: Aggregating and Analysing Opinions for Argument-based Relations. Ph.D. thesis, University of Liverpool, Liverpool (June 2019)
29. Shortliffe, E.H., Davis, R., Axline, S.G., Buchanan, B.G., Green, C.C., Cohen, S.N.: Computer-based consultations in clinical therapeutics: explanation and rule acquisition capabilities of the mycin system. Computers and Biomedical Research **8**(4), 303–320 (1975)
30. Sklar, E.I., Azhar, M.Q.: Argumentation-based dialogue games for shared control in human-robot systems. Journal of Human-Robot Interaction **4**(3), 120–148 (2015)
31. Tolchinsky, P., Cortes, U., Modgil, S., Caballero, F., Lopez-Navidad, A.: Increasing human-organ transplant availability: Argumentation-based agent deliberation. IEEE Intelligent Systems **21**(6) (2006)
32. Walton, D., Krabbe, E.C.W.: Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning. State University of New York Press, Albany, NY (1995)
33. Walton, D., Reed, C., Macagno, F.: Argumentation Schemes. Cambridge University Press (2008)