Cognitive Planning in Motivational Interviewing

Emiliano Lorini[1], Nicolas Sabouret[2], Brian Ravenet[2], Jorge Fernandez Davila[1] and Céline Clavel[2]
[1]*IRIT-CNRS, Toulouse University, France*
[2]*Université Paris-Saclay, CNRS, Laboratoire Interdisciplinaire des Sciences du Numérique, Orsay, France*
lorini@irit.fr, jorge.fernandez@irit.fr, {firstname.lastname}@universite-paris-saclay.fr

Keywords: Artificial Intelligence, Automated Planning, Motivational Interviewing, Human-Computer Interaction, Health Coaching

Abstract: This paper presents a cognitive planning model that implements the principles of motivational interviewing, a counseling method used to guide people in adopting behavior changes. This planning system is part of a wider dialogical architecture of artificial counseling agent. We present the formal model and planning problem. We show how it can be used to plan for dialogue in the architecture. We illustrate its functionalities on a simple example.

1 INTRODUCTION

Motivational Interviewing Motivational interviewing (for short MI) is a counseling method used in clinical psychology for eliciting behavior change (Lundahl and Burke, 2009). One crucial aspect of MI consists in exploring the participant's subjectivity through open questions to identify her desires and personal values (*e.g.*, conformity, independence, carefulness, etc.) (Miller and Rollnick, 2012). This exploration allows the participant to become aware of the inconsistency between her desires or personal values (*e.g.*, being in good health), and her current behavior (*e.g.*, not doing enough physical activity). However, MI does not necessarily try to induce beliefs about positive aspects of the behavior change (*e.g.*, most people are already aware that reasonable physical activity is good for health and would like to practice a sport regularly). It rather helps the participant to identify the reasons why she did not convert her mere desires (*e.g.*, I would like to practice a sport) into intentions (*e.g.*, I commit to do sport regularly) and reassures her that these limitations can be overcome. To this aim, the counselor rephrases the ideas expressed by the participant so as to provoke reflections about the connection between her beliefs, desires and intentions.

Several automated MI systems have been proposed in recent times (da Silva et al., 2018; Kanaoka and Mutlu, 2015; Lisetti et al., 2013; Olafsson et al., 2019; Schulman et al., 2011). However, all these systems use predefined dialogue trees to conduct the MI. In this paper, we propose a model based on cognitive planning for driving MI in a human-agent interaction system.

Cognitive planning. Classical planning in artificial intelligence (AI) is the general problem of finding a sequence of actions (or operations) aimed at achieving a certain goal (Ghallab et al., 2004). It has been shown that classical planning can be expressed in the propositional logic setting whereby the goal to be achieved is represented by a propositional formula (Bylander, 1994). In recent times, epistemic planning was proposed as a generalization of classical planning in which the goal to be achieved can be epistemic, *i.e.*, the goal of inducing a certain agent to believe or to know something (Bolander and Andersen, 2011; Löwe et al., 2011). The standard languages for epistemic planning are epistemic logic (EL) (Halpern and Moses, 1992) and its dynamic extension, the so-called dynamic epistemic logic (DEL) (van Ditmarsch et al., 2007). A variety of epistemic logic languages and fragments of DEL with different levels of expressivity and complexity have been introduced to formally represent the epistemic planning problem and efficiently automate it (see, *e.g.*, (Muise et al., 2015; Muise et al., 2021; Kominis and Geffner, 2015; Cooper et al., 2016; Cooper et al., 2021)).

In a recent paper (Fernandez et al., 2021), cognitive planning was introduced as a further generalization of epistemic planning. In cognitive planning, it is not only some knowledge or belief state of a target agent that is to be achieved, but more generally a cognitive state. The latter could involve not only knowledge and beliefs, but also desires, intentions and, more generally, motivations. The cognitive planning (CP) approach is well-suited for modeling interaction whereby an agent tries to trigger attitude change in another agent through the execution of a sequence of speech acts. CP takes into consideration resource boundedness and limited rationality of the interlocutor agent. This makes CP a very well-suited model for implementing motivational interviewing in human-machine interaction (HMI) applications in which an artificial agent is expected to interact with a human — who is by definition resource-bounded — through dialogue and to induce her to behave in a certain way.

Motivational interviewing is composed of several stages: prior to having the participant change her intentions, one has to make her aware of the inconsistencies between her desires and her actual behavior. The artificial agent has both (i) a model of the human's overall cognitive state, including her beliefs and intentions, and (ii) a goal towards the human's mental attitudes, *e.g.*, the goal of making the human aware of the inconsistency between her desires and her actual behavior. Given (i) and (ii), it tries to find a sequence of speech acts aimed at modifying the human's cognitive state thereby guaranteeing the achievement of its goal.

Outline. The aim of this paper is to explain how to situate the cognitive planning module in a general architecture of an artificial agent which is expected to interact with a human user through dialogue and to motivate her to behave in a certain way or to change/adopt a certain style of life through motivational interviewing methods. In Section 2, we provide a birds-eye view of the architecture. In Section 3, we present the formal framework on which the cognitive planning approach is built. In Section 4, a variety of cognitive planning problems are formalized. Section 5 is devoted to describe the belief revision module of the architecture. Finally, in Section 6, the cognitive planning problem is instantiated in a concrete example of motivational interviewing between an artificial agent and a human.

2 General Architecture

The general architecture of the system is detailed in Figure 1.

Data structures The artificial planning agent, that for simplicity we call the machine, is endowed with three kinds of data structure: its belief base, the goal to be achieved and the repertoire of speech acts (or communicative actions) it can perform. We assume the machine's action repertoire includes two types of speech act: assertions and questions. The machine can have persuading goals, aimed at changing the human's beliefs, or influencing goals, aimed at inducing the human to form a certain intention or to behave in a certain way. The machine's belief base includes both information about the environment and information about the human's overall cognitive state and its way of functioning. In other words, the machine has a theory of the human's mind. The machine's belief base evolves during its dialogue with the human.

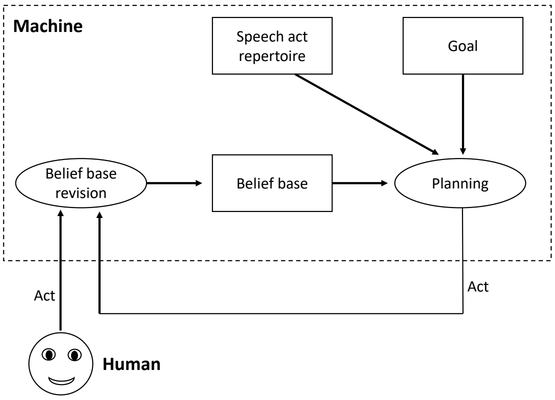

Figure 1: General architecture

Interrogative and informative phase The interaction between the machine and the human is structured in two phases the *interrogative* (or *exploratory*) phase and the *informative* phase. In the interrogative phase the machine gathers information about the human's cognitive state. This includes information about the human's beliefs, desires and preferences. The interrogative phase is identified with a sequence of questions by the machine to the human. The informative phase is the core of the influence process. In this phase, the machine performs a sequence of assertions aimed at modifying the human's cognitive state (her beliefs and/or intentions). The interrogative phase is propaedeutic to the informative phase. Indeed, for the machine to be able to lead the human to change her behavior, it must have information about the human's cognitive state. Such an information is acquired during the interrogative phase. In this work, we assume that the two phases are unified at the planning level: the machine includes in its plan not only the assertions but also also the questions. In particular, the machine has to find a sequence of questions followed by a sequence of assertions such that, for some possible answer by the human, the composition of the two sequences guarantees that the persuading or influencing goal will be achieved. It is reasonable to assume that the machine first tries to find a plan with only assertions. (why asking questions to the human if what the machine knows about the human's cognitive state is already sufficient to persuade or influence her). However, in most cases, the machine has uncertainty and lacks information about the human's cognitive state so that it must ask questions to the human before trying to induce her attitude change. In Section 6, we will show how some aspects of the motivational interviewing (MI) methodology can be naturally captured in the two phases of the cognitive planning approach.

Execution of the plan After having selected a plan, the machine executes it. The machine can either execute the entire plan or execute it one piece after the other by waiting the reply of the human before executing the next piece. We assume that how the plan is executed depends on the application under consideration and on the type of speech act in the plan to be executed. It is reasonable to suppose that when executing the interrogative part of the plan, the machine asks a single question at each step and waits the answer by the human before moving to the next question. After each question by the machine, the human gives an answer and the machine expands or revises, when necessary, its belief base accordingly. Indeed, the information provided by the human in response to the machine's question can enrich the machine's belief base with new facts about the environment (objective facts) or about the human's cognitive state (mental facts) or make the machine's belief base inconsistent. In the latter case, the machine must revise its belief base after having incorporated the new information.

3 Formal Framework

In this section, we present the epistemic language on which the cognitive planning approach is based. The language is a two-agent fragment of the multi-agent epistemic language presented in (Lorini, 2020). The language distinguishes explicit from implicit belief: an agent's belief of explicit type is a piece of information contained in the agent's belief base, while a belief of implicit type corresponds to a piece of information that is derivable from the agent's belief base.

Assume a countably infinite set of atomic propositions *Atm* and a finite set of agents $Agt = \{\mathfrak{h}, \mathfrak{m}\}$, with $\mathfrak{h}$ denoting the human and $\mathfrak{m}$ the machine. The language is defined in two steps. First, the language $\mathcal{L}_0$ is defined by the following grammar in BNF:

$$\alpha \quad ::= \quad p \mid \neg\alpha \mid \alpha_1 \wedge \alpha_2 \mid \triangle_i \alpha,$$

where p ranges over *Atm* and i ranges over *Agt*. $\mathcal{L}_0$ is the language for representing agents' explicit beliefs. The formula $\triangle_i\alpha$ is read "agent i explicitly believes that α". Then, the language $\mathcal{L}$ extends the language $\mathcal{L}_0$ by a modal operator of implicit belief and a dynamic operator for belief expansion for the machine. It is defined by the following grammar:

$$\varphi \quad ::= \quad \alpha \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \Box_{\mathfrak{m}}\alpha \mid [+_{\mathfrak{m}}\alpha]\varphi,$$

where α ranges over $\mathcal{L}_0$. The formula $\Box_{\mathfrak{m}}\alpha$ is read "agent $\mathfrak{m}$ implicitly believes that α". The formula $[+_{\mathfrak{m}}\alpha]\varphi$ is read "φ holds after agent $\mathfrak{m}$ has privately expanded its belief base with α". The other Boolean constructions $\top$, $\bot$, $\rightarrow$, $\vee$ and $\leftrightarrow$ are defined in the standard way.

In $\mathcal{L}$, both agent $\mathfrak{h}$ and agent $\mathfrak{m}$ have explicit beliefs but only agent $\mathfrak{m}$ has implicit beliefs, and moreover the latter are restricted to $\mathcal{L}_0$ formulas of type α. So there are no nested implicit beliefs for agent $\mathfrak{m}$. Agent $\mathfrak{m}$ is assumed to be the unique artificial agent in the system which is endowed with unbounded reasoning and planning capabilities. The cognitive planning problem will be modeled from agent $\mathfrak{m}$'s perspective.

The interpretation of language $\mathcal{L}$ exploits the notion of belief base. While the notions of possible state (or world) and epistemic alternative are primitive in the standard semantics for epistemic logic (Fagin et al., 1995), they are defined from the primitive concept of belief base in our semantics. In particular, a state is a composite object including a description of both the agents' belief bases and the environment.[1]

Definition 1 (State). *A state is a tuple* $B = (B_{\mathfrak{h}}, B_{\mathfrak{m}}, V)$ *where: for every* $i \in Agt$, $B_i \subseteq \mathcal{L}_0$ *is agent* i*'s belief base;* $V \subseteq Atm$ *is the actual environment. The set of all states is noted* **S**.

Note that an agent's belief base B_i can be infinite. The sublanguage $\mathcal{L}_0(Atm, Agt)$ is interpreted w.r.t. states, as follows:

Definition 2 (Satisfaction). *Let* $B = (B_{\mathfrak{h}}, B_{\mathfrak{m}}, V) \in \mathbf{S}$. *Then:*

$$\begin{aligned} B \models p &\iff p \in V, \\ B \models \neg\alpha &\iff B \not\models \alpha, \\ B \models \alpha_1 \wedge \alpha_2 &\iff B \models \alpha_1 \textit{ and } B \models \alpha_2, \\ B \models \triangle_i\alpha &\iff \alpha \in B_i. \end{aligned}$$

Observe in particular the set-theoretic interpretation of the explicit belief operator: agent i explicitly believes that α if and only if α is included in its belief base.

[1] This is similar to the way states are modeled in the interpreted system semantics for multi-agent systems (Lomuscio et al., 2017).

A model is defined to be a state supplemented with a set of states, called *context*. The latter includes all states that are compatible with the common ground (Stalnaker, 2002), *i.e.*, the body of information that the agents commonly believe to be the case.

Definition 3 (Model). *A model is a pair (B, Cxt), where $B \in \mathbf{S}$ and $Cxt \subseteq \mathbf{S}$. The class of all models is noted* $\mathbf{M}$.

Note that we do not impose that $B \in Cxt$. When $Cxt = \mathbf{S}$ then (B, Cxt) is said to be *complete*, since $\mathbf{S}$ is conceivable as the complete (or universal) context which contains all possible states. We compute agent $\mathfrak{m}$'s set of epistemic alternatives from the agent $\mathfrak{m}$'s belief base, as follows.

Definition 4 (Epistemic alternatives). *$\mathcal{R}_{\mathfrak{m}}$ is the binary relation on the set $\mathbf{S}$ such that, for all $B = (B_{\mathfrak{h}}, B_{\mathfrak{m}}, V), B' = (B'_{\mathfrak{h}}, B'_{\mathfrak{m}}, V') \in \mathbf{S}$:*

$$B\mathcal{R}_{\mathfrak{m}}B' \text{ if and only if } \forall \alpha \in B_{\mathfrak{m}} : B' \models \alpha.$$

$B\mathcal{R}_{\mathfrak{m}}B'$ means that B' is an epistemic alternative for agent $\mathfrak{m}$ at B. So $\mathfrak{m}$'s set of epistemic alternatives at B includes exactly those states that satisfy all $\mathfrak{m}$'s explicit beliefs.

Definition 5 extends Definition 2 to the full language $\mathcal{L}$. Its formulas are interpreted with respect to models as follows. (We omit Boolean cases that are defined in the usual way.)

Definition 5 (Satisfaction). *Let $B = (B_{\mathfrak{h}}, B_{\mathfrak{m}}, V) \in \mathbf{S}$ and $(B, Cxt) \in \mathbf{M}$. Then:*

$$\begin{aligned}
(B, Cxt) \models \alpha &\iff B \models \alpha,\\
(B, Cxt) \models \Box_{\mathfrak{m}}\varphi &\iff \forall B' \in Cxt, \text{ if } B\mathcal{R}_{\mathfrak{m}}B'\\
&\qquad \text{then } (B', Cxt) \models \varphi,\\
(B, Cxt) \models [+_{\mathfrak{m}}\alpha]\varphi &\iff (B^{+_{\mathfrak{m}}\alpha}, Cxt) \models \varphi,
\end{aligned}$$

with $B^{+_{\mathfrak{m}}\alpha} = (B_{\mathfrak{h}}^{+_{\mathfrak{m}}\alpha}, B_{\mathfrak{m}}^{+_{\mathfrak{m}}\alpha}, V^{+_{\mathfrak{m}}\alpha})$, $V^{+_{\mathfrak{m}}\alpha} = V$, $B_{\mathfrak{m}}^{+_{\mathfrak{m}}\alpha} = B_{\mathfrak{m}} \cup \{\alpha\}$ and $B_{\mathfrak{h}}^{+_{\mathfrak{m}}\alpha} = B_{\mathfrak{h}}$.

According to the previous definition, agent $\mathfrak{m}$ implicitly believes that φ if and only if, φ is true at all states in the context that $\mathfrak{m}$ considers possible. Moreover, the private expansion of $\mathfrak{m}$'s belief base by α simply consists in agent $\mathfrak{m}$ adding the information α to its belief base, while agent $\mathfrak{h}$ keeps her belief base unchanged.

A formula $\varphi \in \mathcal{L}$ is said to be valid in the class $\mathbf{M}$, noted $\models_{\mathbf{M}} \varphi$, if and only if $(B, Cxt) \models \varphi$ for every $(B, Cxt) \in \mathbf{M}$; it is said to be satisfiable in $\mathbf{M}$ if and only if $\neg\varphi$ is not valid in $\mathbf{M}$. Finally, given a finite $\Sigma \subset \mathcal{L}_0$, we say that φ is a logical consequence of Σ in the class $\mathbf{M}$, noted $\Sigma \models_{\mathbf{M}} \varphi$, if and only if, for every $(B, Cxt) \in \mathbf{M}$ such that $Cxt \subseteq \mathbf{S}(\Sigma)$ we have $(B, Cxt) \models \varphi$, with $\mathbf{S}(\Sigma) = \{B \in \mathbf{S} : \forall \alpha \in \Sigma, B \models \alpha\}$. In (Fernandez et al., 2021), it is proved that the satisfiability checking problem and the logical consequence problem so defined are, respectively, NP-complete and co-NP-complete.

4 Planning Problems

The cognitive planning problem is specified in the context of the language $\mathcal{L}$. It consists in finding a sequence of questions or informative actions for agent $\mathfrak{m}$ which guarantees that it believes that its goal α_G is satisfied. As we emphasized above, agent $\mathfrak{m}$ is assumed to be an artificial agent which interacts with the resource-bounded human agent $\mathfrak{h}$.

Informative actions Let $Act_{\mathfrak{m}} = \{+_{\mathfrak{m}}\alpha : \alpha \in \mathcal{L}_0\}$ be agent $\mathfrak{m}$'s set of belief expansion operations (or informative actions) and let elements of $Act_{\mathfrak{m}}$ be noted $\varepsilon, \varepsilon', \ldots$ Speech acts of type 'assertion' are formalized as follows:

$$assert(\mathfrak{m},\mathfrak{h},\alpha) \stackrel{\text{def}}{=} +_{\mathfrak{m}}\triangle_{\mathfrak{h}}\triangle_{\mathfrak{m}}\alpha.$$

The event $assert(\mathfrak{m},\mathfrak{h},\alpha)$ captures the speech act "agent $\mathfrak{m}$ asserts to agent $\mathfrak{h}$ that α". The latter is assumed to coincide with the perlocutionary effect (Searle, 1969, Sect. 6.2) of the speaker learning that the hearer has learnt that the speaker believes that α.[2] We distinguish simple assertions from convincing actions:

$$convince(\mathfrak{m},\mathfrak{h},\alpha) \stackrel{\text{def}}{=} +_{\mathfrak{m}}\triangle_{\mathfrak{h}}\alpha.$$

The event $convince(\mathfrak{m},\mathfrak{h},\alpha)$ captures the action "agent $\mathfrak{m}$ convinces agent $\mathfrak{h}$ that α". We have $assert(\mathfrak{m},\mathfrak{h},\alpha) = convince(\mathfrak{m},\mathfrak{h},\triangle_{\mathfrak{m}}\alpha)$. We assume 'to assert' and 'to convince' correspond to different utterances. While 'to assert' corresponds to the speaker's utterances of the form "I think that α is true!" and "In my opinion, α is true!", 'to convince' corresponds to the speaker's utterances of the form "α is true!" and "it is the case that α!".

The previous abbreviations and, more generally, the idea of describing speech acts of a communicative plan performed by agent $\mathfrak{m}$ with $\mathfrak{m}$'s private belief expansion operations is justified by the fact that we model cognitive planning from the perspective of the planning agent $\mathfrak{m}$. Therefore, we only need to represent the effects of actions on agent $\mathfrak{m}$'s beliefs.

[2] We implicitly assume that, by default, $\mathfrak{m}$ believes that $\mathfrak{h}$ trusts its sincerity, so that $\mathfrak{h}$ will believe that $\mathfrak{m}$ believes what it says.

Questions We consider binary questions by the machine $\mathfrak{m}$ to the human $\mathfrak{h}$ of the form $?_{\mathfrak{m},\mathfrak{h}}\alpha$.[3] The set of binary questions is noted $Que_{\mathfrak{m}}$. Intuitively, $?_{\mathfrak{m},\mathfrak{h}}\alpha$ is the utterance performed by agent $\mathfrak{m}$ to agent $\mathfrak{h}$ of the form "Do you think that α is true?". Let elements of $Que_{\mathfrak{m}}$ be noted $\lambda, \lambda', \dots$ Each question is associated with its set of possible answers. The answer function $\mathcal{A} : Que_{\mathfrak{m}} \longrightarrow 2^{Act_{\mathfrak{m}}}$ is used to map each binary question to its set of possible answers and is defined as follows:

$$\mathcal{A}(?_{\mathfrak{m},\mathfrak{h}}\alpha) = \{ +_{\mathfrak{m}} \triangle_{\mathfrak{h}}\alpha, +_{\mathfrak{m}}\neg\triangle_{\mathfrak{h}}\alpha \}.$$

Answers to binary questions are noted $\rho, \rho', \dots$ The operation $+_{\mathfrak{m}}\triangle_{\mathfrak{h}}\alpha$ captures agent $\mathfrak{h}$'s positive answer to agent $\mathfrak{m}$'s binary question $?_{\mathfrak{m},\mathfrak{h}}\alpha$ ("I think that α is true!"), while $+_{\mathfrak{m}}\neg\triangle_{\mathfrak{h}}\alpha$ captures agent $\mathfrak{h}$'s negative answer ("I don't think that α is true!"). Note that if agent $\mathfrak{h}$ answers negatively to the consecutive questions $?_{\mathfrak{m},\mathfrak{h}}\alpha$ and $?_{\mathfrak{m},\mathfrak{h}}\neg\alpha$, then she expresses her uncertainty about the truth value of α.

We assume that the positive answer is the *default* answer to a question. Indeed, when agent $\mathfrak{m}$ asks question $?_{\mathfrak{m},\mathfrak{h}}\alpha$, it wants to verify whether agent $\mathfrak{h}$ endorses the belief that α and presupposes that agent $\mathfrak{h}$ will answer positively to the question. In this perspective, the speaker expects a confirmation by the interlocutor. Thus, for notational convenience, we write $da(?_{\mathfrak{m},\mathfrak{h}}\alpha)$ to denote the default answer $+_{\mathfrak{m}}\triangle_{\mathfrak{h}}\alpha$ to the question $?_{\mathfrak{m},\mathfrak{h}}\alpha$.

The following abbreviation defines a dynamic operator capturing the necessary effects of agent $\mathfrak{m}$'s question:

$$[\lambda]\varphi \overset{\text{def}}{=} \bigwedge_{\rho\in\mathcal{A}(\lambda)} [\rho]\varphi,$$

with $\lambda \in Que_{\mathfrak{m}}$. Note that, unlike the basic belief expansion operator $[+_{\mathfrak{m}}\alpha]$, the operator $[\lambda]$ is non-deterministic, as it represents the consequences of *all possible answers* to question λ. In fact, while the formula $[+_{\mathfrak{m}}\alpha]\neg\varphi \vee [+_{\mathfrak{m}}\alpha]\varphi$ is valid in the class $\mathbf{M}$, the formula $[\lambda]\neg\varphi \vee [\lambda]\varphi$ is not.

Executability preconditions The set of events includes both informative actions and questions, and is defined as follows: $Evt_{\mathfrak{m}} = Act_{\mathfrak{m}} \cup Que_{\mathfrak{m}}$. Elements of $Evt_{\mathfrak{m}}$ are noted $\gamma, \gamma', \dots$ They have executability preconditions that are specified by the following function: $\mathcal{P} : Evt_{\mathfrak{m}} \longrightarrow \mathcal{L}$. We assume that an event γ *can* take place if its executability precondition $\mathcal{P}(\gamma)$ holds.

We use the executability precondition function $\mathcal{P}$ to define the following operator of possible occurrence of an event:

$$\langle\langle\gamma\rangle\rangle\varphi \overset{\text{def}}{=} \mathcal{P}(\gamma) \wedge [\gamma]\varphi,$$

with $\gamma \in Evt$. The abbreviation $\langle\langle\gamma\rangle\rangle\varphi$ has to be read "the event γ can take place and φ necessarily holds after its occurrence".

Informative and interrogative planning problems We conclude this section with a formal specification of two planning problems, informative planning and interrogative planning.

Definition 6 (Informative planning problem)**.** *An informative planning problem is a tuple* $\langle\Sigma, Op_{\mathsf{inf}}, \alpha_G\rangle$ *where:*

- $\Sigma \subset \mathcal{L}_0$ *is a finite set of agent* $\mathfrak{m}$*'s available information,*
- $Op_{\mathsf{inf}} \subset Act_{\mathfrak{m}}$ *is a finite set of agent* $\mathfrak{m}$*'s informative actions,*
- $\alpha_G \in \mathcal{L}_0$ *is agent* $\mathfrak{m}$*'s goal.*

Informally speaking, an informative planning problem is the problem of finding an executable sequence of informative actions which guarantees that, at the end of the sequence, the planning agent $\mathfrak{m}$ believes that its goal α_G is achieved. Typically, α_G is a persuading or influencing goal, i.e., the goal of affecting agent's $\mathfrak{h}$ cognitive state (including her beliefs and intentions) in a certain way. A solution plan to an informative planning problem $\langle\Sigma, Op_{\mathsf{inf}}, \alpha_G\rangle$ is a sequence of informative actions $\varepsilon_1, \dots, \varepsilon_k$ from Op_{inf} for some k such that $\Sigma \models_{\mathbf{M}} \langle\langle\varepsilon_1\rangle\rangle \dots \langle\langle\varepsilon_k\rangle\rangle\Box_{\mathfrak{m}}\alpha_G$.

In an interrogative planning problem, the machine can perform both informative actions and questions. This problem is specified in the following definition.

Definition 7 (Interrogative planning problem)**.** *An interrogative planning problem is a tuple* $\langle\Sigma, Op_{\mathsf{inf}}, Op_{\mathsf{quest}}, \alpha_G\rangle$ *where:*

- $\Sigma \subset \mathcal{L}_0$ *is a finite set of agent* $\mathfrak{m}$*'s available information,*
- $Op_{\mathsf{inf}} \subset Act_{\mathfrak{m}}$ *is a finite set of agent* $\mathfrak{m}$*'s informative actions,*
- $Op_{\mathsf{quest}} \subset Que_{\mathfrak{m}}$ *is a finite set of agent* $\mathfrak{m}$*'s questions,*
- $\alpha_G \in \mathcal{L}_0$ *is agent* $\mathfrak{m}$*'s goal.*

Intuitively, an interrogative planning problem is the problem of finding a sequence of questions as a means of understanding the interlocutor's cognitive state and, consequently, of being able to identify the inconsistencies that she must be made aware of, via a sequence of informative actions. In other words, the sequence of questions serves the purpose of "exploring" the interlocutor's cognitive state and of building

[3] In speech act theory, binary (yes-no) questions are usually distinguished from open questions.

a representation of it in order to being able to find a plan to reach the motivational interviewing (MI) goal.

A strong solution plan to an interrogative planning problem $\langle \Sigma, Op_{\mathsf{inf}}, Op_{\mathsf{quest}}, \alpha_G \rangle$ is a sequence of questions $\lambda_1, \dots, \lambda_m$ from Op_{quest} such that

$$\Sigma \models_{\mathbf{M}} \langle\langle \lambda_1 \rangle\rangle \dots \langle\langle \lambda_m \rangle\rangle \top,$$

and $\forall \rho_1 \in \mathcal{A}(\lambda_1), \dots, \forall \rho_m \in \mathcal{A}(\lambda_m)$, $\exists \tau_1, \dots, \tau_k \in Op_{\mathsf{inf}}$ such that

$$\Sigma \models_{\mathbf{M}} [\rho_1] \dots [\rho_m] \langle\langle \tau_1 \rangle\rangle \dots \langle\langle \tau_k \rangle\rangle \Box_{\mathfrak{m}} \alpha_G.$$

A weak solution plan to an interrogative planning problem $\langle \Sigma, Op_{\mathsf{inf}}, Op_{\mathsf{quest}}, \alpha_G \rangle$ is a sequence of questions $\lambda_1, \dots, \lambda_m$ from Op_{quest} such that

$$\Sigma \models_{\mathbf{M}} \langle\langle \lambda_1 \rangle\rangle \dots \langle\langle \lambda_m \rangle\rangle \top,$$

and $\exists \tau_1, \dots, \tau_k \in Op_{\mathsf{inf}}$ such that

$$\Sigma \models_{\mathbf{M}} [da(\lambda_1)] \dots [da(\lambda_m)] \langle\langle \tau_1 \rangle\rangle \dots \langle\langle \tau_k \rangle\rangle \Box_{\mathfrak{m}} \alpha_G.$$

It is easy to verify that checking existence of a weak solution for an interrogative planning problem (EWS-INT problem) is reducible to checking existence of a solution for an informative planning problem (ES-INF problem). In (Fernandez et al., 2021) it was proved that the ES-INF problem is in $\mathsf{NP}^{\mathsf{NP}} = \Sigma_2^{\mathsf{P}}$. Thus, we get the following complexity upper bound for the EWS-INT problem.

Theorem 1. *The EWS-INT problem is in* $\mathsf{NP}^{\mathsf{NP}} = \Sigma_2^{\mathsf{P}}$.

Checking existence of a strong solution for an interrogative planning problem (ESS-INT problem) is not comparable to the ES-INF problem or the EWS-INT problem. Indeed, it requires to take all possible answers to the questions and their possible ramifications into account. The EWS-INT problem considers a single sequence of answers (the sequence of default answers) instead.

5 Belief Revision Module

In this section, we describe the belief revision module of the architecture we sketched in Section 2. As we emphasized above, such a module is necessary for updating the machine's belief base after the human has replied to its questions.

Let $\mathcal{L}_{\mathsf{PROP}}$ be the propositional language built from the following set of atomic formulas:

$$Atm^+ = Atm \cup \{p_{\triangle_i \alpha} : \triangle_i \alpha \in \mathcal{L}_0\}.$$

Moreover, let tr_{PROP} be the following translation from the language $\mathcal{L}_0$ defined in Section 3 to $\mathcal{L}_{\mathsf{PROP}}$:

$$\begin{aligned} tr_{\mathsf{PROP}}(p) &= p, \\ tr_{\mathsf{PROP}}(\neg \alpha) &= \neg tr_{\mathsf{PROP}}(\alpha), \\ tr_{\mathsf{PROP}}(\alpha_1 \wedge \alpha_2) &= tr_{\mathsf{PROP}}(\alpha_1) \wedge tr_{\mathsf{PROP}}(\alpha_2), \\ tr_{\mathsf{PROP}}(\triangle_i \alpha) &= p_{\triangle_i \alpha}. \end{aligned}$$

For each finite $X \subseteq \mathcal{L}_0$, we define $tr_{\mathsf{PROP}}(X) = \{tr_{\mathsf{PROP}}(\alpha) : \alpha \in X\}$. Moreover, we say that X is propositionally consistent if and only if $\perp \not\in Cn\big(tr_{\mathsf{PROP}}(X)\big)$, where Cn is the classical deductive closure operator over the propositional language $\mathcal{L}_{\mathsf{PROP}}$. Clearly, the latter is equivalent to saying that $\bigwedge_{\alpha \in X} tr_{\mathsf{PROP}}(\alpha)$ is satisfiable in propositional logic.

Let $\Sigma_{core}, \Sigma_{mut} \subseteq \mathcal{L}_0$ denote, respectively, the core (or, immutable) information in agent $\mathfrak{m}$'s belief base and the volatile (or, mutable) information in agent $\mathfrak{m}$'s belief base. Agent $\mathfrak{m}$'s core beliefs are stable and do not change under belief revision. On the contrary, volatile beliefs can change due to a belief revision operation . Moreover, let $\Sigma_{input} \subseteq \mathcal{L}_0$ be agent $\mathfrak{m}$'s input information set. We define $\Sigma_{base} = \Sigma_{core} \cup \Sigma_{mut}$. The revision of $(\Sigma_{core}, \Sigma_{mut})$ by input Σ_{input}, noted $Rev(\Sigma_{core}, \Sigma_{mut}, \Sigma_{input})$, is formally defined as follows:

1. if $\Sigma_{core} \cup \Sigma_{input}$ is not propositionally consistent then $Rev(\Sigma_{core}, \Sigma_{mut}, \Sigma_{input}) = (\Sigma_{core}, \Sigma_{mut})$,
2. otherwise, $Rev(\Sigma_{core}, \Sigma_{mut}, \Sigma_{input}) = (\Sigma'_{core}, \Sigma'_{mut})$, with $\Sigma'_{core} = \Sigma_{core}$ and
$$\Sigma'_{mut} = \bigcap_{X \in MCS(\Sigma_{core}, \Sigma_{mut}, \Sigma_{input})} X,$$

where $X \in MCS(\Sigma_{core}, \Sigma_{mut}, \Sigma_{input})$ if and only if:

- $X \subseteq \Sigma_{mut} \cup \Sigma_{input}$,
- $\Sigma_{input} \subseteq X$,
- $X \cup \Sigma_{core}$ is propositionally consistent, and
- there is no $X' \subseteq \Sigma_{mut} \cup \Sigma_{input}$ such that $X \subset X'$ and $X' \cup \Sigma_{core}$ is propositionally consistent.

The revision function Rev has the following effects on agent $\mathfrak{m}$'s beliefs: (i) the core belief base is not modified, while (ii) the input Σ_{input} is added to the mutable belief base only if it is consistent with the core beliefs. If the latter is the case, then the updated mutable belief base is equal to the intersection of the subsets of the mutable belief base which are maximally consistent with respect to the core belief base and which include the input Σ_{input}.[4] This guarantees that belief

[4] Note that the revision function Rev does not expand agent $\mathfrak{m}$'s core belief set Σ_{core} with the input information set Σ_{input}. It would be interesting to introduce a function $f_{\mathsf{appr}} : \mathcal{L}_0 \longrightarrow \{0, 1\}$ which specifies for every formula α in $\mathcal{L}_0$ whether the information α is completely apprehensible by agent $\mathfrak{m}$ (*i.e.*, $f_{\mathsf{appr}}(\alpha) = 1$) or not (*i.e.*, $f_{\mathsf{appr}}(\alpha) = 0$). Specifically, $f_{\mathsf{appr}}(\alpha) = 1$ means that if agent $\mathfrak{m}$ learns that α is true then, as a consequence, it will firmly believe that α is true thereby adding α not only to its set of mutable beliefs but also to its set of core beliefs. The function f_{appr} would allow us to define a variant of belief revision according to which if $\Sigma_{core} \cup \Sigma_{input}$ is propositionally consistent, then the core belief set Σ_{core} is expanded by all formulas α in Σ_{input} such that $f_{\mathsf{appr}}(\alpha) = 1$, that is, $\Sigma'_{core} = \Sigma_{core} \cup \{\alpha \in \Sigma_{input} : f_{\mathsf{appr}}(\alpha) = 1\}$.

revision satisfies minimal change. The function *Rev* is a screened revision operator as defined in (Makinson, 1997). The latter was recently generalized to the multi-agent case (Lorini and Schwarzentruber, 2021). Let $Rev(\Sigma_{core},\Sigma_{mut},\Sigma_{input}) = (\Sigma'_{core},\Sigma'_{mut})$.

For notational convenience, we write $Rev^{core}(\Sigma_{core},\Sigma_{mut},\Sigma_{input})$ to denote Σ'_{core} and $Rev^{mut}(\Sigma_{core},\Sigma_{mut},\Sigma_{input})$ to denote Σ'_{mut}. Note that, if Σ_{base} is propositionally consistent, then $Rev^{core}(\Sigma_{core},\Sigma_{mut},\Sigma_{input}) \cup Rev^{mut}(\Sigma_{core},\Sigma_{mut},\Sigma_{input})$ is propositionally consistent too.

6 Example

In this section, we illustrate the use of the cognitive planning and belief revision module of the architecture with the aid of a human-machine interaction (HMI) scenario. We assume $\mathfrak{m}$ is a virtual coaching agent which has to motivate the human agent $\mathfrak{h}$ to practice a physical activity. We suppose agent $\mathfrak{m}$ complies with the general principles of the theory of motivational interviewing (MI) to find a persuasive strategy aimed at changing the human's attitude.

One of the central cornerstones of MI is the postulate that for eliciting behavior change in a person, she has to become aware of the inconsistency between her current behavior and her desires. In other words, she has to recognize the fact that her current behavior will prevent her from obtaining what she likes.

Let us assume the disjoint sets *CondAtm*, *DesAtm* and *ActAtm* are subsets of the set of atomic propositions *Atm*. Elements of *CondAtm* are atoms specifying conditions, while elements of *DesAtm* are atoms specifying desirable properties, that is, properties that agent $\mathfrak{h}$ may wish to achieve (i.e., agent $\mathfrak{h}$'s possible desiderata). Finally, atoms in *ActAtm* are used to describe agent $\mathfrak{h}$'s behavior. Specifically, we define $ActAtm = \{\mathsf{does}(\mathfrak{h},a) : a \in Act\}$, where *Act* is a finite a set of action names. The atom $\mathsf{does}(\mathfrak{h},a)$ has to be read "agent $\mathfrak{h}$ behaves in conformity with the requirement *a*" or, simply, "agent $\mathfrak{h}$ does action *a*".

The sets of literals from *CondAtm*, *DesAtm* and *ActAtm* are defined in the usual way as follows:

$$DesLit = DesAtm \cup \{\neg p : p \in DesAtm\},$$

$$CondLit = CondAtm \cup \{\neg p : p \in CondAtm\},$$

$$ActLit = ActAtm \cup \{\neg p : p \in ActAtm\},$$

$$Lit = DesLit \cup CondLit \cup ActLit.$$

We define $LitSet = 2^{Lit}$ and $LitSet_0 = LitSet \setminus \{\emptyset\}$.

We moreover assume that the set of atomic propositions *Atm* includes one atom $\mathsf{des}(\mathfrak{h},l)$ for each $l \in DesLit$ standing for "agent $\mathfrak{h}$ desires l to be true".

For the sake of illustration, we suppose that $Act = \{ps\}$ where *ps* is the action (or requirement) "to practice regularly a sport or physical activity". Therefore, $ActAtm = \{\mathsf{does}(\mathfrak{h},ps)\}$. Moreover, $DesAtm = \{dr,pw,lw,at,gh,st\}$ and $CondAtm = \{ow,sl,co\}$, with the atoms having the following intuitive meaning: *dr*: "agent $\mathfrak{h}$ has dietary restrictions"; *pw*: "agent $\mathfrak{h}$ puts on weight"; *lw*: "agent $\mathfrak{h}$ loses weight"; *at*: "agent $\mathfrak{h}$ is attractive"; *gh*: "agent $\mathfrak{h}$ is in good health"; *st*: "agent $\mathfrak{h}$ is stressed"; *ow*: "agent $\mathfrak{h}$ has an office work"; *sl*: "agent $\mathfrak{h}$ has a sedentary life style"; *co*: "agent $\mathfrak{h}$ is a commuter and spends quite some time in the traffic everyday".

The following abbreviation captures a simple notion of necessity for $X \in LitSet$ and $l \in Lit$:

$$\mathsf{nec}(X,l) \overset{\text{def}}{=} \bigwedge_{l' \in X} l' \to l.$$

$\mathsf{nec}(X,l)$ has to be read "the facts in X will not be true unless l is true" or more shortly "l is necessary for X".

Agent $\mathfrak{m}$'s initial knowledge about agent $\mathfrak{h}$'s cognitive state is specified by the following six abbreviations:

$$\alpha_1 \overset{\text{def}}{=} \bigwedge_{l \in Lit} \triangle_{\mathfrak{h}} \mathsf{nec}(\{l\},l),$$

$$\alpha_2 \overset{\text{def}}{=} \bigwedge_{l \in Lit} \big(\triangle_{\mathfrak{h}} \mathsf{nec}(\emptyset,l) \leftrightarrow \triangle_{\mathfrak{h}} l\big),$$

$$\alpha_3 \overset{\text{def}}{=} \bigwedge_{l \in Lit, X, X' \in LitSet : X' \subseteq X} \Big(\triangle_{\mathfrak{h}} \mathsf{nec}(X,l) \to \big(\bigwedge_{l' \in X'} \triangle_{\mathfrak{h}} l' \to \triangle_{\mathfrak{h}} \mathsf{nec}(X \setminus X', l)\big)\Big),$$

$$\alpha_4 \overset{\text{def}}{=} \bigwedge_{l \in Lit, X, X' \in LitSet : X \subseteq X'} \big(\triangle_{\mathfrak{h}} \mathsf{nec}(X,l) \to \triangle_{\mathfrak{h}} \mathsf{nec}(X',l)\big),$$

$$\alpha_5 \overset{\text{def}}{=} \bigwedge_{l \in Lit} \Big(\big(\mathsf{des}(\mathfrak{h},l) \leftrightarrow \triangle_{\mathfrak{h}} \mathsf{des}(\mathfrak{h},l)\big) \wedge \big(\neg\mathsf{des}(\mathfrak{h},l) \leftrightarrow \triangle_{\mathfrak{h}} \neg\mathsf{des}(\mathfrak{h},l)\big)\Big),$$

$$\alpha_6 \overset{\text{def}}{=} \bigwedge_{a \in Act} \Big(\big(\mathsf{does}(\mathfrak{h},a) \leftrightarrow \triangle_{\mathfrak{h}} \mathsf{does}(\mathfrak{h},a)\big) \wedge \big(\neg\mathsf{does}(\mathfrak{h},a) \leftrightarrow \triangle_{\mathfrak{h}} \neg\mathsf{does}(\mathfrak{h},a)\big)\Big),$$

Hypotheses α_1-α_4 are general properties about agent $\mathfrak{h}$'s conception of necessity. According to α_1, agent $\mathfrak{h}$ believes that every fact is necessary for itself while, according to α_2, agent $\mathfrak{h}$ believes a fact is true regardless of the circumstances if and only if she believes that it is true. According to α_3, if agent $\mathfrak{h}$ believes that l is necessary for the facts in X being true and believes every fact in $X' \subseteq X$, then she believes that l

is necessary for the facts in the remaining set $X \setminus X'$ being true. According to α_4, if $X \subseteq X'$ and agent $\mathfrak{h}$ believes that l is necessary for X then she believes that l is necessary for X' as well. Hypotheses α_5 and α_6 capture agent $\mathfrak{h}$'s introspection over her desires (hyphothesis α_5) and agent $\mathfrak{h}$'s perfect knowledge about her actions and inactions (hyphothesis α_6).

We moreover suppose that agent $\mathfrak{m}$ has the following information in its belief base capturing the necessity relations between conditions, desirable properties and actions:

$$\begin{aligned}\alpha_7 \overset{\text{def}}{=} &\, \mathsf{nec}\big(\{\neg dr, \neg pw, sl\}, \mathsf{does}(\mathfrak{h}, ps)\big) \wedge \\ &\, \mathsf{nec}\big(\{at, \neg dr\}, \mathsf{does}(\mathfrak{h}, ps)\big) \wedge \\ &\, \mathsf{nec}\big(\{sl, gh\}, \mathsf{does}(\mathfrak{h}, ps)\big) \wedge \\ &\, \mathsf{nec}\big(\{gh\}, \neg st\big) \wedge \\ &\, \mathsf{nec}\big(\{co, ow\}, sl\big).\end{aligned}$$

For example, $\mathsf{nec}\big(\{\neg dr, \neg pw, sl\}, \mathsf{does}(\mathfrak{h}, ps)\big)$ means that practicing regularly a sport is necessary for not having dietary restrictions and not putting weight, while having a sedentary work style (i.e., a person cannot pretend to not put weight and not have dietary restrictions without practicing a sport, if she has a sedentary work style).

The following abbreviation defines the concept of agent $\mathfrak{h}$'s awareness of the inconsistency between the actual state of affairs α and her desires:

$$\begin{aligned}\mathsf{AwareIncon}(\mathfrak{h}, \alpha) \overset{\text{def}}{=} \bigvee_{X \in LitSet} \Big(& \bigwedge_{l' \in X} \mathsf{des}(\mathfrak{h}, l') \wedge \\ & \triangle_{\mathfrak{h}} \mathsf{nec}(X, \neg\alpha) \wedge \triangle_{\mathfrak{h}} \alpha \Big).\end{aligned}$$

According to the previous definition, agent $\mathfrak{h}$ is aware of the inconsistency between the actual state of affairs α and her desires, noted $\mathsf{AwareIncon}(\mathfrak{h}, \alpha)$, if she believes that the satisfaction of her desires is jeopardized by the fact that α is true. More precisely, (i) agent $\mathfrak{h}$ believes that she will not achieve her desires unless α is false and (ii) she believes that α is actually true.

We suppose that the pieces of information $\alpha_1, \ldots, \alpha_7$ constitute agent $\mathfrak{m}$'s initial core belief base, that is, $\Sigma_{core} = \{\alpha_1, \ldots, \alpha_7\}$. Moreover, we suppose that agent $\mathfrak{m}$'s initial mutable belief base is empty, that is, $\Sigma_{mut} = \emptyset$. We consider the planning problem in which agent $\mathfrak{m}$ tries to motivate agent $\mathfrak{h}$ to practice regularly a sport. To this aim, agent $\mathfrak{m}$ tries to achieve the following goal:

$$\alpha_G \overset{\text{def}}{=} \neg\mathsf{does}(\mathfrak{h}, ps) \rightarrow \mathsf{AwareIncon}\big(\mathfrak{h}, \neg\mathsf{does}(\mathfrak{h}, ps)\big).$$

In other words, agent $\mathfrak{m}$ tries to make it the case that if agent $\mathfrak{h}$ does not practice a sport, then she becomes aware of the inconsistency between her actual desires and the fact that she does not practice a sport.

Let $X \subseteq DesLit$, $X' \subseteq CondLit$ and $l \in ActLit \cup CondLit$. We assume agent $\mathfrak{m}$'s action $convince\big(\mathfrak{m}, \mathfrak{h}, \mathsf{nec}(X \cup X', l)\big)$ to be concretely realized through the utterance "since condition X' holds, you will not satisfy your desires X unless l is true!". For example, $convince\big(\mathfrak{m}, \mathfrak{h}, \mathsf{nec}(X \cup X', \mathsf{does}(\mathfrak{h}, ps))\big)$ corresponds to the utterance "since condition X' holds, you will not satisfy your desires X unless you do action ps!", while $convince\big(\mathfrak{m}, \mathfrak{h}, \mathsf{nec}(X \cup X', \neg\mathsf{does}(\mathfrak{h}, ps))\big)$ corresponds to the utterance "since condition X' holds, you will not satisfy your desires X unless you refrain from doing action ps!". For notational convenience, we abbreviate $convince\big(\mathfrak{m}, \mathfrak{h}, \mathsf{nec}(X \cup X', l)\big)$ by $!_{\mathfrak{m},\mathfrak{h}}(X, X', l)$. We assume the following repertoires of informative and interrogative actions for agent $\mathfrak{m}$:

$$\begin{aligned}Op_{\mathsf{inf}} =& \big\{ !_{\mathfrak{m},\mathfrak{h}}(X, X', l) : X \subseteq DesLit, X' \subseteq CondLit \\ & \text{and } l \in ActLit \cup CondLit \big\}, \\ Op_{\mathsf{quest}} =& \big\{ ?_{\mathfrak{m},\mathfrak{h}} \mathsf{des}(\mathfrak{h}, l) : l \in DesLit \big\} \cup \\ & \big\{ ?_{\mathfrak{m},\mathfrak{h}} l : l \in ActLit \cup CondLit \big\},\end{aligned}$$

with the following executability preconditions for their elements:

$$\begin{aligned}\mathcal{P}\big(!_{\mathfrak{m},\mathfrak{h}}(X, X', l)\big) =& \Box_{\mathfrak{m}} \Big(\mathsf{nec}(X \cup X', l) \wedge \\ & \bigwedge_{l' \in X} \mathsf{des}(\mathfrak{h}, l') \wedge \bigwedge_{l'' \in X'} \triangle_{\mathfrak{h}} l'' \Big), \\ \mathcal{P}\big(?_{\mathfrak{m},\mathfrak{h}} \mathsf{des}(\mathfrak{h}, l)\big) =& \mathcal{P}\big(?_{\mathfrak{m},\mathfrak{h}} l\big) = \top.\end{aligned}$$

In other words, a question is always executable. Moreover, agent $\mathfrak{m}$ can perform the action $!_{\mathfrak{m},\mathfrak{h}}(X, X', l)$ — i.e., "since condition X' holds, you will not satisfy your desires X unless l is true!" — only if (i) it believes that agent $\mathfrak{h}$ desires every fact in X to be true, (ii) it believes that agent $\mathfrak{h}$ believes every fact in X', and (iii) it believes that l is necessary for X when X' holds. Thus, by performing the speech act $!_{\mathfrak{m},\mathfrak{h}}(X, X', l)$, agent $\mathfrak{m}$ informs agent $\mathfrak{h}$ that, in view of the fact that condition X' holds, l is necessary for the satisfaction of her desires X, since it presupposes that agent $\mathfrak{h}$ has indeed such desires and believes that the condition holds.

We suppose that at every step k of the interaction with agent $\mathfrak{h}$, agent $\mathfrak{m}$ tries to find a solution for the informative planning problem $\langle \Sigma_{base}^{k}, Op_{\mathsf{inf}}^{k}, \alpha_G \rangle$. If it can find it, it proceeds with its execution and then interaction stops. Otherwise, it tries to find a weak solution for the interrogative planning problem $\langle \Sigma_{base}^{k}, Op_{\mathsf{inf}}^{k}, Op_{\mathsf{quest}}^{k}, \alpha_G \rangle$. If it cannot find it, the interaction stops. Otherwise, it executes the corresponding sequence of questions and revises its belief base according to agent $\mathfrak{h}$'s set of responses $Resp_{\mathfrak{h}}^{k}$. Then, it moves to step $k+1$. We suppose that $\Sigma^0_{core} = \Sigma_{core}$, $\Sigma^0_{mut} = \Sigma_{mut}$, $Op^0_{\mathsf{inf}} = Op_{\mathsf{inf}}$ and $Op^0_{\mathsf{quest}} = Op_{\mathsf{quest}}$. Moreover,

$$\begin{aligned}
\Sigma^{k+1}_{core} &= Rev^{core}(\Sigma^k_{core}, \Sigma^k_{mut}, Resp^k_{\mathfrak{h}}),\\
\Sigma^{k+1}_{mut} &= Rev^{mut}(\Sigma^k_{core}, \Sigma^k_{mut}, Resp^k_{\mathfrak{h}}),\\
Op^{k+1}_{\mathsf{inf}} &= Op^k_{\mathsf{inf}},\\
Op^{k+1}_{\mathsf{quest}} &= Op^k_{\mathsf{quest}} \setminus Selected(Op^k_{\mathsf{quest}}),
\end{aligned}$$

where $Selected(Op^k_{\mathsf{quest}})$ is the set of questions included in the interrogative plan selected at step k. We remove them because we want to avoid that agent $\mathfrak{m}$ keeps asking the same question indefinitely.

Let us illustrate an example of interaction. At step 0, agent $\mathfrak{m}$ cannot find a solution for the informative planning problem. Thus, it decides to go with questions. It finds $?_{\mathfrak{m},\mathfrak{h}}\mathsf{does}(\mathfrak{h},ps)$ as solution for the interrogative planning problem. We suppose agent $\mathfrak{h}$'s response to agent $\mathfrak{m}$'s question is $+_{\mathfrak{m}}\neg\triangle_{\mathfrak{h}}\mathsf{does}(\mathfrak{h},ps)$. At step 1, again agent $\mathfrak{m}$ cannot find a solution for the informative planning problem. Thus, it moves to the interrogative planning problem and finds the following sequence of questions as a weak solution:

$$?_{\mathfrak{m},\mathfrak{h}}\mathsf{des}(\mathfrak{h},gh), ?_{\mathfrak{m},\mathfrak{h}}co, ?_{\mathfrak{m},\mathfrak{h}}ow.$$

Agent $\mathfrak{m}$ executes the interrogative plan. We suppose agent $\mathfrak{h}$'s set of responses to agent $\mathfrak{m}$'s questions at step 1 is $\big\{+_{\mathfrak{m}}\triangle_{\mathfrak{h}}\mathsf{des}(\mathfrak{h},gh), +_{\mathfrak{m}}\triangle_{\mathfrak{h}}co, +_{\mathfrak{m}}\triangle_{\mathfrak{h}}ow\big\}$.

Thus, at step 2, agent $\mathfrak{m}$ can find a solution for the informative planning problem. The solution is the following sequence of assertive speech acts of length 2:

$$!_{\mathfrak{m},\mathfrak{h}}(\emptyset,\{co,ow\},sl), !_{\mathfrak{m},\mathfrak{h}}\big(\{gh\},\{sl\},\mathsf{does}(\mathfrak{h},ps)\big).$$

Agent $\mathfrak{m}$ executes the informative plan. The previous interaction between agent $\mathfrak{m}$ and agent $\mathfrak{h}$ is illustrated in Figure 2 in which every speech act is associated with its corresponding utterance.

Speaker	Utterance	Speech act
$\mathfrak{m}$	Do you practice a sport regularly?	$?_{\mathfrak{m},\mathfrak{h}}\mathsf{does}(\mathfrak{h},ps)$
$\mathfrak{h}$	I don't	$+_{\mathfrak{m}}\neg\triangle_{\mathfrak{h}}\mathsf{does}(\mathfrak{h},ps)$
$\mathfrak{m}$	Do you wish to be in good health?	$?_{\mathfrak{m},\mathfrak{h}}\mathsf{des}(\mathfrak{h},gh)$
$\mathfrak{h}$	Yes	$+_{\mathfrak{m}}\triangle_{\mathfrak{h}}\mathsf{des}(\mathfrak{h},gh)$
$\mathfrak{m}$	Do you spend quite some time in the traffic everyday as a commuter?	$?_{\mathfrak{m},\mathfrak{h}}co$
$\mathfrak{h}$	Yes	$+_{\mathfrak{m}}\triangle_{\mathfrak{h}}co$
$\mathfrak{m}$	Do you have an office work?	$?_{\mathfrak{m},\mathfrak{h}}ow$
$\mathfrak{h}$	Yes	$+_{\mathfrak{m}}\triangle_{\mathfrak{h}}ow$
$\mathfrak{m}$	You spend quite some time in the traffic everyday as a commuter and you have an office work. Therefore, your life style is sedentary!	$!_{\mathfrak{m},\mathfrak{h}}(\emptyset,\{co,ow\},sl)$
$\mathfrak{m}$	Your life style is sedentary. Therefore, you will not satisfy your desire to be in good health unless you practice a sport regularly!	$!_{\mathfrak{m},\mathfrak{h}}\big(\{gh\},\{sl\},$ $\mathsf{does}(\mathfrak{h},ps)\big)$

Figure 2: Human-machine dialogue

7 Conclusion

Let's take stock. We have presented a model of cognitive planning and shown that it can elegantly formalize some principles of the motivational interviewing (MI) methodology, a counseling method used in clinical psychology for eliciting attitude and behavior change in humans.

Directions of future work are manifold. An important strategy of MI consists in helping the participant to overcome the obstacles that prevent her from converting her mere desires into intentions and then into effective behavior. Some of these obstacles are of cognitive nature. For example, the participant could hesitate whether to start to practice a sport regularly since she fears that practicing a sport increases the risk of getting injured. In this situation, the counselor can try to reassure the participant that her fear is unfounded. More generally, it can try to make the participant to revise her beliefs that a certain action has negative consequences. Another cognitive obstacle could be the participant's belief that she does not have the right capabilities and potential to change her behaviour. The counselor can again try to make the participant revise her belief by providing counter-evidence. We plan to extend our analysis to these aspects of MI that we were neglected in the paper.

In future work, we also plan to experimentally validate our approach to MI based on cognitive planning. To this aim, we plan to implement the scenario described in Section 6 and to evaluate the performance of the artificial agent in its interaction with the human.

The work presented in this paper is part of a larger project which is devoted to development of an artificial agent with persuasive capabilities which can promote positive behavior change in the human. The next step of our investigation is to endow the artificial agent with multimodal communicative capabilities which go beyond verbal behavior. As shown in (Potdevin et al., 2021), non-verbal behavior in communication including facial expressions is fundamental for increasing the machine's believability and trustworthiness thereby making the human more willing to believe what the machine says.

ACKNOWLEDGEMENTS

Support from the ANR project CoPains "Cognitive Planning in Persuasive Multimodal Communication" (grant number ANR-18-CE33-0012) is gratefully acknowledged.

REFERENCES

Bolander, T. and Andersen, M. B. (2011). Epistemic planning for single- and multi-agent systems. *Journal of Applied Non-Classical Logics*, 21(1):9–34.

Bylander, T. (1994). The computational complexity of propositional STRIPS planning. *Artificial Intelligence*, 69(1-2):165–204.

Cooper, M. C., Herzig, A., Maffre, F., Maris, F., Perrotin, E., and Régnier, P. (2021). A lightweight epistemic logic and its application to planning. *Artificial Intelligence*, 298.

Cooper, M. C., Herzig, A., Maffre, F., Maris, F., and Régnier, P. (2016). Simple epistemic planning: generalised gossiping. In *Proceedings of the 22nd European Conference on Artificial Intelligence (ECAI 2016)*, pages 1563–1564.

da Silva, J. G. G., Kavanagh, D. J., Belpaeme, T., Taylor, L., Beeson, K., Andrade, J., et al. (2018). Experiences of a motivational interview delivered by a robot: qualitative study. *Journal of medical Internet research*, 20(5):e7737.

Fagin, R., Halpern, J., Moses, Y., and Vardi, M. (1995). *Reasoning about Knowledge*. MIT Press, Cambridge.

Fernandez, J., Longin, D., Lorini, E., and Maris, F. (2021). A simple framework for cognitive planning. In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-2021)*, pages 6331–6339. AAAI Press.

Ghallab, M., Nau, D., and Traverso, P. (2004). *Automated planning: theory and practice*. Morgan Kaufmann.

Halpern, J. Y. and Moses, Y. (1992). A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54(3):319–379.

Kanaoka, T. and Mutlu, B. (2015). Designing a motivational agent for behavior change in physical activity. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1445–1450.

Kominis, F. and Geffner, H. (2015). Beliefs in multiagent planning: from one agent to many. In *ICAPS 2015*, pages 147–155. AAAI Press.

Lisetti, C., Amini, R., Yasavur, U., and Rishe, N. (2013). I can help you change! an empathic virtual agent delivers behavior change health interventions. *ACM Transactions on Management Information Systems (TMIS)*, 4(4):1–28.

Lomuscio, A., Qu, H., and Raimondi, F. (2017). MCMAS: an open-source model checker for the verification of multi-agent systems. *International Journal on Software Tools for Technology Transfer*, 19:9–30.

Lorini, E. (2020). Rethinking epistemic logic with belief bases. *Artificial Intelligence*, 282.

Lorini, E. and Schwarzentruber, F. (2021). Multi-agent belief base revision. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI 2021)*. ijcai.org.

Löwe, B., Pacuit, E., and Witzel, A. (2011). DEL planning and some tractable cases. In *Proceedings of the 3rd International International Workshop on Logic, Rationality and Interaction (LORI 2011)*, pages 179–192. Springer Berlin Heidelberg.

Lundahl, B. and Burke, B. L. (2009). The effectiveness and applicability of motivational interviewing: A practice-friendly review of four meta-analyses. *Journal of clinical psychology*, 65(11):1232–1245.

Makinson, D. (1997). Screened revision. *Theoria*, 63:14–23.

Miller, W. R. and Rollnick, S. (2012). *Motivational interviewing: Helping people change*. Guilford press.

Muise, C., Belle, V., Felli, P., McIlraith, S. A., Miller, T., Pearce, A. R., , and Sonenberg, L. (2021). Efficient multi-agent epistemic planning: Teaching planners about nested belief. *Artificial Intelligence*, 302.

Muise, C., Belle, V., Felli, P., McIlraith, S. A., Miller, T., Pearce, A. R., and Sonenberg, L. (2015). Planning over multi-agent epistemic states: A classical planning approach. In *AAAI 2015*, pages 3327–3334. AAAI Press.

Olafsson, S., O'Leary, T., and Bickmore, T. (2019). Coerced change-talk with conversational agents promotes confidence in behavior change. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*, pages 31–40.

Potdevin, D., Clavel, C., and Sabouret, N. (2021). Virtual intimacy in human-embodied conversational agent interactions: the influence of multimodality on its perception. *Journal on Multimodal User Interfaces*, 15(1):25–43.

Schulman, D., Bickmore, T., and Sidner, C. (2011). An intelligent conversational agent for promoting long-term health behavior change using motivational interviewing. In *2011 AAAI Spring Symposium Series*.

Searle, J. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press, Cambridge.

Stalnaker, R. (2002). Common ground. *Linguistics and Philosophy*, 25(5-6):701–721.

van Ditmarsch, H. P., van der Hoek, W., and Kooi, B. (2007). *Dynamic Epistemic Logic*. Kluwer Academic Publishers.