

Towards Human-Aware Epistemic Planning For Promoting Behavior-Change

Andreas Brännström, Timotheus Kampik, Juan Carlos Nieves

Department of Computing Science
Umeå university
SE-901 87, Umeå, Sweden
{andreasb, tkampik, jcnieves} @cs.umu.se

Abstract

This paper introduces an approach to human-aware epistemic planning in which a rational intelligent agent plans its actions for encouraging a human to proceed in a social virtual reality (VR) environment. In order to persuade the human user to execute specific actions, the agent adapts the virtual environment by adjusting motivators in the environment. The agent's model of the human is based on the theory of planned behavior (TPB), a cognitive theory to explain and predict human behavior. The intelligent agent manipulates the environment, a process where the agent conducts epistemic actions, i.e., adapting the environment and observing human responses, in order to understand the human's behavior and encourage human actions. An action reasoning framework is introduced that defines transitions between goal-oriented human activities in the virtual scenario. The proposed human-aware planning architecture can also be applied in environments that are not virtual, by utilizing modern mobile devices which have built-in sensors that measure motion, orientation, and various environmental conditions.

Introduction

Making a lasting behavior-change is rarely a simple process. It usually involves a substantial commitment of time and effort to overcome behavioral challenges. One key stage in a behavior-change process is to take action by confronting the situations in which the unwanted behavior arises (Prochaska and Velicer 1997). This is a goal-driven activity, aimed towards a human's contemplated goals. Assuming one final goal, along the way there are a series of sub-goals that may lie in a sequence, successively pursued to finally reach the final goal. However, the sub-goals can also be mutually exclusive, more or less necessary, recurring, ongoing, temporary, and with certain probability due to the stochastic human behavior. A problem in human-aware planning is whether these dynamic sub-goals of the human can be preserved in an autonomous system's predictive model of the human (Giorgini et al. 2002). This requires a model that can identify the current situation, which sub-goals that are present in the human's plan, which sub-goal the human is currently pursuing, which goals are achieved, and which are coming next.

Epistemic planning (Bolander 2017) concerns the problem where an agent has gaps in its current state of knowledge. These gaps must be filled in order to reach a desired state of knowledge. Thus, the agent might have to acquire additional knowledge, to reason, or to draw conclusions that fill the gaps. To collect the missing knowledge, epistemic actions, e.g., sensing or world altering actions (Shvo and McIlraith 2020), can be taken towards the environment, by testing how the environment behaves in response to these actions. The agent can also involve other agents that may possess the relevant knowledge, by asking or reasoning about the other agents' beliefs. In a scenario where an agent's task is to promote human actions, the human's mental state can only be interpreted; the agent can merely attempt to encourage human actions, and observe the human's behavior in response. In such a scenario, epistemic actions can be taken to acquire a theory of mind (Frith and Frith 2005) of the human to figure out the current intentions, behavioral motivators and behavioral inhibitors of the human, to provide timely interventions. This paper approaches this problem through human-aware planning.

Human-aware planning is a way to improve the ability of autonomous systems to plan its actions in a space that is populated and affected by humans (Chakraborti 2018). In human-aware planning, the intelligent system is required to make alternative hypotheses of the humans' plans, i.e., predict the actions that the human might perform in the future (Cirillo, Karlsson, and Saffiotti 2010). In addition, the intelligent system is required to manage human goal achievement, i.e., to understand the relationship between human goals and which partial goals are required to be achieved for reaching a future goal.

This paper introduces an theoretical action reasoning framework¹ to deal with scenarios where the intelligent agent's task is to promote human behavior-change by assisting a human user in reaching their goals in a virtual reality environment. In these virtual scenarios, dynamic sub-goals of the human must be kept in the agent's planning process while pursuing the human's higher goal. The human has a final goal, but along the way there are a series of sub-goals

¹<https://github.com/Interactive-Intelligent-Systems/ha-tpb-planner>

that may lie in a sequence, successively pursued to finally reach the goal. However, the sub-goals can also be mutually exclusive, more or less necessary, recurring, ongoing, temporary, and with certain probability due to the stochastic human behavior. Given this human-aware planning problem, the following research questions arise:

- By considering an agent that has control over manipulating an environment, how can the agent reason about dynamic goals in human-aware planning?
- How can an environment be dynamically adapted based on the interactions with a human user in order to promote human behavior-change?

An action reasoning (Gelfond and Lifschitz 1998) framework is proposed, which is prototyped in the disjunctive logic planner DLV-K (Eiter et al. 2003). The architecture is modelled in accordance with the Theory of Planned Behavior (TPB) (Ajzen and others 1991), a cognitive theory to explain and predict an individual's intention to engage in a behavior at a specific time and place. The general idea is that the individual's beliefs about a behavior have causal effects on the individual's attitudes, subjective norms, and perceived behavioral control in the behavior, which in turn promotes or inhibits engagement in the behavior.

In order to formalize a computational model, Action Reasoning (Gelfond and Lifschitz 1998) is utilized. To reason about the goals of the human, the agent perceives the current state of the environment and generates an action plan based on the specified constraints; e.g., temporal and sequential constraints between activities, human preferences, subjective norms, and by perceiving and identifying the human's mental and physical state. By defining the human's physical and mental state based on variables of the environment, plans can be generated for providing goal-oriented interventions.

A final step in the proposed architecture is the conduction of *counterfactual actions*. The system starts an evaluation process in which it generates hypotheses about the human's future behavior. In this evaluation process, the alternative plans generated by the logic planner are evaluated in relation to weights tailored for the individual human, selecting a final plan for adapting the virtual scenario. The planning agent thus utilizes *counterfactual* actions to reflect upon the human's mental state, and then *epistemic* actions by which the environment is adapted. The epistemic actions are followed by an observation of the human's actual behavior, which provides input for the next counterfactual planner execution.

A general model of the proposed architecture can be applied in any human-centered scenario for promoting behavior change, given that relevant knowledge for the specific domain has been elicited and incorporated in the model. In this paper, the use-case for evaluating the architecture is a VR game, developed for children with autism for practicing social scenarios, scenarios that children with autism often find stressful or scary. The particular use-case represents a cafeteria environment (see Figure 1 and 2). For this use-case, stressors and motivators must be incorporated in the model, which the intelligent agent can utilize in its epistemic planning for promoting behavior change. The environment



Figure 1: The virtual environment represents a school cafeteria. The environment is built in the Unity 3D Game Engine.

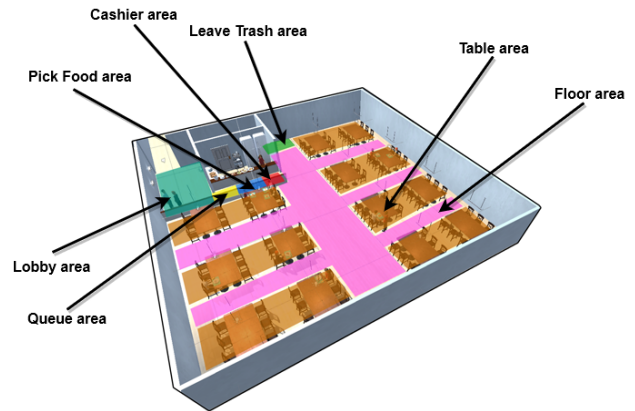


Figure 2: The virtual cafeteria is represented by a grid in which smaller areas are defined. These areas can be related to certain activities in the scenario in order to help the system in activity recognition.

model, as well as the generic user model, is based on interviews with domain experts, i.e., psychologists and special education teachers. The VR game has been evaluated in controlled user-tests to find directions for further development.

The rest of this paper is organized as follows. First, we briefly present the theoretical framework; the theory of planned behavior (TPB), and a definition of transition systems and action reasoning. The state-of-the-art in human-aware planning is then presented and how this is related to epistemic planning. Finally, a specification of the proposed intelligent system architecture is presented which introduces an action reasoning approach to human-aware planning that utilizes the theory of planned behavior. The paper is concluded by a discussion of the architecture's potential, limitations, possible use-cases, and directions for future work.

Theoretical Framework

This section explains the cognitive theory that has influenced the modeling of the proposed action reasoning framework for human-aware planning; Theory of Planned Behavior (Ajzen and others 1991). Computational paradigms

of the proposed architecture are then explained; Transition systems (Gelfond and Lifschitz 1998) and Action reasoning (Gelfond and Lifschitz 1998).

Theory of Planned Behavior

Theory of Planned Behavior (TPB) (Ajzen and others 1991), later reformed as The Reasoned Action Approach (RAA) (Fishbein and Ajzen 2011), is a cognitive theory explaining and predicting an individual's intention to engage in a behavior at a specific time and place. The general idea is that the individual's beliefs about a behavior have causal effects on the individual's attitudes, subjective norms, and perceived behavioral control in the behavior, which in turn promotes or inhibits engagement in the behavior. The key component in this model is as such behavioral intentions which are influenced by (1) the individual's attitude about a behavior, (2) the individual's subjective norms in relation to the behavior, and (3) the individual's perceived behavioral control in conducting the behavior.

Attitude (A) refers to the degree to which an individual has a positive or negative evaluation of the behavior. This entails a consideration of the outcomes of performing the behavior. The overall attitude towards the behavior is a consideration of each expected outcome b of the behavior, multiplied with the individual's valuation e of that outcome. These expectancy-value pairs in the behavior are summarized, resulting in the overall attitude towards the behavior.

Definition 1 (Attitude) Let $OUT_a = \{out_0, \dots, out_n\}$ be a set of propositional atoms that denote the expected outcomes of an activity a . Let $OUT'_a = \{(out_0, e_0, b_0), \dots, (out_n, e_n, b_n)\}$, where $\forall (out_i, b_i, e_i) \in OUT'_a, 0 \leq i \leq n, b_i \in [0, 1]$ is the quantification of the expected outcome out_i and $e_i \in [0, 1]$ is an agent's subjective quantitative evaluation of out_i . We call OUT'_a the agent's outcome assessment of a . We define the agent's attitude towards the behavior outcomes, denoted by $A(OUT'_a)$, as follows:

$$A(OUT'_a) := \sum_{(x,b,e) \in OUT'_a} b * e$$

Subjective norm (SN) refers to the belief about whether people approve or disapprove of the behavior i . The individual's beliefs about what people of importance to the person think of their, or the individual's, engagement in the behavior. This involves social norms, which are normative behavior for a group of people. The overall subjective norms towards the behavior is a consideration of each normative belief n of the behavior, multiplied with the individual's motivation m to comply with that norm. All norm-compliance pairs in the behavior are then summarized, resulting in the individual's overall evaluation of subjective norms towards the behavior.

Definition 2 (Subjective norm)

Let $INF_a = \{inf_0, \dots, inf_u\}$ be a set of propositional atoms that denote the social influences in an activity a . Let $INF'_a = \{(inf_0, m_0, n_0), \dots, (inf_u, m_u, n_u)\}$, where $\forall (n_i, m_i) \in INF'_a, 0 \leq i \leq u, n_i \in [0, 1]$ is the

quantification of the social influences inf_i and $m_i \in [0, 1]$ is an agent's subjective quantitative motivation to comply with inf_i . We call INF'_a the agent's norm assessment of a . We define the agent's subjective norm towards the behavior's social influences, denoted by $SN(INF'_a)$, as follows:

$$SN(INF'_a) := \sum_{(x,n,m) \in INF'_a} n * m$$

Perceived behavioral control (PBC) refers to an individual's perception of the ease or difficulty of performing the behavior i . Perceived behavioral control can vary for an individual depending on the activity or action that is to be performed. The overall perceived behavioral control towards the behavior is a consideration of each performance p aspect of the behavior, multiplied with the individual's perceived controllability c of that aspect. All performance-controllability pairs in the behavior are then summarized, resulting in the overall perceived behavioral control towards the behavior.

Definition 3 (Perceived behavioral control)

Let $PER_a = \{per_0, \dots, per_n\}$ be a set of propositional atoms that denote the performance aspects of an activity a . Let $PER'_a = \{(per_0, c_0, p_0), \dots, (per_n, c_n, p_n)\}$, where $\forall (p_i, c_i) \in PER'_a, 0 \leq i \leq n, p_i \in [0, 1]$ is the quantification of the performance aspects per_i and $c_i \in [0, 1]$ is an agent's subjective perceived controllability of per_i . We call PER'_a the agent's control assessment of a . We define the agent's perceived behavioral control towards the behavior's performance aspects, denoted by $PBC(PER'_a)$, as follows:

$$PBC(PER'_a) := \sum_{(x,p,c) \in PER'_a} p * c$$

According to TPB, behavioral intention (BI) is the motivational factor that influences a given behavior, the likelihood that the human will initiate in the behavior. Behavioral intention is the sum of the overall attitude, subjective norm, and perceived behavioral control. Specific activities can be more or less affected by these three predictors. Thus, an activity specific weight w is added to each predictor.

Definition 4 (Behavioral intention) Let a be an activity, let OUT'_a be an agent's outcome assessment of a , let INF'_a be the agent's norm assessment of a , and let PER'_a be the agent's control assessment of a . We define the agent's behavioral intention towards an activity a , denoted by $BI(a)$, as follows:

$$BI(a) := w_1 A(OUT'_a) + w_2 SN(INF'_a) + w_3 PBC(PER'_a)$$

, where $\forall w \in \{w_1, w_2, w_3\}, w \in [0, 1]$.

Let us introduce a running example to first demonstrate an application of TPB; later, we will continue this example to demonstrate our human-aware epistemic planning approach.

Example 1 Suppose that John, 13 years old, has social anxiety in new situations involving unknown people and unpredictable events. Such a place is a school cafeteria. The environment is usually noisy, with sounds of people talking, laughing, and shouting; chairs are being pulled and hitting each other. Different kinds of social interactions may be involved, such as with the cashier. The queuing and payment procedure is stressful to John, due to people in the queue standing in front and behind; he may reason about how these people will think of John's behavior, and possible complications that may arise in the payment procedure. There could be a large number of people in the cafeteria and most tables could be occupied. While looking for a table, John may occasionally get eye contact with people, and several people may walk past John as he searches the cafeteria for a table. John gets anxiety when he is unsure about the procedures. Any new or unpredictable aspect of the situation can lead to stress, such as not knowing how to queue and pay for the food, how to find a table, and what to do with the tray after the meal is finished.

In the following running example, scores are presented in a 7 point bipolar scale ranging from -3 to 3, which later on can be normalized to a score between 0.0 and 1.0.

As John enters the school cafeteria with his teacher, he can hear the noisy sound of people talking and laughing. Due to the noise, he expects there to be a large number of people in there (+3) and evaluates this as unsettling (-1). As he walks into the lobby area, he can see multiple unknown people, and expect it to be likely (+2) that these people may be looking at him as he passes them. This is stressful (-2) for John, as he evaluates people looking at him as negative. John is not sure about the queuing procedure. He expects the queuing procedure to be complicated (+3) and evaluate his controllability of this procedure to be low (-1). However, John values his teacher's expertise in this situation (+3), and he has confidence in following her lead (+3).

By analysing these simplified sets of beliefs through the theory of planned behavior, John's attitude $A(OUT'_a)$ toward the behavior, the expectancy of people, can be calculated as $-3 = (+3) * (-1)$; John's subjective norms $SN(INF'_a)$, influence of people looking at him, can be calculated as $-4 = (+2) * (-2)$; and John's perceived behavioral control $PBC(PER'_a)$ of standing in the queue, can be calculated as $6 = ((+3) * (-1)) + ((+3)*(+3))$. Together, these values can be summed to calculate the behavioral intention $BI(a)$, i.e., $-9 = (-9)+(-6)+(6)$. This intention-score can possibly be turned positive by removing factors of uncertainty, and adding factors of motivation.

The system can approach this by, e.g., decreasing sounds, decreasing people, and adding guides that show how to proceed in the queue, how to pick food, and how to interact with the cashier. The changes in expectancy and value of these adaptations can then be scored in a similar way as seen in the above example, and an estimation of John's behavioral intention can be achieved.

Human-Aware Transition Systems and Action Reasoning

Let us start introducing some basic concepts of action specification languages. In particular, we will introduce a generic syntax and semantics of an action language called C_{TPB} that is based on other actions languages such as C_{TAIL} that was introduced by Dworshak *et al.* (Dworschak *et al.* 2008).

C_{TPB} is an extension of the language \mathcal{A} , which was introduced by Gelfond and Vladimir (Gelfond and Lifschitz 1993).

The alphabet of C_{TPB} consists of two nonempty disjoint sets of symbols \mathbf{F} and \mathbf{A} . They are called the set of fluents \mathbf{F} and the set of actions \mathbf{A} . A *fluent* expresses the property of an object in a world, and forms part of the description of states of this world. A *fluent literal* is a fluent or a fluent preceded by \neg . A *state* σ is a collection of fluents. We say a fluent f holds in a state σ if $f \in \sigma$. We say a fluent literal $\neg f$ holds in σ if $f \notin \sigma$.

Definition 5 (Human-aware alphabet) Let \mathbf{A} be a non-empty set of actions and \mathbf{F} be a non-empty set of fluents.

- $\mathbf{F} = \mathbf{F}^E \cup \mathbf{F}^H$ such that \mathbf{F}^E is a non-empty set of fluents describing observable items in an environment and \mathbf{F}^H is a non-empty set of fluents describing the mental-states of humans: attitude, subjective norm and control.
- $\mathbf{A} = \mathbf{A}^E \cup \mathbf{A}^H$ such that \mathbf{A}^E is a non-empty set of actions that can be performed by a software agent and \mathbf{A}^H is non-empty set of actions that can be performed by a human agent.

C_{TPB} is defined by three sub-languages: an action description language, an action observation language and an action query language.

Definition 6 A human-aware domain description language $D^h(\mathbf{A}, \mathbf{F})$ in C_{TPB} consists of expressions of the following form:

- (1) $(a \text{ causes } f_1, \dots, f_n \text{ if } g_1, \dots, g_m)$
- (2) $(a \text{ influences attitude } f \text{ if } g_1, \dots, g_m)$
- (3) $(a \text{ influences subjective norm } f \text{ if } g_1, \dots, g_m)$
- (4) $(a \text{ influences control } f \text{ if } g_1, \dots, g_m)$
- (5) $(f_1, \dots, f_n \text{ influences attitude } f)$
- (6) $(f_1, \dots, f_n \text{ influences subjective norm } f)$
- (7) $(f_1, \dots, f_n \text{ influences control } f)$
- (8) $(f_1, \dots, f_n \text{ if } g_1, \dots, g_m)$
- (9) $(f_1, \dots, f_n \text{ triggers } a)$
- (10) $(f_1, \dots, f_n \text{ allows } a)$
- (11) $(f_1, \dots, f_n \text{ inhibits } a)$
- (12) $(f_1, \dots, f_n \text{ promotes } a)$
- (13) $(f_1, \dots, f_n \text{ demotes } a)$
- (14) $(\text{noconcurrency } a_1, \dots, a_n)$
- (15) $(\text{default } f)$

where $a_i \in \mathbf{A}$ ($0 \leq i \leq n$) and $f_j \in \mathbf{F}$, ($0 \leq j \leq n$).

The semantics of a domain description $D(A, F)$ is defined in terms of transition systems. An interpretation I over F is a complete and consistent set of fluents.

Definition 7 A state $s \in S$ of the domain description $D^h(A, F)$ is an interpretation over F such that for every static clausal law $f_1, \dots, f_n \text{ if } g_1, \dots, g_m \in D^h(A, F)$,

we have $\{f_1, \dots, f_n\} \subseteq s$ whenever $\{g_1, \dots, g_m\} \in s$. S denotes all the possible states of $D^h(A, F)$.

Definition 8 The action observation language of C_{TPB} consists of expressions of the following form:

$$(f \text{ at } t_i) \quad (a \text{ occurs_at } t_i) \quad (16)$$

where $f \in \mathbf{F}^H$, a is an action and t_i is a point of time.

Definition 9 (Action Theory) Let D be a domain description and O be a set of observations. The pair (D, O) is called an action theory.

Definition 10 (Trajectory Model) Let (D, O) be an action theory. A trajectory $s_0, A_1, s_1, A_2, \dots, A_n, S_n$ of D is a trajectory model of (D, O) , if it satisfies all observations of O in the following way:

1. if $(f \text{ at } t) \in O$, then $f \in s_t$
2. if $(f \text{ occurs_at } t) \in O$, then $a \in A_{t+1}$.

Let us observe that giving a trajectory $s_0, A_1, s_1, A_2, \dots, A_n, S_n$ where $A_i \subseteq \mathbf{A}$ ($0 \leq i \leq n$) and \mathbf{A} is a set of actions that can be performed either by a human-agent or a software planner-agent. Actions performed by a human-agent can be movement between areas in the environment, while actions by a software planner-agent are adaptations of the environment that indirectly influences a human-agent's mental-state fluents, i.e., attitude, subjective norm and control.

Human-aware theoretical framework

The aim of the human-aware planner agent is to adapt the environment in order to change the mental state of the human. The planner agent is modeled as an action theory (D, O) , utilizing a domain description D , which explains how human mental states relate to fluents of the environment, and a set of observations O , i.e., the observable environment, such as lights, sounds, and people, and the partly observable mental state of the human, i.e., attitude stress, normative stress, and control stress.

The agent models a human agent, utilizing the observed human actions in the environment and causal laws, specified for the planner agent's adaptive actions of the environment, to estimate the human's future actions. The planner agent's generated plan can be represented as a trajectory P ; a sequence of actions a_0, \dots, a_n leading from a starting configuration of fluents s_0 towards a goal configuration of fluents s_n denoted by $Plan P = \langle s_0, a_0, s_1, a_1, \dots, a_n, s_n \rangle$. Each action directly changes fluents of the environment and indirectly changes fluents of the mental state of the human. See Table 1 and 2 for a demonstration of the planner agent's planning process in response to different sets of observations. Each table presents a transition in the cafeteria scenario. Table 1 presenting a transition from the lobby area to the queue area, and Table 2 a transition from the queue area to the food area. Each transition has a set of observations and a set of planner actions, i.e., adaptations of the environment.

Human-Aware Epistemic Planning

The states of the environment, according to the theory of planned behavior, is based on three dimensions; attitudes, subjective norms and perceived behavioral control. According to TPB, a change in any of these dimensions can affect the behavioral intention, which in turn is a predictor to behavioral achievement.

Each state of the environment is a unique configuration of these environment variables, which describes the motivational factor of the state. This definition of motivation can be used to try to predict the human's behavioral intention. Changing these variables can configure states that provide appropriate levels of difficulty, assisting the human in completing the scenario while still providing sufficient practice.

In order to successfully provoke behavior change, the adaptation should provoke a change in the evaluation of one of three sets of beliefs, i.e., the evaluation of the attitudes, the subjective norms, or the perceived behavioral control. This can be achieved through appropriate actions that take in consideration the current state of the human and the environment, and does proper adaptations which results in a change of beliefs.

The human enters the scenario with a goal. In order to achieve this goal there are a variety of sub-goals along the way. Whether or not the human will be able to meet these interim goals depends on the difficulty of the environment and on the state of the human. The intelligent system thus has the task of adapting the environment at every stage of the scenario to help the human reach the next intermediate goal, and finally reach the final goal.

A goal is represented by a configuration of fluents, i.e., variables of the environment, and the representation of the environment is composed by these possible configurations, defining the state space. The intelligent system's task is to find a configuration of fluents that best represents the next goal. In every such goal, some fluents are required, e.g., a certain human position in the environment or that certain objects must be present in the scene, etc. However, fluents can also be dynamic and instead have a preferred value, e.g., the human's stress can vary but is preferred to be as low as possible. Similarly, the number of stressors can vary but is preferred to meet the learning objective. In this way, the intelligent system aims to find a plan that leads the human from the current configuration of fluents to the next goal.

The intelligent architecture can through a representation of the environment, interaction constraints and a set of causal rules make a plan on how to adapt the environment in order to promote human behavior change. The planning starts with a disjunctive logic planner (DLV-K) modelled by concepts based on the theory of planned behavior. The DLV-K planner generates alternative plans for adaptations of the virtual environment. The next step, which is a central strategy in this architecture, is the conduction of *counterfactual actions*. The system creates a *counterfactual instance* of the scenario, i.e., an evaluation process in which the system makes a specific assumption w.r.t. the human's future behavior. In this evaluation process, the alternative plans generated by the DLV-K planner are scored in relation to weights tailored to the individual human. In the counterfactual instance,

Test case	Stress observation	Planner agent actions with plan length 4
1	A(1), N(1), C(1)	<i>(no action); (no action); (no action); move(queue)</i>
2	A(1), N(1), C(3)	<i>increase_guides(2); (no action); (no action); move(queue)</i>
3	A(1), N(3), C(3)	<i>increase_guides(2); decrease_light(2); decrease_light(1); move(queue)</i>
4	A(2), N(3), C(3)	<i>increase_guides(2); decrease_light(2); decrease_sound(2); move(queue)</i>
5	A(3), N(3), C(1)	<i>decrease_sound(2); decrease_light(2); decrease_light(1); move(queue)</i>
6	A(3), N(3), C(3)	<i>increase_guides(2); decrease_sound(2); decrease_light(2); move(queue)</i>

Table 1: This table presents test runs of the human-aware planning process, presenting 6 combinations of observed attitude stress (A), normative stress (N), and control stress (C). Value 3 = high; Value 2 = medium; Value 1 = low. The initial observation of the environment in all cases are: at(human, lobby), lights(3), sound(3), guides(1). The goal in all cases are: at(human, queue). This means that the planner’s goal is to move the figurative human from the Lobby area to Queue area, while following the restrictions specified for the transition. The transition from Lobby to Queue requires all three stress fluents to have values below high. The plan length is set to 4, i.e., 3 planner actions, and 1 figurative move of the human. Each adaptation of the environment causes a change in stress, i.e., in A, N or C, of the figurative human, which finally results in an acceptable move to the next area, i.e., the next goal of the planner. The action decrease_sound(X) relieves attitude stress, the action decrease_light(X) relieves normative stress, and the action increase_guides(X) relieves control stress.

Test case	Stress observation	Planner agent actions with plan length 6
7	A(1), N(1), C(1)	<i>(no action); (no action); (no action); (no action); (no action); move(food)</i>
8	A(1), N(1), C(3)	<i>increase_guides(2); increase_guides(3); (no action); (no action); (no action); move(food)</i>
9	A(1), N(3), C(3)	<i>increase_guides(2); decrease_light(2); increase_guides(3); decrease_light(1); (no action); move(food)</i>
10	A(2), N(3), C(3)	<i>decrease_sound(2); increase_guides(2); increase_guides(3); (no action); decrease_light(2); move(food)</i>
11	A(3), N(3), C(1)	<i>decrease_sound(2); decrease_light(2); decrease_sound(1); (no action); (no action); move(food)</i>
12	A(3), N(3), C(3)	<i>increase_guides(2); decrease_sound(2); decrease_light(2); increase_guides(3); (no action); move(food)</i>

Table 2: This table presents test runs of the human-aware planning process, presenting 6 combinations of observed attitude stress (A), normative stress (N), and control stress (C). Value 3 = high; Value 2 = medium; Value 1 = low. The initial observation of the environment in all cases are: at(human, queue), lights(3), sound(3), guides(1). The goal in all cases are: at(human, food). This means that the planner’s goal is to move the figurative human from the Queue area to Food area, while following the restrictions specified for the transition. The transition from Queue to Food requires attitude stress fluents and normative stress fluents to have values below high, and control stress fluents to have value below medium. The plan length is set to 6, i.e., 5 planner actions, and 1 figurative move of the human. Each adaptation of the environment causes a change in stress, i.e., in A, N or C, of the figurative human, which finally results in an acceptable move to the next area, i.e., the next goal of the planner. The action decrease_sound(X) relieves attitude stress, the action decrease_light(X) relieves normative stress, and the action increase_guides(X) relieves control stress. A higher focus on control stress relief can be seen in this transition.

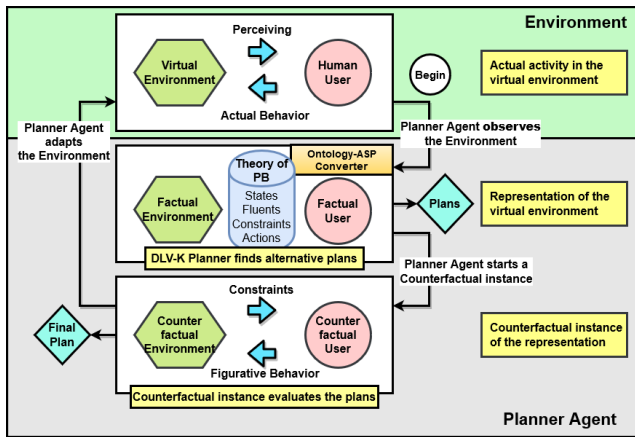


Figure 3: System flowchart. The agent holds a representation of the virtual environment and the state of the user. In order to generate a plan, the agent creates a counterfactual instance of the representation of the environment. In the counterfactual instance, the agent can make forecasts of the user’s activity and stress reactions by utilizing the specified constraints inspired by theory of planned behavior. When a plan is generated, the planner agent can finally adapt the virtual environment. Actual behavior of the user can then be observed as the user perceives and reacts to the actual changes of the environment.

the system reflects upon future “what if” scenarios in which a *counterfactual human* walks through the scenario in ways promoted by the system. The counterfactual human’s stress is aggregated or relieved depending on counterfactual adaptations of the environment. In this way, the counterfactual adaptations can explore what could happen if an actual adaptation was made. When this counterfactual process is complete, a final plan is selected specifying how to do actual adaptations of the environment for promoting the human’s actual behavior (see Figure 3).

The system takes the current state of the environment as input, and matches this state against the causal rules that are specified, the User Model (UM) and the Environment Model (EM). The UM represents the human’s current stress level, divided into three fluents based on the theory of planned behavior, i.e., attitude stress, normative stress, and behavioral control stress. The UM also defines an estimation of the human’s current position in the scenario. The EM represents the environment’s current composition of fluents, i.e., stressors or motivators in the environment. These models, are comprised of (1) fluents that can be changed through adaptive actions of the system, (2) weights that are tailored and learned for the individual human, and (3) static variables, e.g., the position linked to a specific state.

Environment fluents Environment fluents (EF) are variables that can be observed and adapted in the environment, such as objects, sounds, locations, behaviors, etc. Environment fluents are grouped into three sets of fluents: (1) attitude stress-provoking fluents, (2) normative stress-provoking fluents, and (3) behavioral control stress-

provoking fluents. An environment variable has the role of provoking certain changes of attitudes, subjective norms, or perceived behavioral control in the UM. In the pursuit of reaching goals in the environment, and in this process changing the variables of the UM, the variables of the EM must be adapted.

Tailored weights Tailored weights (TW) are values associated with each set of environment fluents in a state. The tailored weights are: (1) attitude weights, (2) normative weights, and (3) control weights. The tailored weights represent the impact of the corresponding set of fluents on the human’s stress in a state. A specific sub-activity can in this way be more susceptible to certain stressors for an individual. The attitude weights specify the impact of attitude fluents, the normative weights specify the impact of normative fluents, and the control weights specify the impact of behavioral control fluents. In addition, a specific tailored weight is associated with each fluent in the set, providing a greater degree of personalization. Thus, a state can have a more or less, e.g., normative impact on stress, and a specific stressor in that state can have its own weight that enhances or inhibits this specific fluent in regard to stress.

Counterfactual actions and Epistemic actions The actions of the system can be grouped in two sets of actions: (1) counterfactual actions and (2) epistemic actions.

Counterfactual actions are executed in the counterfactual instance by the system. In this phase, the human is represented as an agent that can move in the scenario according to the specified constraints and causal rules of the environment. In this process, the human’s stress levels are affected by counterfactual adaptations of the forecasted scenario. The counterfactual process attempts to determine the best plan that the system should perform in order to promote behavior-change, or to encourage the human doing an activity.

Epistemic actions are the actions that are executed in the environment in order to promote behavior. The fluents of the environment can be changed through epistemic actions, an exploration process that observes actual human behavior in response to adaptations. The logic planner generates alternative plans that follow the causal effects of the actions that are specified. A general challenge with logic planners is to select one of the generated plans. Here is where the counterfactual instance comes into play. In the counterfactual instance, the alternative plans that are generated by the logic planner are evaluated to find the best suitable plan. Evaluations are done in accordance with the equations specified in the theory of planned behavior (see Definitions 1, 2, and 3). The counterfactual instance is created, in which each plan’s potential outcome is evaluated in relation to attitude fluents, normative fluents, and control fluents, and their corresponding tailored weights. A plan is selected and finally executed through epistemic actions that changes the environment. New observations can then be made to do further planning.

Causal effects of actions Actions are defined by a set of preconditions and post-conditions. Preconditions specify when actions are executable, i.e., in which configuration of

fluents can, e.g., an increase or decrease of fluents be executed. Post-conditions specify which fluents will change after the action is executed, e.g., an increase of a fluent value. Each action has a set of causal effects defined by causal laws.

Causal effects of actions determine changes in the environment after the execution of planner actions. In addition, a set of indirect causal effects are defined which specify the estimated result of changes in the environment on the human's stress levels. These rules eventually help to predict behavioral intention and behavioral achievement of the human through the planning process. The planner agent conducts these counterfactual actions, i.e., it simulates what would be the result if the human agent executes a certain action with the current configuration of environment fluents. The system makes counterfactual adaptations to the environment, leading to counterfactual movements of the human, in order to generate a plan that is finally executed resulting in epistemic actions, i.e., actual adaptations of the environment.

Discussion

This work has defined a human-aware planning architecture that is utilizing action reasoning to represent behaviors, constraints, states, goals, and causal rules for actions to adapt the environment. The composite structure presented in this work is novel in the area of human-aware planning, using the theory of planned behavior as its theoretical base, in order to understand the human's behavior and provide suitable adaptations of the environment.

An early-stage proof of concept prototype has been implemented to demonstrate the potential of the intelligent architecture. The prototype's focus is on one specific scenario, representing a cafeteria environment. This can limit the potential insights from an evaluation. In order to get a wider perspective, a set of scenarios should be implemented, evaluated and compared. A general model of the proposed architecture can be applied in any human-centered scenario for promoting behavior change, given that relevant knowledge for the specific domain has been elicited and incorporated in the model. The architecture can thus be evaluated in a variety of settings.

Related Work

The human-aware planning problem has in general been explored in scenarios where a robot is situated in an environment involving humans (Chakraborti, Sreedharan, and Kambhampati 2018). A survey of recent efforts in human-aware planning is presented in (Chakraborti, Sreedharan, and Kambhampati 2018) and in (Ahrendt, Fährndrich, and Albayrak 2014). In contrast, this paper has explored how a software agent can understand the human's behavior in an environment, and generate a plan on how to adapt the environment in order to promote human actions.

There is a diverse body of research related to the ideas presented in the current work. Plan recognition as planning, originally introduced by Ramirez and Geffner (Ramirez and Geffner 2009), use planning algorithms to enable an agent recognize the goals and plans of other agents. This relates

to the current study where the agent generates plans for estimating and promoting human behavior.

Empathetic Planning (Shvo and McIlraith 2019) is introduced. In their work, empathy is defined as the ability to understand and share the thoughts and feelings of another. Following this definition, an assistive empathetic agent is formalized able to reason about the preferences of an empathizee (Shvo and McIlraith 2019). The current study uses a formalization of the theory of planned behavior (TPB) for modeling the human, i.e., in an attempt to understand the human's preferences and feelings in relation to the environment.

In Active Goal Recognition (AGR) (Shvo and McIlraith 2020), an AGR agent actively senses and acts as part of the goal recognition process. While pursuing its goal, the agent executes sensing and world altering actions (Shvo and McIlraith 2020). This relates to the notion of epistemic actions in the current work, where the agent actively senses the environment and the human's behavior in response to world altering actions.

Goal Recognition Design (Keren, Gal, and Karpas 2014) presents way to modify a domain model in order for an agent acting in the model to reveal its objective as early as possible (Keren, Gal, and Karpas 2014). This is related to the current work where an agent adapts the environment in order to observe, and estimate, human responses and promote human actions.

In the academic community, a commonly applied concept for implementing cognitive agents is the so-called Belief-Desire-Intention (BDI) approach (Rao and Georgeff 1995). BDI agents can be considered, roughly speaking, *rational agents* that, based on their beliefs about the world, act fulfill their desires to the greatest extent possible. However, in the context of our approach to human-aware planning, a theory of human cognition is needed that focusses on the non-rational aspects of human cognition. Consequently, TPB, which bears some resemblance to Kahneman's and Tversky's prospect theory (Daniel, Amos, and others 1979), but has a stronger focus on social and emotional aspects can be considered a suitable *boundedly rational* formal model. In that, our work attempts to solve a different problem than works on BDI agents and theory of mind (Panisson et al. 2019; Kampik, Nieves, and Lindgren 2019), which, even though they might relax the agents' rationality properties, but provide primarily "computationally"-oriented models and not models rooted in behavioral psychology.

Conclusion and Future Work

In contrast to typical applications of human-aware planning, where a robot is situated in an environment populated by humans (Köckemann, Pecora, and Karlsson 2014), this study has explored how a software agent can promote human actions by adapting the environment. In the use-case of this work, in contrast to plan merging (Gravot and Alami 2001) in robot architectures, the agent can only adapt the environment attempting to encourage human actions.

Possible use-cases are training application using virtual reality and augmented reality technologies (Kumar et al.

2020), enabling a human to practice in realistic environments. The architecture is personalizable in that stressors of the environment can be weighted for each individual. Future work concerns the integration with automated learning approaches, e.g., reinforcement learning (Sutton, Barto, and others 1998), to shape the weights, i.e., the impact of certain stressors on the specific individual, in a dynamic manner. In this way, the weights can be adjusted during training sessions.

References

- [Ahrndt, Fährndrich, and Albayrak 2014] Ahrndt, S.; Fährndrich, J.; and Albayrak, S. 2014. Human-aware planning: A survey related to joint human-agent activities. In *Trends in Practical Applications of Heterogeneous Multi-Agent Systems. The PAAMS Collection*. Springer. 95–102.
- [Ajzen and others 1991] Ajzen, I., et al. 1991. The theory of planned behavior. *Organizational behavior and human decision processes* 50(2):179–211.
- [Bolander 2017] Bolander, T. 2017. A gentle introduction to epistemic planning: The del approach. *arXiv preprint arXiv:1703.02192*.
- [Chakraborti, Sreedharan, and Kambhampati 2018] Chakraborti, T.; Sreedharan, S.; and Kambhampati, S. 2018. Human-aware planning revisited: A tale of three models. In *IJCAI-ECAI XAI/ICAPS XAIP Workshops*.
- [Chakraborti 2018] Chakraborti, T. 2018. *Foundations of Human-Aware Planning-A Tale of Three Models*. Ph.D. Dissertation, Arizona State University.
- [Cirillo, Karlsson, and Saffiotti 2010] Cirillo, M.; Karlsson, L.; and Saffiotti, A. 2010. Human-aware task planning: An application to mobile robots. *ACM Transactions on Intelligent Systems and Technology (TIST)* 1(2):1–26.
- [Daniel, Amos, and others 1979] Daniel, K.; Amos, T.; et al. 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47(2):263–291.
- [Dworschak et al. 2008] Dworschak, S.; Grell, S.; Niki-forova, V.; Schaub, T.; and Selbig, J. 2008. Modeling Biological Networks by Action Languages via Answer Set Programming. *Constraints* 13(1):21–65.
- [Eiter et al. 2003] Eiter, T.; Faber, W.; Leone, N.; Pfeifer, G.; and Polleres, A. 2003. A logic programming approach to knowledge-state planning, ii: The dlvc system. *Artificial Intelligence* 144(1-2):157–211.
- [Fishbein and Ajzen 2011] Fishbein, M., and Ajzen, I. 2011. *Predicting and changing behavior: The reasoned action approach*. Taylor & Francis.
- [Frith and Frith 2005] Frith, C., and Frith, U. 2005. Theory of mind. *Current biology* 15(17):R644–R645.
- [Gelfond and Lifschitz 1993] Gelfond, M., and Lifschitz, V. 1993. Representing Action and Change by logic programs. *Journal of Logic Programming* 17(2,3,4):301.
- [Gelfond and Lifschitz 1998] Gelfond, M., and Lifschitz, V. 1998. Action languages.
- [Giorgini et al. 2002] Giorgini, P.; Mylopoulos, J.; Nicchiarelli, E.; and Sebastiani, R. 2002. Reasoning with goal models. In *International Conference on Conceptual Modeling*, 167–181. Springer.
- [Gravot and Alami 2001] Gravot, F., and Alami, R. 2001. An extension of the plan-merging paradigm for multi-robot coordination. In *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No. 01CH37164)*, volume 3, 2929–2934. IEEE.
- [Kampik, Nieves, and Lindgren 2019] Kampik, T.; Nieves, J. C.; and Lindgren, H. 2019. Empathic autonomous agents. In Weyns, D.; Mascardi, V.; and Ricci, A., eds., *Engineering Multi-Agent Systems*, 181–201. Cham: Springer International Publishing.
- [Keren, Gal, and Karpas 2014] Keren, S.; Gal, A.; and Karpas, E. 2014. Goal recognition design. In *Twenty-Fourth International Conference on Automated Planning and Scheduling*.
- [Köckemann, Pecora, and Karlsson 2014] Köckemann, U.; Pecora, F.; and Karlsson, L. 2014. Grandpa hates robots-interaction constraints for planning in inhabited environments. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*.
- [Kumar et al. 2020] Kumar, R.; Samarasekera, S.; Acharya, G.; Wolverton, M. J.; Ayan, N. F.; Zhu, Z.; and Villamil, R. 2020. Method and apparatus for mentoring via an augmented reality assistant. US Patent 10,573,037.
- [Panisson et al. 2019] Panisson, A. R.; Sarkadi, S.; McBurney, P.; Parsons, S.; and Bordini, R. H. 2019. On the formal semantics of theory of mind in agent communication. In Lujak, M., ed., *Agreement Technologies*, 18–32. Cham: Springer International Publishing.
- [Prochaska and Velicer 1997] Prochaska, J., and Velicer, W. 1997. The transtheoretical model of health behaviour change. *Am J Health Promot* 12:38–48.
- [Ramirez and Geffner 2009] Ramirez, M., and Geffner, H. 2009. Plan recognition as planning. In *Proceedings of the 21st international joint conference on Artificial intelligence. Morgan Kaufmann Publishers Inc*, 1778–1783.
- [Rao and Georgeff 1995] Rao, A. S., and Georgeff, M. 1995. Bdi agents: from theory to practice. In *Proceedings of the First International Conference on Multiagent Systems*, volume 95, 312–319.
- [Shvo and McIlraith 2019] Shvo, M., and McIlraith, S. A. 2019. Towards empathetic planning. *arXiv preprint arXiv:1906.06436*.
- [Shvo and McIlraith 2020] Shvo, M., and McIlraith, S. A. 2020. Active goal recognition. In *AAAI*, 9957–9966.
- [Sutton, Barto, and others 1998] Sutton, R. S.; Barto, A. G.; et al. 1998. *Introduction to reinforcement learning*, volume 135. MIT press Cambridge.