Simulations of a Computational Model for a Virtual Medical Assistant

Aryana Collins Jackson[1][a], Marlène Gilles[1][b], Eimear Wall[2], Elisabetta Bevacqua[1][c], Pierre De Loor[1] and Ronan Querrec[1]

[1]*ENIB, Lab-STICC UMR 6285 CNRS, Brest, France*

[2]*Health Service Executive, Ireland*

Keywords: Intelligent Agents, Embodied Conversational Agents, Human-computer Interaction, Situational Leadership, Medicine, Simulation.

Abstract: We propose a virtual medical assistant to guide both novice and expert caregivers through a procedure without the direct help of medical professionals. Our medical assistant uses situational leadership to handle all interaction with a caregiver, which works by identifying the readiness level of the caregiver in order to match them with an appropriate style of communication. The agent system (1) obtains caregiver behavior during the procedure, (2) calculates a readiness level of the caregiver using that behavior, and (3) generates appropriate agent behavior to progress the procedure and maintain a positive interaction with the caregiver.

1 INTRODUCTION

Due to advancements in virtual agents and telemedicine, caregiving in various capacities can now be done at a distance. For example, virtual agents can perform health assessments (Montenegro et al., 2019) and act as liaisons with medical staff (Bickmore et al., 2015). In these examples, the agent interacts directly with the patient. However, there are cases in which an agent should interact only with a caregiver. When a medical procedure is necessary and there are no trained medical staff nearby, the individual present must assume a caregiving role, and they will need guidance on preserving the patient's health. During these situations, a permanent virtual medical assistant could be useful (Nakhal, 2017).

The assistant will communicate with the caregiver and will also communicate to a team of medical experts standing by on a different site monitoring the procedure. Because of the possible latency and disruptions in communication between the remote site and the site where the experts are, the virtual assistant must be capable of guiding the caregivers through the entire procedure without relying on the help of the medical experts.

A huge part of organizing a successful procedure is maintaining a positive interaction with the caregiver while maintaining the health of the patient (Hjortdahl et al., 2009; Henrickson et al., 2013; Yule et al., 2008; Flin et al., 2010). Both the caregiver and the agent must work towards a common goal: the completed procedure that preserves the health of the patient. To foster this positive working alliance between agent and caregiver, situational leadership is employed (Hersey et al., 1988). In situational leadership, a leader adapts their behavior to the follower according to their experience level and expertise. Situational leadership involves four readiness levels which directly correspond to four leadership styles. The readiness levels involve varying degrees of ability, which refers to a person's competence when completing tasks, and willingness, which refers to their confidence and interest in completing those tasks. In this scenario, the virtual agent assumes the role of leader, and the caregiver assumes the role of follower. Situational leadership is covered in more depth in section 2.

Previously when situational leadership has been studied, it has been implemented (Sims et al., 2009;

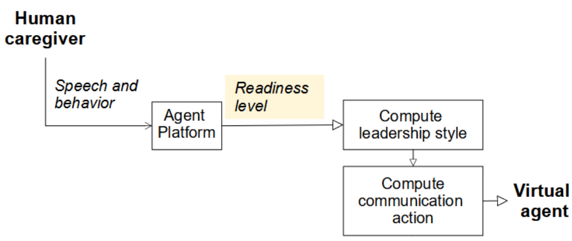


Figure 1: A basic representation of the agent framework.

[a] https://orcid.org/0000-0002-4385-5258
[b] https://orcid.org/0000-0003-1806-1672
[c] https://orcid.org/0000-0002-7117-3748

Jackson, A., Gilles, M., Wall, E., Bevacqua, E., De Loor, P. and Querrec, R.
Simulations of a Computational Model for a Virtual Medical Assistant.
DOI: 10.5220/0010910700003116
In *Proceedings of the 14th International Conference on Agents and Artificial Intelligence (ICAART 2022) - Volume 1*, pages 94-105
ISBN: 978-989-758-547-0; ISSN: 2184-433X

Hersey et al., 1988; Lourdeaux et al., 2019) or simulated for intended use in human-human relationships (Bosse et al., 2017; Bernard, 2020). Virtual leaders have used different methods of guiding humans and of choosing behavior that is personalized to the individual humans. However, our work is the first to present both a computational model of situational leadership and an agent framework in which a virtual agent leads humans (see Figure 1 for a basic representation of the agent framework and Figure 6 for a detailed view). We contribute an implementation of situational leadership in which we (1) quantify the factors that determine readiness level, (2) create an algorithm for calculating readiness level based on these criteria, and (3) create a virtual agent system that uses readiness level and human behavior to create personalized behavior.

While our work is conducted with the context of a medical procedure in mind, the model itself is flexible enough to be applicable to any situation in which a human being must be led through a series of steps by a virtual assistant.

In this paper, we detail what situational leadership is and how it works, we discuss the current state of the art regarding virtual leaders, we detail our computational model of situational leadership, we explain how our model is implemented using Mascaret, and we discuss our conclusions and future work.

2 SITUATIONAL LEADERSHIP

As discussed in the introduction, situational leadership encompasses four follower readiness levels and four leadership styles. The four readiness levels (denoted with an *R* and a number) are listed below (Hersey et al., 1988):

1. R1: Low ability, low willingness;
2. R2: Low to some ability, high willingness;
3. R3: High ability, variable willingness;
4. R4: High ability, high willingness.

Readiness levels R1 and R2 need the leader's help while readiness levels R3 and R4 should be able to complete tasks without the leader's help.

When a leader employs situational leadership, they perform both task and relationship behavior to guide the follower. Task behavior corresponds to a follower's competence (Bedford and Gehlert, 2013). Relationship behavior refers to how the leader fosters and maintains a positive working alliance with the follower and corresponds to a follower's willingness, or confidence (Bedford and Gehlert, 2013). The leadership styles that correspond to these readiness levels are (Hersey et al., 1988):

1. Directing: high task, low relationship behavior;
2. Coaching: high task, high relationship behavior;
3. Supporting: low task, high relationship behavior;
4. Delegating: low task, low relationship behavior.

Note that low relationship behavior is required for followers with low ability and low willingness because the person may not be interested in being encouraged. For followers with high ability and low to some willingness, the agent motivates the person because they have the skills necessary to complete the procedure successfully.

Follower behavior that can indicate a particular readiness level includes both prior experience and knowledge and non-anatomical behavior such as the number of questions asked, the amount of hesitation after receiving tasks, etc. (Collins Jackson et al., 2019; Bosse et al., 2017; Hersey et al., 1988):

Situational leadership is an effective method of managing followers because it accounts for followers with both high and low ability and willingness. This model has been proven effective in an educational environment in which medical instructors implemented leadership styles in their classes and with individual students (Sims et al., 2009), work environments in which managers adopted leadership styles with their direct reports (Hersey et al., 1988), and in a clinical medical supervision context (Bedford and Gehlert, 2013). Situational leadership has also been implemented in a virtual environment to help train medical staff (Lourdeaux et al., 2019).

There is very limited work on computational models using situational leadership, and existing work involves human-human interaction only (Bosse et al., 2017; Bernard, 2020). Rarely are agents in positions of leadership in human-agent interaction. Our work models the work on situational leadership and also provides a new method in which agents and humans can interact by allowing an agent to lead a human. Our model permits an agent to maintain a positive interaction with a follower and interact with the human in a way that is most beneficial for them in addition to maintaining the health of the patient.

In the following section, we discuss previous work on virtual agent leaders.

3 VIRTUAL LEADERS

This work involves a virtual agent acting as a medical assistant who leads a caregiver through a medical procedure. Within the medical domain, there has been a lot of work regarding human-human relationships in the emergency room. The person leading a medical procedure holds an important role in that they manage both the procedure tasks, the health of the patient, and the interaction with the caregiver(s). Thus having the trust of the caregivers and being competent at their work are two of the most important qualities a leader can have (Hjortdahl et al., 2009). Additional qualities that a medical leader should embody include communication, negotiation, autonomy, creativity, and appreciation of the caregivers (Araszewski et al., 2014). Many of these qualities are therefore included in medical professional behavior taxonomies (Henrickson et al., 2013; Yule et al., 2008; Flin et al., 2010).

In terms of virtual agents in medicine, there are a variety of examples of virtual agents created for use specifically in medical situations: agents which act as liaisons between patients and physicians using prescripted speech (Bickmore et al., 2015), agents which are meant to connect emotionally with patients for mental health benefits (Yang and Fu, 2016), agents which act as companions for the elderly (Montenegro et al., 2019), and agents which conduct psychiatric interviews (Philip et al., 2020). In each of these examples, user engagement with the agent was dependent on whether they trusted the agent.

The concept of developing trust is inherent in all types of human-computer interactions, regardless of domain, as it is a prerequisite to a positive interaction (Hoegen et al., 2019; Kulms and Kopp, 2016; Lee et al., 2021). Additionally, trust leads to greater efficiency when the human is completing tasks (Kulms and Kopp, 2016). Situational leadership was developed knowing that followers' trust of their leader is hugely important to successful interactions, and that trust can be built and maintained when leaders interact in an appropriate manner to followers (Hersey et al., 1988). Thus employing situational leadership itself is a method of building and maintaining trust between a human and an agent.

Because our work involves an agent leading a human, we examined previous work in which agents exist in pedagogical scenarios. We again found that trust was important to student-tutor relationships and was a product of a supporting environment (Castellano et al., 2013; Saerbeck et al., 2010). These relationships are most successful when the agent adapts to the learner by responding based on his or her previous experience and level of knowledge (Pecune et al., 2010; Taoum et al., 2018; Cisneros et al., 2019; Querrec et al., 2018).

As mentioned in the introduction, few computational models on situational leadership have been developed (Bosse et al., 2017; Bernard, 2020). While these models have involved human-human interaction only, they have provided a foundation for computing readiness level and therefore determining leadership style from various behavioral parameters. One model is based entirely on follower behavior and allows the leader to adapt their leadership style by monitoring a follower's advancement through the different readiness levels (Bosse et al., 2017). This work in particular formed the backbone of our own research.

In the following section, we outline our own computational model of situational leadership when a virtual agent is a medical leader.

4 A COMPUTATIONAL MODEL OF SITUATIONAL LEADERSHIP

Readiness level is determined by examining follower behavior. Other research has analyzed behavior such as facial actions and head movements (Dermouche and Pelachaud, 2019). However, we do not utilize a human activity monitor in this project. While these kinds of non-verbal behaviors are certainly useful, we chose to prioritize the caregiver's ability to move around the environment without worrying about whether their faces were reliably detected. Instead, we focus on human behavior that can be monitored and input from a keyboard.

In this section, we discuss the parameters used in our model, we detail the follower behavior that are used, we discuss why we chose certain parameter values, we thoroughly explain how the model works, and we explain how our model can be used to adhere to guidelines set by the existing work on situational leadership.

4.1 Model Parameters

Our model uses several parameters and various caregiver behaviors to determine readiness level. From now on, we refer to individual behaviors as behavior criteria or simply criteria.

Two of the parameters needed in this algorithm, persistence and valence, are adapted from an existing model (Bosse et al., 2017), but the others are our own contributions. The parameters explained below are used in equations 1 and 2 (detailed in section 4.4) to calculate a *performance* value which describes how high the follower's performance is in regards to each criterion. These values change depending on the procedure. Before explaining how that cal-

Table 1: The behavior criteria used for computing readiness level and each of their parameters. The *persistence* values change depending on whether the follower has low or high ability and willingness.

	Criterion Name	Domain	Persistence: high	Persistence: low	Weight	Performance threshold
1	Error: action in task	ability	0.7	0.85	0.05	0
2	Error: action outside task	ability	0.1	0.85	0.25	1
3	Wrong resource chosen	ability	0.7	0.85	0.15	0.75
4	No resource chosen	ability	0.2	0.85	0.25	1
5	Action duration too short	ability	0.1	0.85	0.06	0
6	Action duration too long	ability	0.1	0.85	0.07	1
7	Question for help	ability	0.4	0.85	0.17	1
8	Hesitation	willingness	0.9	0.9	0.5	0.5
9	Question for reassurance	willingness	0.9	0.9	0.5	0

culation works, we explain what each parameter is:

Extent refers to the extent to which the follower exhibits the criterion and is described by either 0 or 1 where 0 indicates that the follower does not embody that criterion at all, and 1 indicates that the follower embodies that criterion to the fullest extent possible.

Domain refers to whether the criterion is an indicator of ability or willingness; criteria can either exist in the ability domain or the willingness domain.

Persistence describes how much the behavior during a previous task influences the current readiness level with a float on the interval $[0,1]$. A value of 0 indicates low persistence while a value of 1 indicates high persistence. The lower the persistence value, the faster the performance value will rise or fall when the extent to which the follower exhibits that criterion is high or low respectively (Bosse et al., 2017).

Weight refers to the importance of the criterion within the ability and willingness domains when determining readiness level and is described by a float on the interval $[0,1]$.

Performance threshold refers to the lowest performance value possible of that criterion before the follower could be considered to have low ability or low willingness.

4.2 Follower Behavior Criteria

In previous research, the extent to which a follower exhibits thirty-three different behaviors determines readiness level (Bosse et al., 2017). That model is based on an interaction between a student and a supervisor, and thus many of these behaviors were either irrelevant to a medical scenario (such as *feeling over-obligated* and *lacking self-esteem*) or were impossible to compute within a virtual environment (such as *defensive behavior* and *discomfort in body language*). Additionally, these thirty-three behaviors are categorized by readiness level, indicating that each behavior is indicative of only one readiness level. While our work is based on this research, in order to adapt it to the context of a medical procedure led by a virtual agent, some changes had to be made.

Using that existing list of behaviors (Bosse et al., 2017), we created our own list of behavior criteria (see Table 1). Instead of grouping them by readiness level, we group them by whether they are indicators of ability or willingness. Because all four readiness levels can be described as having ability and willingness that is high, low, or somewhere in between, followers of any style can embody each of the behaviors to varying extents. The list of criteria are explained below. Note that in the procedure, there are both tasks and actions. A task is a complete step, while an action is a sub-step of a task. Thus each task may involve multiple actions. A resource is a tool that can be used to perform an action. Each action has an appropriate duration that it is expected to take. All of this information exists within Mascaret, our agent framework, which is explained further in section 6.

1. *Error: action in task*: The follower has chosen to do an action out of order;
2. *Error: action outside task*: The follower has chosen to do an action that does not exist in the current task;
3. *Wrong resource chosen*: The follower has taken the wrong resource for the action;
4. *No resource chosen*: The follower has neglected to take a resource when one is required *or* tries to take a resource when none is required;
5. *Action duration too short*: The action duration is less than 0.9 times the expected action duration (as devised by our medical professional);
6. *Action duration too long*: The action duration is more than 1.1 times the expected action duration (as devised by our medical professional);
7. *Question for help*: The follower has asked a question because they do not know how to proceed;
8. *Hesitation*: The follower has hesitated for more than five seconds before beginning an action;

9. *Question for reassurance*: Also referred to as a clarifying question; the follower has asked a question to ensure what they are doing is correct.

4.3 Expert-defined Parameter Values

The values for the parameters in Table 1 were devised by consulting with the medical doctor on our team. The procedure that these values were created for is the diagnosis of abdominal pain, and with these values, we assume that the case is not urgent.

Because our context is rather specific, it was important to create our model with the guidance of our team's medical professional. Therefore, all of these values are very procedure specific. The values for persistence, weight, and performance threshold would change depending on the procedure.

For a non-urgent diagnosis of abdominal pain, the most important ability criteria are *Error: action outside task* and *No resource chosen* because they indicate that the follower does not understand the current task and how to reach the end goal. Each are given the highest weight of 0.25. If the follower is considered to be high-ability, these two errors change their performance values considerably with low persistence values of 0.1 and 0.2 respectively.

Question for help is the next most important criterion in the ability domain because it indicates that the follower does not know how to proceed without the agent's help. This criterion is given a moderate persistence of 0.4 when the follower is considered to have high-ability because while questions for help are indicative of more novice behavior, the follower's history should be more important than it is for *Error: action outside task* and *No resource chosen*.

Unlike the previous three criteria, there is room for human error in *Wrong resource chosen* where a follower may accidentally reach for the wrong resource, hence the higher persistence value of 0.7.

Error: action in task, *Action duration too short*, and *Action duration too long* have the lowest weights. The most important thing is that the task has been completed, and so errors out of order do not matter so much as long as the task is being completed. Persistence for this error is set quite high. Action completion duration times are not important unless there is a life-threatening emergency. However, the persistence is set low for the action duration errors as these can co-occur with other criteria, such as *Error: action in/outside task* and *Question for help* and therefore they should be taken quite seriously. Because of their low weight, short or long action duration times on their own do not generally affect the overall ability performance unless they happen repeatedly throughout the procedure.

Finally, *Hesitation* and *Question for reassurance* are weighted equally in the willingness domain. Each of these are equally indicative of a follower's willingness. Additionally, the persistence for both is set quite high at 0.9 because a follower's willingness should be calculated from their history rather than from a single task.

As shown in Table 1, there are two different values for persistence, one for followers are deemed to have low ability or willingness and one for those who are deemed to have high ability or willingness. These values were devised in order to account for a novice who does everything correctly versus an expert who makes one or two mistakes. During a medical procedure, even one small mistake can lead to serious consequences. When a follower displays high-ability behavior, the mistakes they make should have more impact on the performance value. The *Persistence: high* values are used when a follower begins the procedure with high ability or willingness. When the ability performance drops below the first ability threshold of 0.8525, then the *Persistence: low* values are used instead. *Persistence: low* values are also used for followers who begin the procedure with low ability or willingness.

Additionally, persistence changes based on the difficulty of the current action so that a follower who is only easily able to do easier actions and a follower who is able to do harder actions are not considered to be the same level. Each action in the procedure is assigned a value of 1 (most difficulty), 2 (medium difficulty), and 3 (least difficulty). If the action is assigned a 1, then the persistence value for that task decreases by 0.05 to ensure that the persistence is never 1 or 0. If the action is assigned a 3, then the persistence value for that task increases by 0.05. If the action is assigned a 2, there is no change to persistence. This ensures that more difficult tasks are weighted more when the performance value is calculated.

4.4 The Model Explained

The first step to determining readiness level is to calculate the performance value for each criterion. The performance value is a float on the interval $[0,1]$ where a value of 0 indicates low ability or willingness and a value of 1 indicates high ability or willingness.

Equation 1 is used to calculate the performance value v at time t for criterion c, where p represents the persistence and e represents the extent (Bosse et al., 2017).

$$v_{c,t} = (p_c * v_{c,t-1}) + ((1-p_c) * (1-e_{c,t})) \quad (1)$$

Equation 1 requires the performance value at time $t-1$. At the very beginning of the procedure, default performance values are used in place of $v_{c,t-1}$. Values of 0 or 1 are established for each criterion which correspond to the follower's previous experience and knowledge, called the follower profile. The follower profile establishes which readiness level the individual has prior to beginning the procedure and is based on previous procedures, external evaluations of the follower, and self-evaluations. For a follower profile indicating low or variable ability or willingness, the default performance value is 0, and for a follower profile indicating high ability or willingness, the default performance value is 1.

For example, if a follower profile indicates that the individual's readiness level is style 3, then they have high ability and variable willingness (see Figure 2). The default performance values for all the criteria in the ability domain are 1, and the default performance values for the criteria in the willingness domain are 0. Calculating the performance value v for *Error: action in task* for a follower of style 3 who has made an error at the very beginning of the procedure would look like this: $(0.1*1)+((1-0.1)*(1-1)) = 0.1$.

The performance value v is then fed back into the model in equation 1 the next time readiness level is calculated.

The next step of determining readiness level is to average all the performance values within each domain. Equation 2 is used where n refers to the number of criteria in the ability and willingness domains, $weight_c$ refers to the weight that each criterion holds, and $v_{c,t}$ refers to the performance value at time t of each criterion as calculated by equation 1. Equation 2 results in two values representing the follower's overall ability and willingness behavior O.

$$O_t = \sum_{c=1}^{n} weight_c * v_{c,t} \tag{2}$$

There are two different thresholds each for determining whether a follower has low or high ability and willingness. For the first, the performance thresholds shown in Table 1 are used to determine whether a follower has high or low ability and willingness. When using the performance thresholds in equation 1 for both e_c and $v_{c,t-1}$, and then using the resulting $v_{c,t}$ value in equation 2, we achieve the overall ability performance threshold of 0.8525 and the overall willingness performance threshold of 0.625. If a follower has an ability value below 0.8525, they have low ability, and if that value is above or equal to 0.8525, they have high ability. If a follower has a willingness value below 0.625, they have low willingness, and if that value is at or above 0.625, they have high willingness.

The second threshold is defined by ability and willingness equal to or greater than 0.95. This value was devised by examining example follower behavior and determining which were able to self-lead without the leader's help. Followers with both an ability and a willingness value equal to or greater than 0.95 are considered to be in R4. See Figure 2 for a visualization of the thresholds and readiness levels.

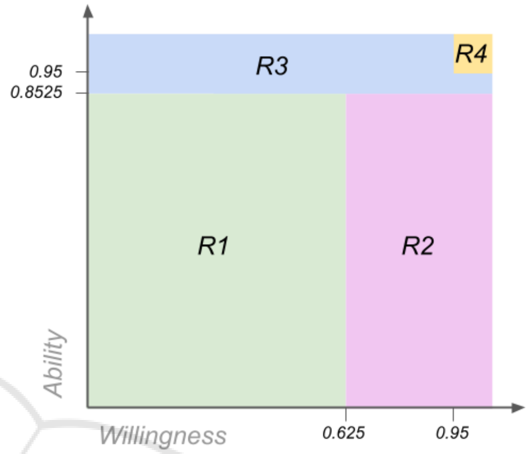


Figure 2: A visualization of the four readiness levels and the performance values needed in order to move between them.

As shown, each readiness level is color-coded uniquely. A follower's ability is contained on the vertical axis while willingness is described on the horizontal axis. Once the initial thresholds of 0.825 and 0.625 are reached for ability and willingness respectively, a follower is considered to be R3. Because R4 is designated for completely self-sufficient followers, the thresholds for R4 are higher. Ensuring that movement between the readiness levels is linear is discussed later in this section.

For examples of how this all works, refer to Figure 3. In this example procedure, *Error: action outside task* is examined. There are ten actions with the following action difficulties: 3, 3, 3, 3, 2, 2, 1, 2, 1, and 2. A follower with an existing readiness level of R1 and a follower with an existing readiness level of R4 perform the exact same steps; they each try to do an action that is outside the scope of the task for actions 5, 7, and 9. As shown, the performance value v dips for both the R1 and R4 follower, but the errors have a greater effect on the R4 follower who is supposed to be an expert.

Note that the extent values for each criterion remain the same until there is another opportunity to change that value. For example, the extent value e of *Error: action outside task* only changes when the

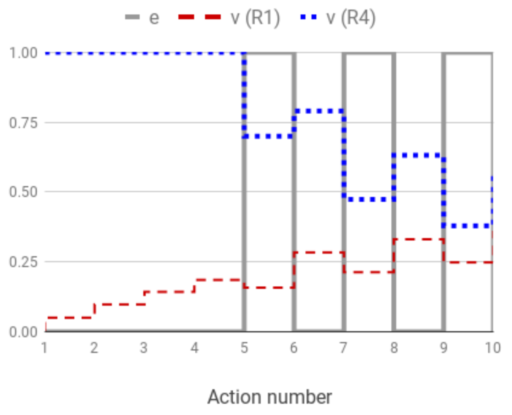


Figure 3: Two examples of the evolution of the performance value v in a procedure in which an R1 and an R4 follower perform the criterion *Error: action outside task* at actions 5, 7, and 9.

follower has the opportunity to make or not make another error (i.e., when it is time for the caregiver to begin a new action). The performance values v are only calculated when e changes. The overall ability and willingness values are updated when the individual v values change.

4.5 Adhering to Situational Leadership

Finally, there are two possible issues that should be avoided when using this model: (1) the skipping of readiness levels (followers that move directly between R1 and R3 as well as R2 and R4) and (2) followers moving between levels too quickly (Hersey et al., 1988; Bosse et al., 2017). To combat these issues, we implement a method of artificially lowering ability and lowering or raising willingness values as needed, which is similar to dynamic range compression (Kates, 2005). This allows for one domain's overall performance to wait while the other domain's performance catches up.

There are several of instances in which the ability or willingness value would need to be artificially changed:

- An R1 follower whose ability rises faster than willingness will need their ability value lowered when their willingness crosses the high-willingness threshold to ensure they move from R1 to R2 (see Figures 4 and 5c);
- An R3 follower whose ability drops below the high-ability threshold but whose willingness is still low will need their ability value lowered and may need their willingness value raised to ensure they move from R3 to R2 (see Figures 5j, 5k, 5m, and 5o);
- An R2 follower whose willingness drops below the high-willingness threshold but whose ability is high will need their ability value lowered to ensure they move from R2 to R1;
- An R4 follower whose ability or willingness drops below the R4 thresholds of 0.95 will need the other domain's value lowered to either 0.8525 (ability) or 0.625 (willingness) in order to ensure that the follower moves from R4 to R3 and from R3 to R2 smoothly (see Figure 5m.

Because our context is a medical procedure in which a patient's health is at risk, when in doubt, we always assume that a follower's ability is lower than it might be (Hjortdahl et al., 2009). For this reason, followers' ability values are only artificially lowered. As explained in section 6, the agent's communication content does not vary much between levels R1 and R2 or R3 and R4, and so willingness values can be lowered or raised when necessary.

As an example of how values might be artificially changed, examine an R1 follower whose ability and willingness increase at different rates. Before examining what this looks like, we can refer to Figure 2 and see that this would mean that the follower would move directly from R1 to R3. Because a follower can only progress from one readiness level to the next, we artificially lower the overall ability value in order to wait for the willingness performance to catch up. See Figure 4 for a demonstration of how this works.

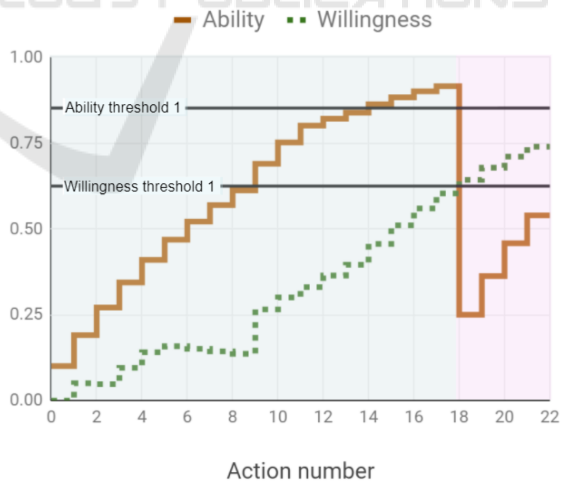


Figure 4: An example of how ability values are artificially lowered in order to ensure the follower progresses from one level to the next. Green represents readiness level R1 while pink represents readiness level R2.

In Figure 4, the follower passes the threshold for high ability at action 14. At this point, however, the willingness value is still low, so the follower is still considered to be R1. The procedure continues, and at action 18, the follower's willingness passes the threshold for high willingness. Since their willingness performance has increased above the threshold, the follower can now be considered to be in R2. Readiness level R2 is defined by low to some ability and high willingness, and so the follower's ability value is artificially decreased to 0.25, a value we chose in order to allow the system to "remember" the follower's performance history somewhat. In Figure 4, we can see that for actions 18-22, the follower is clearly in R2. In this way, the follower can progress until their ability once again passes the ability threshold. If, at that point, willingness has remained high, then the follower will progress to R3.

By manually lowering the ability and willingness values as needed, the system acts similarly to dynamic range compression in which signals are limited once they reach a certain threshold (Kates, 2005). Unlike compression, we lower the values in order to allow the follower to progress themselves through to the next readiness level. This is how we ensure that not only do followers progress from one readiness level to the next without skipping levels, but we also ensure that a follower does not move between readiness levels too quickly.

In the following section, we demonstrate how readiness level evolves in a variety of scenarios.

5 SIMULATIONS OF FOLLOWER PROGRESSION

In order to visualize how readiness level might change during a medical procedure and evaluate those changes to ensure they are indicative of the correct readiness level in practicality, we establish four starting states (R1, R2, R3, and R4 followers) and four possible progressions: (1) ability increases or decreases, willingness stays the same; (2) ability stays the same, willingness increases or decreases; (3) ability increases or decreases, willingness increases or decreases; and (4) ability remains the same, willingness remains the same. This results in sixteen different scenarios.

As discussed in section 4.3, the criteria in each scenario are created with help from the medical professional on our team and are based on real-world errors that followers in each readiness level would make. Therefore, creating these scenarios is also a method of testing our model and the values for the parameters described in section 4.3. The simulations included in this section are not exhaustive; many more were tested in order to ensure that the values for the parameters in our model returned accurate readiness levels.

The scenarios use a sample of a procedure which comprises a total of 22 actions. The first 9 actions have the least difficulty, actions 10-12 have the most difficulty, actions 13-14 have the least difficulty, and actions 15-22 have medium difficulty.

Figures 5a-5p display the 16 scenarios and are color-coded according to Figure 2: green represents R1, pink represents R2, blue represents R3, and yellow represents R4. Note that unless a follower is making every possible error constantly throughout the procedure, which often isn't possible, ability and willingness will increase from 0.

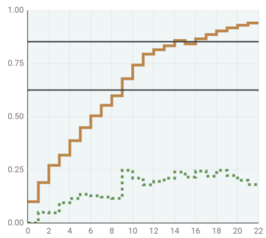
6 AGENT FRAMEWORK

Figure 6 describes the flow of information and how interaction is possible in our agent framework based on Mascaret. The agent platform contains the means to calculate readiness level and the information regarding the nominal procedure. The nominal procedure contains all information regarding resources, roles, procedure steps, and also allows for changes to be made to the procedure.

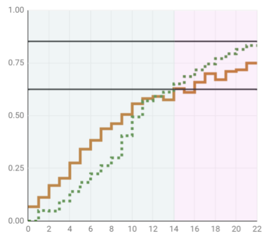
When a caregiver makes an error in Table 1, a criteria event is created which describes the error. The agent platform calculates readiness level from the human caregiver's speech and behavior, and then readiness level and criteria events stemming from caregiver behavior are created and work together to create communication intentions. The modality for agent communication is both verbal and non-verbal behavior, so communication intentions (verbal behavior) are paired with appropriate non-verbal behavior in the behavior planner (Collins Jackson et al., 2020).

The Mascaret framework permits the modeling of semantic, structural, geometric, and topological properties of the entities in the virtual environment and their behaviors. Mascaret also defines the notion of a virtual agent by their behaviors, their communications, and their organisation. Essentially, it is a framework in which an embodied virtual human can interact with a user.

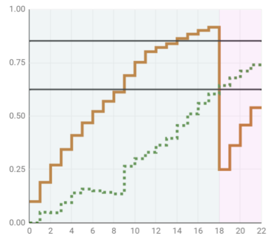
In Mascaret, human activities can be described in the virtual environment by using predefined collaborative scenarios (called procedures) which represent plans of actions for agents or instructions provided to users for assisting them. With Mascaret, a whole medical procedure with roles, resources, and trajectories can be formalized and virtually executed to assist a caregiver to go through this same procedure in the real world.



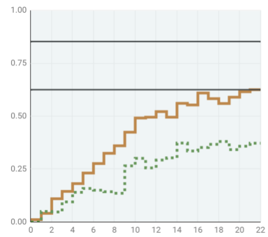
(a) R1: Ability rises, willingness remains low.



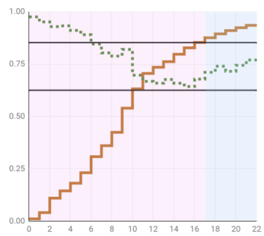
(b) R1: Willingness rises above the threshold.



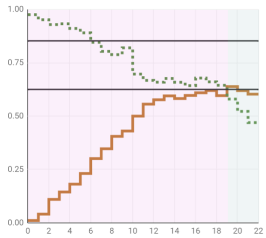
(c) R1: Ability and willingness rise above the thresholds.



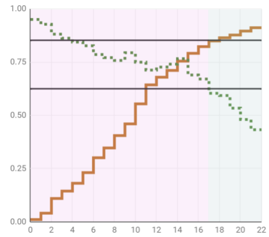
(d) R1: Ability and willingness remain low.



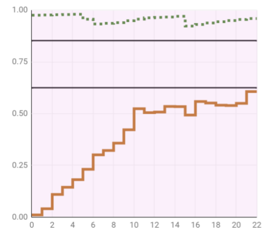
(e) R2: Ability rises above the threshold, willingness remains high.



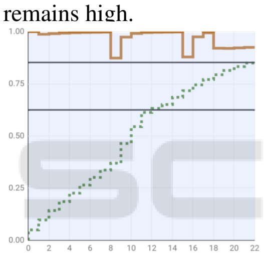
(f) R2: Ability remains low, willingness drops below the threshold.



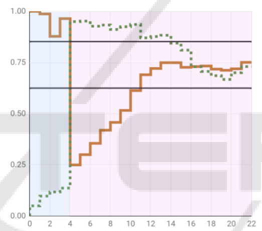
(g) R2: ability rises above the threshold, willingness drops below the threshold.



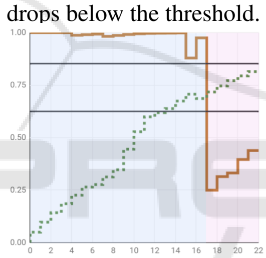
(h) R2: Ability remains low, willingness remains high.



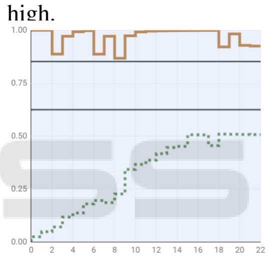
(i) R3: Ability remains high, willingness rises above the threshold.



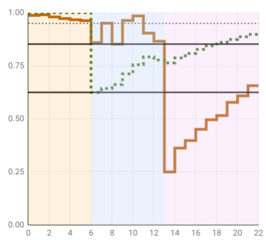
(j) R3: Ability drops below the threshold, willingness remains low (but is artificially raised).



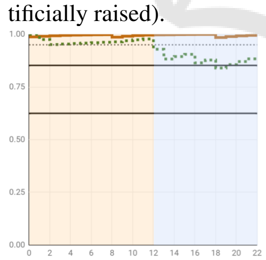
(k) R3: Ability drops below the threshold, willingness rises above the threshold.



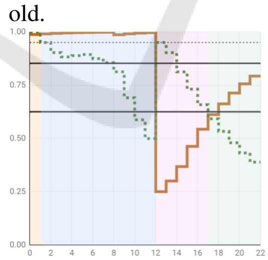
(l) R3: Ability remains high, willingness remains low.



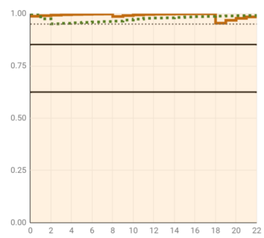
(m) R4: Ability drops below the threshold, willingness remains high (but is artificially lowered).



(n) R4: Ability remains high, willingness drops below the threshold.



(o) R4: Ability and willingness drop below the thresholds.



(p) R4: Ability and willingness remain high.

Figure 5: Sixteen simulations demonstrate how readiness level evolves when ability and willingness increases, decreases, or remains the same, four starting in each readiness level. The solid brown line indicates the overall ability performance, and the dotted green line indicates the overall willingness performance. Green shading refers to R1, pink shading refers to R2, blue shading refers to R3, and yellow shading refers to R4.

To allow Mascaret to interact with the user in a natural way, we introduce an embodied conversational agent (ECA) based on the SAIBA framework (Vilhjálmsson et al., 2007) and which has an intent planner which generates the communicative intentions of the agent, a behavior planner which translates communicative intentions into verbal and nonverbal signals, and a behavior realizer which trans-

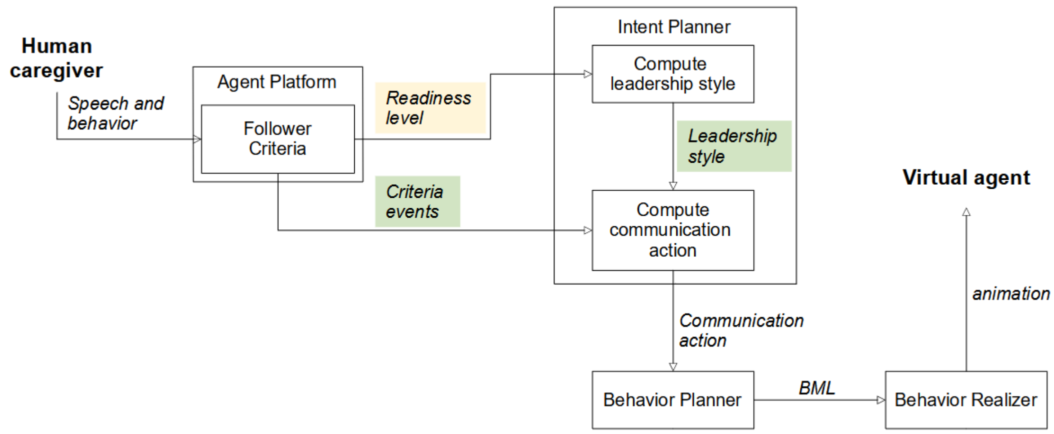


Figure 6: The general flow of information within our agent framework. Readiness level (highlighted in yellow) informs leadership style. Criteria event and leadership style (highlighted in green) inform the communication actions.

forms these signals in animation.

In Mascaret, we introduce the concept of follower criteria (see Table 1). After each action is performed by the follower, each behavior criterion is evaluated to see if it has been triggered (e.g., if the person has tried to do an action that is outside the task, if they have asked a question because they are stuck, etc.). According to the triggered criteria, the readiness level is computed (as explained in section 4).

The readiness level is used to compute the leadership style which is then stored in the intent planner (Collins Jackson et al., 2021)). The intent planner then generates a communicative intention each time a new action in the procedure must be done or when a follower criterion that needs attention has been triggered (for example, the virtual agent must inform the caregiver if they have made a mistake, must answer a question that the caregiver has asked, etc.).

To determine the ECA's communicative intention, we implement a set of rules which dictates how and what the agent will communicate according to its leadership style to be used in conjunction with linguistic rules specified in previous work (Collins Jackson et al., 2022):

Directing and coaching leadership (corresponding to readiness levels 1 and 2):

- Agent communicates every action to do;
- Agent communicates every time there is a criteria event.

Supporting leadership (corresponding to readiness level 3):

- Agent only communicates when criteria events are generated by criteria 1-7 and 9 have occurred (see Table 1 for criteria numbers).

Delegating leadership (corresponding to readiness level 4):

- Agent only communicates when 1, 2, 7, or 9 have occurred.

Thus readiness level and by extension, leadership style, criteria events, and procedure actions inform the communication actions that are created in the intent planner.

Note that unstructured dialogue between the caregiver and the agent is not possible. If the caregiver asks a question, the agent is able to respond, but the agent's speech is limited to only the actions and resources within the procedure.

To ensure additional safety, the caregiver can decline the agent's help at any time. When the caregiver feels competent without the agent's assistance, they can decline the agent's guidance during the procedure as a whole. In these situations, the agent will only act as a conduit of communication from the team of medical experts.

The agent itself is built in Unity and assumes the appearance of a doctor. In Figure 7, the agent is shown next to a dummy patient in augmented reality (AR) which we use for testing. However, our framework also allows for the agent to be displayed on a computer screen or on a tablet. When testing the system in AR, all follower behavior criteria can be monitored in the virtual system. Questions can be asked via the interface as well and are limited to requests for help regarding actions, resources, or roles and requests for clarification regarding trajectory.

Through both the virtual and augmented reality environments, remote medical experts can visualize the patient, the caregiver, and the procedure steps done thus far in 3D physical space and make decisions regarding procedure changes accordingly. In augmented reality, the caregiver can interact with the augmented reality environment by selecting which actions have been completed (as shown in Figure 7). For more information regarding the virtual environment set-up, please refer to (Querrec et al., 2018).

Figure 7: Our agent built with Unity pictured in augmented reality.

7 CONCLUSIONS AND FUTURE WORK

In this paper, we have described a novel model and the architecture of a novel system in which situational leadership is applied in a virtual agent system. Simulations of our model demonstrate how follower readiness level can be calculated for use in an agent framework that generates personalized agent behavior to each unique human user. Using our model, a virtual agent can lead a novice or expert caregiver through a medical procedure. This interaction is possible in our SAIBA-compliant agent framework using Mascaret, where interaction is completed with communication actions actuated by both caregiver behavior and information from the formalized procedure.

Our algorithms for determining readiness level and leadership style can be adapted to other circumstances outside the medical sphere as well when it is appropriate for an agent to lead a human being through any kind of task.

We aim to evaluate this system to determine whether an agent employing situational leadership is more effective (preserves the patient's health, is more efficient, and reduces stress of the caregiver) than an agent using the same leadership style throughout the procedure regardless of readiness level. Measuring trust and engagement from participants might come from non-anatomical behavior such as speed of accepting tasks. We plan to also measure trust and engagement with the agent with survey questions both before and after the procedure.

We also plan to further explore the co-occurrence of criteria to understand better how criteria that happen concurrently affect overall performance and whether these interaction effects should be taken into account when calculating readiness level.

ACKNOWLEDGEMENTS

This work has been carried out within the French project VR-MARS which is funded by the National Agency for Research (ANR).

REFERENCES

Araszewski, D., Bolzan, M. B., Montezeli, J. H., and Peres, A. M. (2014). The exercising of leadership in the view of emergency room nurses. *Cogitare Enferm*, 19(1):40–8.

Bedford, C. and Gehlert, K. M. (2013). Situational supervision: Applying situational leadership to clinical supervision. *The Clinical Supervisor*, 32(1):56–69.

Bernard, C. (2020). *Analyse de l'activité de soignants médicaux et paramédicaux sur simulateur haute-fidélité lors de simulations d'urgence, en vue de la conception d'un environnement de soins ergonomique*. PhD thesis, University of Lorient.

Bickmore, T., Asadi, R., Ehyaei, A., Fell, H., Henault, L., Intille, S., Quintiliani, L., Shamekhi, A., Trinh, H., Waite, K., Shanahan, C., and Paasche-Orlow, M. K. (2015). Context-Awareness in a Persistent Hospital Companion Agent. In *Proceedings of the 15th International Conference on Intelligent Virtual Agents*, IVA, pages 332–342, Delft, The Netherlands. ACM.

Bosse, T., Duell, R., Memon, Z. A., Treur, J., and van der Wal, N. (2017). Computational model-based design of leadership support based on situational leadership theory. *SIMULATION: Transactions of The Society for Modeling and Simulation International*, 93(7).

Castellano, G., Paiva, A., Kappas, A., Aylett, R., Hastie, H., Barendregt, W., Nabais, F., and Bull, S. (2013). Towards empathic virtual and robotic tutors. In *Proceedings of the 16th International Conference on Artificial Intelligence in Education*, AIED, pages 733–736, Memphis, TN, USA. Springer-Verlag.

Cisneros, F., Marín, V. I., Orellana, M. L., Peré, N., and Zambrano, D. (2019). Design process of an intelligent tutor to support researchers in training. In *Proceedings of EdMedia and Innovate Learning*, pages 1079–1084, Amsterdam, Netherlands. AACE.

Collins Jackson, A., Bevacqua, E., De Loor, P., and Querrec, R. (2019). Modelling an embodied conversational agent for remote and isolated caregivers on leadership styles. pages 256–259. IVA '19.

Collins Jackson, A., Bevacqua, E., De Loor, P., and Querrec, R. (2021). A computational interaction model for a virtual medical assistant using situational leadership. WI-IAT, Essendon, VIC, Australia. ACM.

Collins Jackson, A., Bevacqua, E., DeLoor, P., and Querrec, R. (2020). A taxonomy of behavior for a medical coordinator by utilizing leadership styles. pages 532–543. International Conference on Human Behaviour and Scientific Analysis.

Collins Jackson, A., Glémarec, Y., Bevacqua, E., De Loor, P., and Querrec, R. (2022). Speech perception and implementation in a virtual medical assistant. ICAART.

Dermouche, S. and Pelachaud, C. (2019). Engagement modeling in dyadic interaction. In *Proceedings of the 21st International Conference on Multimodal Interaction*, ICMI, page 440–445, Suzhou, Jiangsu, China. ACM.

Flin, R., Patey, R. E., Glavin, R., and Maran, N. (2010). Anaesthetists' non-technical skills. *BJA: British Journal of Anaesthesia*, 105(1):38–44.

Henrickson, P., Flin, R., McKinley, A., and Yule, S. (2013). The surgeons' leadership inventory (sli): a taxonomy and rating system for surgeons' intraoperative leadership skills. *BMJ Simulation and Technology Enhanced Learning*, 205(6):745–751.

Hersey, P., Blanchard, K. H., and Johnson, D. E. (1988). *Management of Organizational Behavior: Leading Human Resources*, chapter Situational Leadership, pages 169–201. Prentice-Hall, 5 edition.

Hjortdahl, M., Ringen, A. H., Naess, A.-C., and Wisborg, T. (2009). Leadership is the essential non-technical skill in the trauma team - results of a qualitative study. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, 17(48).

Hoegen, R., Aneja, D., McDuff, D., and Czerwinski, M. (2019). An end-to-end conversational style matching agent. https://arxiv.org/pdf/1904.02760.pdf.

Kates, J. M. (2005). Principles of digital dynamic-range compression. *Trends in Amplification*, 9:45 – 76.

Kulms, P. and Kopp, S. (2016). Test. In *Proceedings of the 16th International Conference, Intelligent Virtual Agents*, IVA, pages 75–84, Los Angeles, CA, USA. ACM.

Lee, S. K., Kavya, P., and Lasser, S. C. (2021). Social interactions and relationships with an intelligent virtual agent. *International Journal of Human-Computer Studies*, 150.

Lourdeaux, D., Afoutni, Z., Ferrer, M.-H., Sabouret, N., Demulier, V., Martin, J.-C., Bolot, L., Boccara, V., and Lelong, R. (2019). Victeams: A virtual environment to train medical team leaders to interact with virtual subordinates. page 241–243. IVA '19.

Montenegro, C., Zorrilla, A. L., Olaso, J. M., Santana, R., Justo, R., Lozano, J. A., and Torres, M. I. (2019). A dialogue-act taxonomy for a virtual coach designed to improve the life of elderly. *Multimodal Technologies and Interaction*, 3(52).

Nakhal, B. (2017). *Generation of communicative intentions for virtual agents in an intelligent virtual environment : application to virtual learning environment*. PhD thesis.

Pecune, F., Cafaro, A., Ochs, M., and Pelachaud, C. (2010). Evaluating social attitudes of a virtual tutor. In *Proceedings of the 16th International Conference on Intelligent Virtual Agents*, IVA, pages 245–255, Los Angeles, CA, USA. ACM.

Philip, P., Dupuy, L., Auriacombe, M., Serre, F., de Sevin, E., Sauteraud, A., and Franchi, J.-A. M. (2020). Trust and acceptance of a virtual psychiatric interview between embodied conversational agents and outpatients. *NPJ Digital Medicine*, (1).

Querrec, R., Taoum, J., Nakhal, B., and Bevacqua, E. (2018). Model for verbal interaction between an embodied tutor and a learner in virtual environments. pages 197–202.

Saerbeck, M., Schut, T., Bartneck, C., and Janse, M. D. (2010). Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI, page 1613–1622, Atlanta, GA, USA. ACM.

Sims, H. P., Faraj, S., and Yun, S. (2009). When should a leader be directive or empowering? how to develop your own situational theory of leadership. *Business Horizons*, 52(2):149–158.

Taoum, J., Raison, A., Bevacqua, E., and Querrec, R. (2018). An adaptive tutor to promote learners' skills acquisition during procedural learning.

Vilhjálmsson, H., Cantelmo, N., Cassell, J., Chafai, N., Kipp, M., Kopp, S., Mancini, M., Marsella, S., Marshall, A., Pelachaud, C., Ruttkay, Z., Thórisson, K., Welbergen, H., and Werf, R. (2007). The behavior markup language: Recent developments and challenges. volume 4722, pages 99–111.

Yang, P.-J. and Fu, W.-T. (2016). Mindbot: A social-based medical virtual assistant. In *Proceedings of the 2016 International Conference on Healthcare Informatics*, ICHI, pages 319–319, Chicago, IL, USA. IEEE.

Yule, S., Flin, R., Maran, N., Rowley, D., Youngson, G., and Paterson-Brown, S. (2008). Surgeons' non-technical skills in the operating room: Reliability testing of the notss behavior rating system. *World journal of surgery*, 32:548–56.