User Model User-Adap Inter (2011) 21:441–484
DOI 10.1007/s11257-010-9091-3

ORIGINAL PAPER

# A user modeling approach for reasoning about interaction sensitive to bother and its application to hospital decision scenarios

**Robin Cohen · Hyunggu Jung · Michael W. Fleming · Michael Y. K. Cheng**

Received: 3 December 2009 / Accepted in revised form: 22 November 2010 /
Published online: 4 January 2011


**Abstract** In this paper, we present a framework for interacting with users that is sensitive to the cost of bother and then focus on its application to decision making in hospital emergency room scenarios. We begin with a model designed for reasoning about interaction in a single-agent single-user setting and then expand to the environment of multiagent systems. In this setting, agents consider both whether to ask other agents to perform decision making and at the same time whether to ask questions of these agents. With this fundamental research as a backdrop, we project the framework into the application of reasoning about which medical experts to interact with,

We are especially eager to be included in a special issue that is dedicated to the memory of Fiorella de Rosis and Alison Cawsey. De Rosis played a significant role in motivating co-author Fleming towards the completion of his PhD research, when presented at the UM01 conference. Cawsey provided valuable feedback to co-author Cohen, following her investigation of Cohen's model of argumentation, as part of her own PhD research. The co-authors of this paper thus have a connection to the two researchers being honoured with this special issue.

R. Cohen · H. Jung (✉) · M. Y. K. Cheng
Cheriton School of Computer Science, University of Waterloo,
200 University Ave. W., Waterloo, ON, N2L 3G1, Canada
e-mail: h3jung@uwaterloo.ca

R. Cohen
e-mail: rcohen@uwaterloo.ca

M. Y. K. Cheng
e-mail: mycheng@cs.uwaterloo.ca

H. Jung
Department of Computer Science, Stanford University, 353 Serra Mall, Stanford, CA 94305, USA

M. W. Fleming
Faculty of Computer Science, University of New Brunswick, Fredericton, NB, E3B 5A3, Canada
e-mail: mwf@unb.ca

sensitive to possible bother, during hospital decision scenarios, in order to deliver the best care for the patients that arrive. Due to the real-time nature of the application and the knowledge-intensive nature of the decisions, we propose new parameters to include in the reasoning about interaction and sketch their usefulness through a series of examples. We then include a set of experimental results confirming the value of our proposed approach for reasoning about interaction in hospital settings, through simulations of patient care in those environments. We conclude by pointing to future research to continue to extend the model for reasoning about interaction in multiagent environments for the setting of time-critical care in hospital settings.

**Keywords** Reasoning about interaction · Modeling bother · Multiagent systems · Hospital decision making · Coordination of teams of health professionals · Choosing medical experts

## 1 Introduction

A network of researchers in computer science, engineering, nursing and medicine is currently involved in a project aimed at providing effective decision making support in various healthcare contexts, including that of the hospital setting, in a project known as hSITE (Healthcare Support Through Information Technology Enhancements). The central aim of this project is to be able to employ the right person, at the right time, with the right information, for more effective healthcare.

Specific challenges arise in the emergency room setting, in particular.[1] In general, in hospital emergency scenarios (ER), a patient arrives and is seen by the ER triage nurse, who determines in what section of the ER the patient will go. The triage nurse has experience to know whether the patient needs to be in acute, sub-acute, fast-track section or resuscitation room. After this is determined, the patient goes to the respective section and is taken care of and assessed first again by an ER nurse, then a Nurse practitioner/ER resident/doctor. If a patient comes in with a condition that is obvious and needs a specialist, then the nurse would proceed to call the specific specialist right away e.g. a brain injury causes a call for a neurologist. If a condition is not obvious, further assessments with the nurse, nurse practitioner and ER doctor are required to determine which specialist service to call. If the patients brought in by the Emergency Medical Services are in a critical condition, they are immediately brought into the resuscitation room, and nurses assigned to that room as well as an ER doctor attend to them. From there it is determined if other specialists are needed. In some hospitals, an urgency level is determined for each patient in the ER and is recorded with the patient.

The human first clinical assistants (FCAs) thus must make the best decisions possible about which specialists to bring in to assist arriving patients. In this context, we are designing a multiagent reasoning system running with parameters that model the medical experts and the patient, in order to propose the best experts (i.e. to deliver the required care to the patient). Without consideration of the possible bother being

[1] This scenario was outlined for us by health professionals associated with the Strategic Research Network known as hSITE (Plant 2008).

incurred when experts are solicited (so merely focusing on who might have the best expertise for the problem at hand) what results is a significant bottleneck in the effective delivery of the care to the patient. Especially with patients in critical conditions, it becomes important to be making very effective decisions about who should be consulted. In addition, the parameters are constantly changing in this dynamic setting, and reasoning needs to be sensitive to this as well. Our work is in fact being viewed by hSITE as a critical element for this kind of decision making.

More generally, artificial intelligence systems are often faced with situations in which their ability to perform a task (or to collaborate with users on a task) could be improved by soliciting further information from potentially helpful users. However, such interactions have costs associated with them. A decision-theoretic approach for reasoning about whether or not to initiate information gathering dialogues with users is presented by Fleming and Cohen (2004), weighing the benefits and costs of interaction, in order to determine if it is best to continue to operate autonomously or to elicit further input from the user. One of the important costs to consider is the extent to which the user will be bothered, when interaction is initiated.

In a multiagent system setting, there may in fact be several agents each trying to engage the same user in interaction. Modeling the bother endured by a user now becomes a more challenging task, and in essence, requires effective coordination of the possible interaction being conducted by the various agents.

In this paper, we present a model for reasoning about the bother that may be incurred when agents decide to interact with users, while problem solving in a multiagent environment. Other approaches for designing adjustable autonomy multiagent systems focus on having agents reason about determining the entity that can best increase the utility of the problem solving for the system. We argue that the cost introduced when another entity is bothered must be included in the decision making about whether to adjust the autonomy of an agent. In addition, we offer a model for decision making in a multiagent setting that reasons not only about transfering the control for decisions to other agents but also about asking questions of users, in order to gain information of value to improve the decision to be made.

We then discuss the challenge of having multiple agents reasoning about bothering users simultaneously and propose a method of coordination that employs proxy agents representing each user. With this method in place, we proceed to describe an experiment with human users to explore the value of our model for reasoning about bother and presents results to show that our users experienced greater satisfaction overall when their decision making was mediated by our framework for considering bother costs, before interacting with users.

Our approach for modeling the cost of bothering users in multiagent settings is then used as the motivation for a system to advise medical professionals about which experts to approach to assist with patients who are arriving in critical condition, requiring care. In particular, we reason about which expert to request to handle each patient at hand, using the hybrid adjustable autonomy system that we have developed (i.e. reasoning about which entities to transfer control to). In hospital settings, humans are asking other humans to assist. As such, the multiagent algorithms are simply used in the background to allow the first response professionals to select the medical professionals who will be engaged in interaction.

Due to the critical nature of the decision making in hospital settings, we propose new parameters to include in our modeling of users: expertise level (for the medical professionals) and task criticality (for the patients). We also propose a streamlined version of the hybrid transfer of control methodology, resulting in decision making that is less time consuming. In addition, we propose that the strategies for decision making be regenerated to cope with the dynamic environment.

The value of considering bother when reasoning about interaction has in fact been underscored by other researchers in the field. One important reason for monitoring the extent to which users are being burdened by interaction is in order to ensure that these users will continue to be productive participants in the problem solving. In an experiment consisting of fifty participants, conducted by Bailey et al. (2001), the researchers measured the effects of interruption on a user's primary task performance, annoyance, and anxiety. The results of the study revealed that interruptions have a negative effect on both a user's primary task performance (in terms of slower completion times), and emotional state. Interrupted users showed a definite increase in annoyance compared to a set of users who were not disturbed. In addition, there was a demonstrably higher (detrimental) increase in anxiety suffered by users due to interruptions. As such the research by Bailey et al. (2001) serves to emphasize that systems should be sensitive to the effort that needs to be expended by users when reasoning about whether to engage them in interaction.

In the following sections we outline our particular approach to modeling users and reasoning about interaction with users, sensitive to bother. We move forward from this to an adjusted model that focuses more on addressing dynamically changing environments, then projecting the modeling directly into the hospital decision making setting.

## 2 Incorporating bother cost into reasoning about interaction

In this section, we provide insight into the question of how to incorporate bother cost considerations into the process of reasoning about interaction. We begin with a discussion of the single-user, single-decision case and then extend the scope to account for multi-user, multiagent systems.

### 2.1 A single-user, single-decision model

Fleming (2003) and Fleming and Cohen (2004) developed a domain-independent decision-theoretic model for an agent to reason about whether or not it should interact with a human user. The model is aimed at solving 'single decision' problems, defined as "from an initial state, the system decides about interacting with the user, then makes a decision about what action to perform and then takes that action to complete the task" (Fleming and Cohen 2004).

The general algorithm for a system to reason about whether or not it should ask a question is fairly intuitive, and proceeds as follows (as presented by Fleming and Cohen 2004):

1. Determine the expected benefits of interacting with the user. More specifically, determine by how much the system's performance on the task is expected to improve (if at all) after asking the user the question.
2. Determine the expected costs of the interaction.
3. Proceed with the interaction only if the benefits exceed the costs.

The computation of the *benefits* of interaction is simply $Benefits = EU_{ask} - EU_{\neg ask}$, where $EU_{ask}$ represents the expected utility of an agent's decision using information obtained from the user, while $EU_{\neg ask}$ represents the expected utility of an agent's decision made without any more information. Note that the expected utility denoted here does not incorporate the costs incurred, but rather refers only to the value of the decision.

The value of $EU_{\neg ask}$ is the expected utility of the action that the agent believes to be the most promising in the current state, given the information it has without asking the user any further questions and is calculated as follows:

$$EU_{\neg ask} = \max_{a \in Actions} EU(a) \tag{1}$$

For each possible action $a$, the expected utility calculation takes into account the fact that there may be uncertainty about the possible outcomes of the action. For any given action $a$, suppose there are several possible results, each denoted $res_i$, with probability $P(res_i)$ and utility $U(res_i)$. Then,

$$EU(a) = \sum_i P(res_i) \cdot U(res_i) \tag{2}$$

To compute the value of $EU_{ask}$, let $P_{UR}$ denote the probability that the user responds and let $EU_{UR}$ be the expected utility that the agent could achieve if it receives an answer from the user. If the user does not respond or says that he does not know the answer, the agent will choose the action it believes to be the best, with expected utility $EU_{\neg ask}$.

$$EU_{ask} = P_{UR} \cdot EU_{UR} + (1 - P_{UR}) \cdot EU_{\neg ask} \tag{3}$$

Here, $EU_{UR}$ is computed by considering all possible responses $r_j$ that the user could give and the expected utility of the action $a_j$ that the agent would choose, given each response $r_j$.

$$EU_{UR} = \sum_{r_j \in Resp} P(r_j) \cdot EU(a_j \mid r_j) \tag{4}$$

The computation of interaction *costs* is done through a linear model, where the total cost is a weighted sum of individual costs; i.e., $Costs = \sum_i w_i C_i$. Two costs considered in Fleming & Cohen's work are the cost of the time required to interact with the user, and the cost of bothering the user.[2] This research clearly outlines where a

[2] Fleming (2003) also discusses briefly the cost of carrying out certain queries, such as costs in fetching from databases.

model of bother cost can be introduced into the process of reasoning about interaction. Bother cost is in fact included as a key factor in determining whether or not an agent will interact with a human user. How to model bother cost is discussed in greater detail in Sect. 4.

### 2.2 A multi-user, multiagent model

Adjustable autonomy is the idea that we should be "dynamically adjusting the level of autonomy of an agent depending on the situation" (AAAI 1999). In a way, we are neither accepting nor rejecting agent autonomy outright, but instead, selecting the autonomy level that is most appropriate for the situation. For example, for complex, mission-critical tasks, we may want agents to defer the decision-making to a more capable human user, while for routine mundane tasks, we probably want agents to be autonomous, so as to avoid bothering the user. An agent-based adjustable autonomy multiagent system is one where each individual agent reasons about possibly transferring its decision-making control to another entity (agent or user), as part of its overall problem solving process (Maheswaran et al. 2003).

A prominent adjustable autonomy system in the literature is that of the Electric Elves (E-Elves) project (Scerri et al. 2002), where each individual agent develops a transfer-of-control (TOC) strategy, indicating which entities the agent should transfer decision-making control to, and at what times. A notable aspect of a TOC strategy is that typically there is a time limit for an agent to wait for an entity to respond, before the agent gives up, and transfers control to another entity. This prevents an agent from waiting indefinitely for a non-responding entity. For example, a TOC strategy could be $User_1(5)User_2(11)Agent$, representing the strategy where the agent will first transfer control to $User_1$, and if $User_1$ does not respond by time point 5, then the agent will transfer control to $User_2$, and if $User_2$ also does not respond in time, then the agent will take back control and decide autonomously. The entities to include in a transfer of control strategy and the times to specify when reasoning about whether to attempt a different transfer of control are determined by reasoning about the expected utility of the actions taken at various times, by various users and agents, based on real world experience.

As a first step, we extend the E-Elves model (Scerri et al. 2002) to allow each agent to also reason about initiating information gathering interaction with an agent before determining what to do next (e.g., transfer decision-making control, or decide autonomously).[3] In our work, we differentiate between the agent querying an entity for information (which we refer to as a partial transfer-of-control or PTOC), and the agent asking an entity to make a decision (which we refer to as a full transfer-of-control or FTOC). Both of these cases are considered to be interaction from the agent to the entity. Figure 1 shows an example that includes both PTOCs and FTOCs.

Reasoning about this interaction in fact requires an effective model of bother cost as well. The challenge is for each agent to determine its optimal TOC strategy, by

[3] An overview of this proposal is presented as a two-page summary in Cheng and Cohen (2005). This journal paper serves to provide a more extensive description of the proposed model.

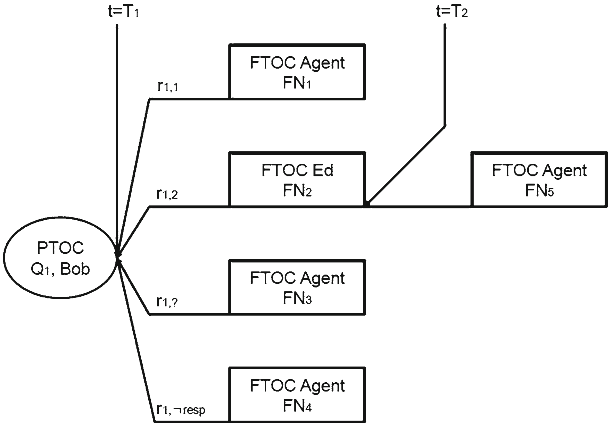


**Fig. 1** Example hybrid TOC strategy

generating possible strategies, evaluating the expected utility of the strategies and then selecting the one with the highest expected utility. The use of the term utility here is consistent with that used in E-Elves and reflects the difference between the benefits and costs.

The expected utility of a strategy $s$ is, in turn, dependent on the expected utility of all the leaf nodes in $s$.[4]

Formula-wise, $EU(s)$ is calculated as follows:

$$EU(s) = \sum_{LN_l} \left[ P(LN_l) \times (EQ(LN_l) - W(T_{LN_l}) - BC_{LN_l}) \right] \tag{5}$$

In the equation, $EQ(LN_l)$ denotes the expected quality of the agent's decision at leaf node $LN_l$, given the information it has gathered along the path to $LN_1$. $W(T_{LN_l})$ denotes the costs of waiting until the time of leaf node $LN_1$ to finish the interaction,[5] and $BC_{LN_l}$ denotes the bother cost accumulated from interacting with entities from all the transfers that the agent has done up to (and including) the current transfer of control under consideration.

The expected utility of the overall strategy is in effect the sum of the utility of each of the individual paths in it; thus, one needs to factor in the probability that the particular path will be taken $P(LN_1)$. This in turn will depend as well on the probability of response to any requests for interaction on that path. In Sect. 7, we extend this

[4] Note that, in Sect. 3.1, we considered only two strategies for the agent: it could ask the user a question (with expected utility $EU_{ask}$) or it could proceed with its reasoning without the user's help (with expected utility $EU_{\neg ask}$). The benefits of asking were calculated by computing $EU_{ask} - EU_{\neg ask}$, and these benefits were then compared to the costs of interaction. In our revised model in this section, we consider the expected utility of many possible interaction strategies, with both the benefits and costs incorporated into a single expected utility measure, $EU(s)$.

[5] Note that $W(t)$ is introduced in the E-Elves model as well and is intended to represent the cost of waiting in order to get the task completed, with respect to the need for the actions to be carried out quickly within the domain of application.

model for reasoning about interaction for dynamic, real-time environments and then provide additional clarification of the formulae that are used to perform the required calculations in Cheng and Cohen (2005).

Since our agents are operating in a multiagent system, the issue of coordination arises. In order to reason about coordinating the transfer-of-control strategies of individual agents, one critical issue is how each agent can best model the expected bother cost of the entity to whom control is being transferred. After all, other agents may be bothering the same entity so that agents cannot simply reason about their own proposed interactions with others.

In Cheng (2005) a framework is presented for reasoning about how to adjust the TOC strategies of agents to be sensitive to the bother initiated by other agents in the multiagent system. This includes four different algorithms for performing that coordination, along with experimental results to show the relative value of the different approaches. In this paper, we focus on one approach, referred to as the Broadcast method. We then use this particular method as the basis for the coordination performed in a study with human users, which is then used to validate the benefits of our particular model for reasoning about bother.

We begin first with background research on how best to model bother, leading up to our specific proposal for how each agent within the multiagent system can model the concept of bother. For the remainder of this paper, we focus entirely on the challenge of reasoning about human users and selecting users as the entities to engage when adjusting autonomy.

## 3 Background research on modeling bother cost

In this section, we give an overview of the related work on modeling bother conducted by various researchers in the field. In the following section, we develop our proposed bother cost model, motivated by the research discussed below.

### 3.1 Bauer et al.'s TrIAs system

In the work by Bauer et al. (2000) on trainable information assistants (TrIAs), annoyance to the user from too many queries is taken into account. In particular, users are modeled as getting more bothered with each successive query. As such, Bauer et al. model the bother cost as *annoy*(*#q, user*), where *#q* is the number of queries posed to the *user*. This cost is then subtracted from the value of the query to obtain the actual value of the query, i.e., $v'(q) := v(q) - annoy(\#q, user)$. Bauer et al. (as cited by Fleming 2003), goes one step further by modeling the *annoy* function based on the user's readiness/patience to interact with the system. In particular, they use a log-like penalty function for disruptions to very patient users, and an exponential function for impatient users.[6]

[6] The characterization of these penalty functions is not included in Bauer et al. (2000) but was discussed in private communication with M. Fleming.

This research motivates our use of a Bother Increasing Function as a user-specific parameter within our bother cost model. As will be shown in the following section, we are able to capture the kind of penalty function curves discussed above.

### 3.2 Fleming's bother cost model

There are two main principles to Fleming's bother cost model (Fleming 2003). First is the idea that "recent interruptions and difficult questions should carry more weight than interruptions in the distant past and very straightforward questions." Second is the notion that a user's willingness to interact with the system is a critical factor in bother cost modeling. Some users are very willing and would prefer to actively help the system achieve a better result, while other users are not willing, and would prefer not to be bothered much.

Fleming's model is as follows (Fleming 2003):

- First estimate how bothersome the dialogue has been so far. This *bother so far* (*BSF*) is given by $BSF = \sum_I c(I) \times \beta^{t(I)}$, where the system computes the sum over all the past interactions with the user (including the currently considered interaction). $c(I)$ is how bothersome the interaction was (e.g., cognitive effort required by the user to answer the question), $t(I)$ is the amount of time that has passed since that interaction, and $\beta$ is a discount factor that diminishes the effect of past interactions as time passes.
- Let $w$ represent the user willingness, with a range of 0 to 10, with higher $w$ meaning more willingness.
- Let $\alpha = 1.26 - 0.05w$ and $Init = 10 - w$. Here, $Init$ is to reflect the cost of bothering a user for the first time.
- Then, $BotherCost = Init + \frac{1-\alpha^{BSF}}{1-\alpha}$. From this formulation, a lower willingness $w$ results in a higher $Init$ cost, and also a higher $\alpha$ value (which amplifies the effect of the bother so far $BSF$). As $BSF$ increases, so too does $BotherCost$, but at different rates, depending on the $\alpha$ value. As shown by Fleming (2003), for low $w$ values, $\alpha$ will be greater than 1, and we will see an exponential-like increase due to $BSF$, while for high $w$ values, $\alpha$ will be less than 1, and we see a log-like increase. The values used for the calculation of $\alpha$ are in order to generate these kinds of curves for users with these characterizations of willingness.

Our proposed bother cost model retains the concept of a Bother So Far parameter, used to set the Bother Increasing Function to reflect the projected tolerance of the user. We also reflect on the extent that the user is willing to be bothered in a general User Unwillingness factor, which results in more bother to the user the greater the user is unwilling to interact.

### 3.3 Raskutti & Zukerman's RADAR system

Raskutti & Zukerman's work on RADAR (Raskutti and Zukerman 1997), a computerized information-providing system for travel planning, looked at two factors when determining which disambiguating query to issue: a *nuisance factor* and an *estimated* *queries factor*. The nuisance factor represents the annoyance caused to the user due to irrelevant questions. The estimated queries factor represents the number of queries that the system expects to ask as a consequence of asking the query currently under consideration. Like the other works described earlier, asking more queries is modeled as reducing the utility experienced by the user.

The suggestion that certain questions may be more bothersome to some users than others motivates our use of a Base Bother factor for each question in consideration, as part of the modeling of the bother cost to the user.

### 3.4 Attention and interruptability

One term that is related to bother is that of attention or interruptability. Vertegaal (2003) asserts that the growing demand on a user's attention is a critical usability issue, in other words that the user's attention is a limited resource. This has led to the study of a new type of user interface design, named Attentive User Interfaces (AUIs), that are sensitive to the user's attention.

The concept of attention is integrated into related research about reasoning about interaction with users. Horvitz et al.'s COORDINATE system (Horvitz et al. 2002), allows users to input assessments of how interruptible they are during different meetings using a range of low, medium, and high interruptability levels. In addition, users can associate a dollar-value cost for interrupting during a certain interruptability level, and a default cost for interrupting during a free period (i.e., not in a meeting).

In a later formulation, Horvitz and Apacible (2003) focus not so much on the interruptability level associated with meetings, but rather on the concept of *attention state*. They consider the utility $u(D_i, A_j)$ to represent the cost to a user in attention state $A_j$ when disrupted by event $D_i$. The formulation of the expected cost of interruption is then $ECI = \sum_j [p(A_j|E) \times u(D_i, A_j)]$ where $p(A_j|E)$ is the probability of the attentional state, conditioned on evidence stream $E$. Motivated by this research, we decide to incorporate a parameter called the Attention State factor into our proposed model of bother to users, which will increase the overall measure of bother when the user's attention is more engaged (thus making the user less interruptable).

Walji et al. (2004) report on several findings by other researchers on the impact of interruptions. One finding that supports our proposal to model the attention state factor is that different people have different abilities to respond to and manage interruptions. In experiments conducted by Bailey et al. (2001) on the effects of interruptions, they found that the degree of disruptiveness depends on the user's mental load at the time of the interruption. This provides additional support for the decision to include in our bother cost model a factor reflecting the extent to which the user is occupied, when interaction is initiated.

## 4 Proposed bother cost model

In our research, "bother factor" is meant to represent the degree to which a user would be annoyed, disrupted or inconvenienced by any interaction with the system.

The exact role of this particular factor will depend on the domain. For some domains, users will almost always enter the collaboration with the understanding that they will be expected to play a major role in the problem solving. In other systems, which will be of greater interest from the perspective of our model, the possibility exists for a wide range of participation levels for users. This latter case arises in applications that are not necessarily meant to be interactive, but where the system has been charged with the responsibility to perform some task and might occasionally benefit from obtaining more information from the user at certain points in its reasoning. This is the case, for instance, in adjustable autonomy multiagent systems (Hexmoor et al. 2003), in which agents can adapt the degree of autonomy they exhibit in different situations.

Most importantly, however, the role of the bother factor will depend on the individual user; we therefore advocate a user modeling approach to the construction of a bother cost model. While some users will prefer to be very actively involved with these systems, doing everything in their power to help the system achieve the best possible results, others will prefer not to be bothered and will be happy with the best solution the system is able to find on its own.

It should be pointed out, however, that even systems that are meant to be completely interactive must take care not to bother a user unnecessarily. Consider the idea of a system that is intended to help a user with planning a vacation. The number of questions that *might* be useful for a system to ask is quite large. However, if the system were to ask every one of these questions at the beginning of a planning session, even the most patient user would very quickly become overwhelmed and frustrated. Instead, it seems reasonable for a system to ask only those questions that are absolutely necessary at the outset. From that point on, decisions should be made about the value of asking a question. The expected cost of bothering the particular user involved should be an important factor in such decisions.

Below, we present a working model that incorporates the current bother cost research in the field. From previous works, we have extracted the following factors which are believed to influence bother cost:

- The difficulty of the interruption query, $TOC\_Base\_Bother\_Cost$. For example, usually, asking a user his/her preference is easier (i.e., cognitively less intense) than asking a user to decide on a plan of action.[7]
- The attention state of the user, $Attention\_State\_Factor$. For instance, a user is more interruptible when resting than when he/she is busy with important work.
- The user's unwillingness to interact with the system, $User\_Unwillingness\_Factor$. This is a measure of how receptive (or rather, unreceptive) the user is towards being TOC'ed, and how disrupted they are by interruptions.[8]

[7] Each question carries a base bother cost that is pre-set; for example, asking a user to make a weighty decision is set at the highest end, a less weighty decision is set somewhat lower, a somewhat challenging preference query is then lower still and a very simplistic preference query is set lowest.

[8] We found it more intuitive to model unwillingness, instead of willingness because (as will be shown shortly) a higher value in this term will result in a higher bother cost.

– The timings of the interruptions, $t(TOC)$, and the discount factor, $\beta$ $(0 < \beta < 1)$, which reduces the bother impact of past TOCs as time passes.[9]

By logically adapting Fleming's bother cost model (Fleming 2003) to incorporate the findings of other researchers, we propose the following enhanced bother cost model:

$$\begin{aligned} Init &= User\_Unwillingness\_Factor \\ &\quad \times Attention\_State\_Factor \times TOC\_Base\_Bother\_Cost \end{aligned} \tag{6}$$

$$\begin{aligned} &BSF(Bother\ So\ Far) \\ &= \sum_{toc \in PastTOC} TOC\_Base\_Bother\_Cost(toc) \times \beta^{t(toc)} \end{aligned} \tag{7}$$

where *PastTOC* is the set of all the past TOCs experienced by the user, $TOC\_Base\_Bother\_Cost(toc)$ is just the $TOC\_Base\_Bother\_Cost$ of $toc$, and $t(toc)$ is the time point at which $toc$ occurred.

To determine the increase to the bother cost due to $BSF$, we have a function, $BC\_Inc\_Fn(BSF, User\_Unwillingness)$, that maps a $BSF$ value to a bother cost increase, based on the user's unwillingness level.

$$BotherCost(BC) = Init + BC\_Inc\_Fn(BSF, User\_Unwillingness). \tag{8}$$

Here are some suggestions for possible bother cost factor values:[10]

– $[TOC\_Base\_Bother\_Cost]$ $Easy = 5$, $Medium = 10$, $Hard = 20$
– $[Attention\_State\_Factor]$ $Relaxed = 0.75$, $Neutral = 1$, $Busy = 1.25$
– $[User\_Unwillingness\_Factor]$ $Willing = 0.5$, $Neutral = 1$, $Unwilling = 2$
– $[\beta]$ 0.90
– $[BC\_Inc\_Fn]$ For Willing, $BC\_Inc\_Fn(x) = x^{0.75}$, for Neutral, $BC\_Inc\_Fn(x) = x^1$, for Unwilling, $BC\_Inc\_Fn(x) = x^{1.25}$. This gives us roughly the same bother cost shape as used by Fleming (2003) and Bauer et al. (2000). Figure 2 shows how the bother cost increases due to bother so far, for the different user willingness types.[11]

9 Note: The value of $\beta$ depends on the size of the time step. If a time step is 'huge', then $\beta$ should be low (to reflect that one time step means a lot of time has elapsed, and so we should discount more), while inversely, if the time step is 'small', then $\beta$ should be high. Also, it is conceivable that the value of $\beta$ will depend on the particular person. For this paper, we assume (for simplicity's sake) that it is the same value for all users.

10 Note: These are only suggestions. In the real world, the system designer would want to tailor the values to the domain. For instance, the domain might require finer granularity in terms of the number of attention states, or, perhaps the differences between willing and unwilling users are greater, necessitating greater differences in $BC\_Inc\_Fn(BSF, User\_Unwillingness)$. The specific values proposed here are to suggest possible scales used by the system designer, for ease of interpretation. Values could be normalized to one scale, to make interpretations more uniform.

11 Note: For $BSF$ less than 1, we might want to use different functions, else we get the somewhat odd result that there is more bother increase for Willing users than Unwilling users. However, with $BSF$ less than 1, the increase in bother cost is negligible and so this slight difference is virtually irrelevant.

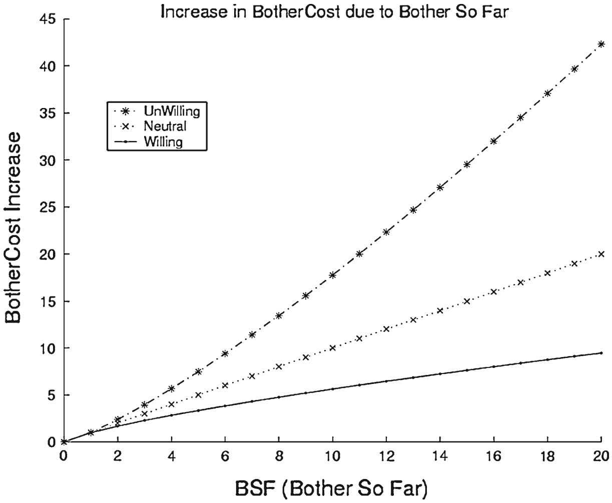


**Fig. 2** Graph showing how much bother cost increases due to bother so far, for the different user willingness types

### 4.1 Example bother cost calculation

Suppose a busy (*Attention_State_Factor* = 1.25) unwilling (*User_Unwillingness_Factor* = 2) user is faced with the following TOCs:

[$TOC_1$] An easy preference query ($TOC\_Base\_Bother\_Cost = 5$) at time 5
[$TOC_2$] An easy preference query ($TOC\_Base\_Bother\_Cost = 5$) at time 7
[$TOC_3$] A medium decision query ($TOC\_Base\_Bother\_Cost = 10$) at time 25

The bother cost (*BC*) for $TOC_1$ is: $Init = 2 \times 1.25 \times 5 = 12.5,\ BSF = 0,\ BC = 12.5 + 0^{1.25} = 12.5$

The bother cost (*BC*) for $TOC_2$ is: $Init = 2 \times 1.25 \times 5 = 12.5,\ BSF = 5 \times 0.9^{7-5} = 4.05,\ BC = 12.5 + 4.05^{1.25} = 18.25$

The bother cost (*BC*) for $TOC_3$ is: $Init = 2 \times 1.25 \times 10 = 25,\ BSF = 5 \times 0.9^{25-5} + 5 \times 0.9^{25-7} = 1.36,\ BC = 25 + 1.36^{1.25} = 26.47$

As a comparison, suppose the same TOCs are given to a user who is relaxed (*Attention_State_Factor* = 0.75) and willing (*User_Unwillingness_Factor* = 0.5). Then, the bother costs are: $BC(TOC_1) = 1.875,\ BC(TOC_2) = 4.73,\ BC(TOC_3) = 5.01$.

As can be seen, there is a significant difference in bother cost between the two different users. We feel that overall, users would benefit from having adjustable autonomy agents that are sensitive to their willingness type, and also their attention state. In general, we want to discourage TOCs to users when they are unwilling or when they are busy.

### 4.2 Discussion of the bother cost model

In developing the bother cost model proposed in this section, we made several design decisions. Instead of the concept of a *bother cost*, another way to incorporate bother into reasoning about interaction is the notion of an *annoyance threshold*, which determines the amount of acceptable bother allowed for an interaction (McCrickard et al. 2003). We designed our model to use cost instead of threshold for the reason that we want to explicitly weigh the benefits of the interaction against the costs. Even if an interaction is costly, it might still be beneficial to interact if the benefits are high (e.g., when an agent requires badly needed information in order to make an important decision).

While we integrated several ideas from Fleming's original model (Fleming 2003) we made some adaptations as well, as follows. First, we introduced a factor to reflect the attentional state of the user, motivated by researchers such as Bailey et al. (2001), Horvitz and Apacible (2003), and Chen and Vertegaal (2004). Second, we factor in the cognitive difficulty of the current query, $TOC\_Base\_Bother\_Cost$ into the $Init$ value, and have it scale with respect to the user's willingness and attention. This is in contrast with the original (Fleming 2003) model, where the current query difficulty gets bundled into $BSF$, and so has a much less direct effect on the final bother cost value. In addition, our model allows the system designer to decide how to map $BSF$ values to bother cost increases, by supplying their own $BC\_Inc\_Fn(BSF, User\_Unwillingness)$ functions.

Note that within our research, we do not focus on how to generate the questions that the agents will ask, and so do not incorporate the *nuisance factor* of asking irrelevant questions that appears in Raskutti & Zukerman's model (Raskutti and Zukerman 1997), making the simplification that all questions asked are relevant.

We do not elaborate on how to obtain values for the parameters in our model. There are, however, some valuable starting points for this research outlined by other researchers. Fleming (2003) suggests that the system initially asks the user a series of questions to determine the user's willingness, to let the willingness level of the user to be adjusted later. For example, if a user finds that he/she is being interrupted too much, then he/she can adjust the willingness level downwards. To determine $TOC\_Base\_Bother\_Cost$ values, one can look at the experiments conducted by McCrickard et al. (2003) and by Altmann and Trafton (2004), which measure the ill-effects of interruptions, such as the longer time to finish the primary task, and the number of errors incurred. Horvitz and Apacible (2003) comment that users are comfortable assessing the interruption cost in terms of a dollar value that they are willing to pay to avoid the disruption, and so this is another way to determine $TOC\_Base\_Bother\_Cost$ values. Fogarty et al. (2004a), Horvitz and Apacible (2003), and Chen and Vertegaal (2004) have done interesting work on how to assess the user's attention state from sensor readings. For the remainder of this paper, we will set aside the question of obtaining the initial values for the variables that model the user. We return to this topic briefly in Sect. 10, reflecting on directions for future research.

Overall, incorporating bother cost into the process of reasoning about adjusting autonomy enables each agent in a multiagent system to be sensitive to the user's state. For instance, implemented correctly, bother cost awareness prevents any one user from being incessantly bombarded by a never-ending series of TOC requests, as this increases the bother cost to a high level. Also, by being aware of the user's willingness and attention state, our agents are more likely to transfer control to willing and relaxed users, instead of to unwilling and busy ones. The motivation is to provide a system that will be attractive to users, for real world applications.

## 5 Coordinating bother cost in the process of reasoning about transfers of control in a multiagent system

In this section, we discuss the importance for agents to more accurately keep track of bother cost to users in the context of a cooperative adjustable autonomy multi-agent system. Each agent in the system must reason about whether to transfer decision-making control to another entity (agent or user) or to retain decision-making control itself, driven by considerations of maximizing the expected utility of the transfer-of-control strategy it selects for the problem solving. Note that deciding to transfer decision-making control to a user also requires selecting which user to engage and thus an estimate of the bother cost imposed on each user becomes relevant.

There are interesting challenges to modeling bother cost in a multiagent system. While an agent has up-to-date records regarding the bother cost of various users in the single agent case, this is no longer true in the multiagent systems case. From an individual agent's perspective, the bother so far (*BSF*) of users may change for reasons other than as a result of its own actions. In particular, $BSF$ of users may change due to the actions of other agents in the system. Unless mechanisms are in place to address this, agents will likely have 'stale' local data about users' bother so far.

The problem with stale bother data is that an agent may plan a TOC strategy which it believes is optimal, but is in fact, not optimal since the users involved in the TOC strategy may have already been bothered too much by other agents in the system. As such, we would like the agents to propagate bother cost information amongst themselves in order to keep each other updated.

The really interesting part about this problem in particular is that $BSF$ gradually decreases with time (assuming no new TOCs to the user). In other words, the passage of time will discount/reduce the effects of a particular TOC to a user. As such, how significant it is that an agent gets notified of another agent's TOC to a user depends on whether or not that agent will consider accessing that user in the near future. For example, suppose there are two agents in the system, $agent_1$ and $agent_2$, and two users, $Bob$ and $Ed$ (with $Bob$ being the better decision maker). Suppose then that $agent_2$ TOCs to $Bob$ at time $t = 2$. Whether or not $agent_1$ really needs to know about this TOC event depends on how soon or late $agent_1$ will consider transferring control to users. If it's soon (e.g., $t = 3$), then it would be good for $agent_1$ to know about $agent_2$'s TOC to $Bob$, so that $agent_1$ can explicitly reason about whether or not it is still worthwhile to TOC to $Bob$, or to just TOC to $Ed$ (who has not been bothered yet). On the other hand, if it's late (e.g., $t = 30$), then $agent_1$ does not need to know about $agent_2$'s TOC to $Bob$, because the bother effect of that particular TOC is probably no longer significant enough to affect $agent_1$'s reasoning about the best TOC strategy to employ.

### 5.1 Information sharing agents

As discussed in the previous section, miscoordination may occur due to agents acting on stale information. As such, part of the coordination solution would involve a method by which agents share information about bother cost. Possible approaches that are discussed in Cheng (2005) include: broadcasting, no information sharing, communicating to verify that bother has been modeled effectively and a relaxed version of the latter algorithm that tolerates being accurate only within a certain threshold. We present broadcasting in some detail and offer a brief description of the verify approach. We revisit the need to consider coordination more extensively when discussing future research on employing our bother cost model to direct decision making in hospital settings.

We make the following simplifications:

- We focus on sharing users' bother cost information amongst the agents, not other types of information such as domain specific information.
- Each user ($User_i$) has his/her own proxy agent ($Proxy_i$). This is a reasonable assumption and is similar in fact to what occurs in the E-Elves project Scerri et al. (2002) where each user has a personal assistant agent.[12] In order for an agent to access (i.e, TOC) a user, it has to go through that user's proxy agent first. To differentiate between the proxy agent and the agent which initiates the TOC, we shall denote the latter as 'requesting agent' or sometimes simply as 'agent', while the former will always be noted as 'proxy agent'.
- To focus on the interplay between expected decision quality ($EQ$), bother cost ($BC$), and communication cost, we have simplified the TOC model by assuming that the probability of response is 1 (i.e., the user to whom control has been transferred will always immediately respond). The TOC strategies considered in this section will be of length one, since one TOC is enough to arrive at a decision. Thus, the key term to consider for the expected utility (EU) of a strategy is $EQ - BC$.

## 5.2 Coordination by broadcast

This type of agent emphasizes up-to-date information, with the benefit of high utility ($EQ - BC$) achieved by the agents, at the cost of high communication overhead. There are two ways to do this: [*Push*] Whenever a user is bothered (i.e., control is transferred to that user), his/her proxy agent will broadcast this news to all agents in the society, and [*Pull*] Whenever an agent is about to enter its optimal TOC strategy reasoning process, it will first broadcast poll all the proxy agents for the current bother cost data of their users. For this paper, we will go with the push approach, since there is no time lag involved (i.e., agent can start planning right away, instead of waiting to receive all the poll information).[13]

When a Broadcast agent needs a decision made, the process is as follows:

1. Using its up-to-date bother cost information, the agent determines an optimal TOC strategy, which specifies transferring control to user $User_i$.
2. The agent sends a TOC request (which includes the TOC question to ask) to proxy agent $Proxy_i$ who will in turn, relay the TOC question to $User_i$.

[12] Note that in E-Elves, the personal assistant agent does not perform the coordination function that we envision our proxy agents doing.

[13] Note that the difference between push and pull strategies could potentially be significant, as the former depends on the number of agents in the system and the latter on the number of users.

3. $Proxy_i$ broadcasts an update/notification message to all agents in the system, to alert them of the TOC event.
4. When an agent receives a notification message, it will update the bother so far ($BSF$) value for $User_i$, so that future TOC strategy planning will be accurate.

### 5.3 Coordination by verify

Below we also provide further insights into the method that verifies bother cost values before proceeding, to offer additional appreciation for the role that proxy agents may play.

The Broadcast method is in fact used within our study with human users, outlined in the following subsection. But the proxy agent can in fact be playing an even greater coordinating role, towards improved calculations of the actual current bother cost to a user. This is fact proposed in the method we label as Verify (with results presented in the following section). In this arrangement, the proxy agent collects requests to bother the user but requires the estimated bother cost of the user to be sent along as part of the message. The proxy verifies whether this calculation from the sending agent is accurate (with respect to the current bother being experienced by the user, which it knows since all requests go through it) and if not, denies the request to contact the user and asks the sender to redo calculations, in order to determine whether this user is in fact still the best one to contact; the current bother so far to the user is conveyed to the sending agent. That agent will then recalculate its optimal TOC strategy. This method carries with it an additional cost (the cost of retrying) but has as a saving fewer requests for assistance simply being turned down because the user who was asked is considerably more busy than was anticipated.

We have conducted extensive experiments (Cheng 2005) to demonstrate the relative value of differing methods for coordinating multiple agents that are calculating bother costs of others and possibly issuing requests, at once. In the following subsection we present a small subset of these results, aimed at simply confirming that the coordination methods provide valuable benefit beyond methods which do not attempt to be careful with their calculations.

An example to demonstrate how the values of the variables determine the agents that are bothered and the interaction that ensues is outlined in Appendix A. Here, we sketch not only how the bother so far and bother costs influence the decisions about interaction but also illustrate more specifically how the Coordinate by Broadcast method ultimately determines which agent is asked, as multiple requests may ensue.

### 5.4 Some empirical results

In this section, we present some empirical results to demonstrate the robustness of our proposed framework for modeling bother costs when reasoning about interacting with users in multiagent settings.

To begin, the method for coordinating the bother cost modeling outlined in Sect. 5.2 is compared to three other models for achieving coordination, described briefly below.

**Table 1** Average utility of methods for coordinating bother cost modeling

| | Experiment: willing users | | | |
|---|---|---|---|---|
| | Broadcast | Private | Verify | Threshold |
| Average utility | 89.25 | 81.62 | 89.25 | 86.85 |
| | Experiment: unwilling users | | | |
| | Broadcast | Private | Verify | Threshold |
| Average utility | 71.82 | 6.06 | 71.82 | 70.45 |

The first is called Private and involves having each agent reasoning about which entity to ask based on private knowledge alone. These "optimists" assume that no one else is trying to bother the same users. The second is called Verify (corresponding to Coordination by Verify as explained in Sect. 5.3) and has an agent check with the proxy whether its estimate for bother is accurate, receiving a yes/no response that may then alter the decision. The third is called Threshold and it operates like the second but with tolerance to be within a threshold on its bother cost estimates, when verifying. The latter two agents serve to filter requests that are forwarded to the agent, relying on the relevant modeling of user bother.

Table 1 below shows the utility averaged over several decision events, using $EQ - BC$ for the utility (where a decision event is an attempt to transfer control to an entity). The experimental setup was that of a simulation which had 50 users and 50 agents, 5 decision classes and for each decision class a randomly generated number from the range [50,100] assigned to serve as that user's $EQ$ value for that decision class. The bother cost parameters set $\beta$ to be 0.9, $Attention\_State\_Factor$ to be 1, $TOC\_Base\_Bother\_Cost$ to be 10. There are 100 timesteps taken per trial and for each timestep and agent there is a 0.05 chance that a decision event will occur for the agent at that timestep.

To determine the value of our model with willing and unwilling users both, we ran two experiments. In the first the users were all willing and in the second the users were all unwilling. The results are as presented in Table 1. The Coordinate by Broadcast bother algorithm is shown to perform well with respect to the utility that is calculated. In addition, as one expected utility decreased with willingness levels. This method is also clearly important compared to the Private algorithm which does not attempt to coordinate. This algorithm suffers when users are unwilling. Note that when the overall utility of decisions is high, this demonstrates that effective choices were made in determining the entities to address the decisions at hand; the tasks are being handled appropriately.

In a third experiment, we varied the value of $\beta$ in order to confirm the robustness of the bother cost modeling. The results are graphed in Fig. 3. As $\beta$ increases, the rate at which bother so far is discounted with respect to time is decreased. As such, the higher $\beta$ values impose higher demands. Still, the average utility values as plotted show our model "holding its own" over the range of $\beta$ values.

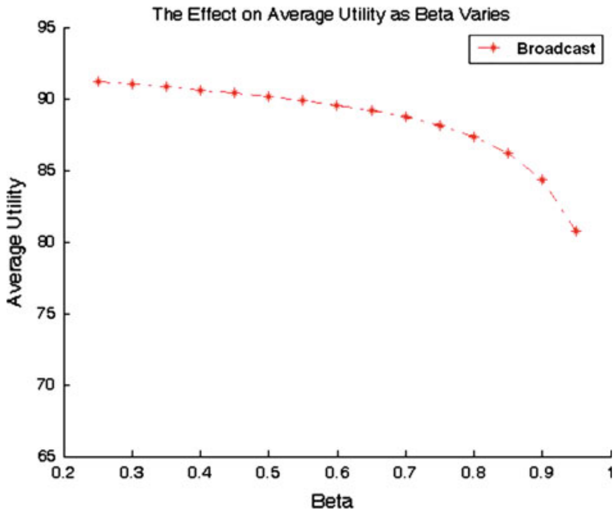


**Fig. 3** The effect on average utility as *beta* varies

On the whole, these experiments demonstrate that our proposals for coordinating the bother cost modeling result in effective decision making (and offer a clear improvement over methods which rely entirely on private information).

## 6 A streamlined model for interaction in dynamic, time critical scenarios

In this section, we introduce a model that can be used specifically for scenarios where an agent is reasoning about which human users to enlist to perform decision making, in an environment where decisions need to be made under critical time constraints and where the parameters that serve to model the human users are changing dynamically, to a significant extent. We offer a hybrid transfer of control strategy, similar to the one presented in Sect. 2.2, but with a restricted set of possible questions and a limit on the transfer of control chain.

As in Sect. 2.2, transfer of control strategies are generated in order for the optimal strategy to be selected for execution. Similar to the approach in Sect. 2.2, when there is no response from the user who is approached to respond, after waiting the indicated period of time, the next user is considered and so on, down the chain within the transfer of control strategy. Different from the approach in Sect. 2.2, attempts at full transfers of control in fact are framed as PTOCs with the question Q: "Can you take over the decision making?".[14] This then enables both a "yes"s response, which results in an FTOC[15] to this user or a "no" response (or silence).

[14] This then introduces a common base bother cost for the interaction, since this is the same question asked each time.

[15] As a simplication, we assume that a "Yes" response results in the user successfully assuming control of the decision.

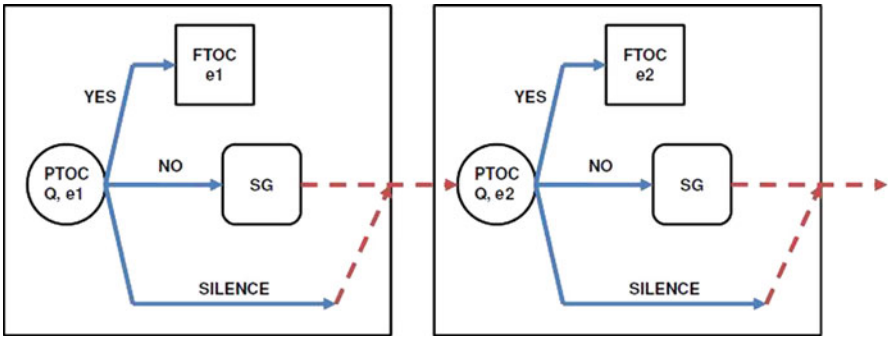


**Fig. 4** Visual representation of strategy with the FTOCs and PTOCs; each world occupies one square

The "no" response is where we introduce a new method aimed at coping with the dynamics of the environment and the criticality of time. Instead of proceeding to ask other potential users who may have been identified as suitable to possibly help, the transfer of control attempt ends and instead there is a new generation of possible strategies. The purpose of requiring this regeneration is in order to reason at this new point in time regarding the users who are available to help and their expected quality of decision, cost of bother, etc. This approach is particularly valuable in order to overcome the challenge of reasoning with stale information about the possible users, in circumstances where choices that are less than optimal can be undesirable, to a dramatic extent. As a result, the quick responses from users who are not available to assist can be used to recompute the best possible users to enlist.

The approach that is followed when there is silence projects continued attempts to contact other users. At the end of this chain of attempts, we inject a final decision of strategy regeneration.[16] Strategy regeneration will then allow for an updating of parameter values. Note that in our current model, we make the simplification that the strategies do not ask different entities within the same chain. This is because we are limiting ourself to only one question, that of asking the expert to help.

A diagram outlining the FTOCs and the PTOCs that we envisage is presented in Fig. 4 where an arrow with a solid line means the stream of time, but a dotted line means there is no break by the end of the arrow. In addition, we introduce a concept of *world* to facilitate the computation of the utility of any given strategy. One *world* consists of one PTOC, one FTOC, and one SG node and includes all the parameters currently used to calculate benefits and costs to reason about interaction with entities. Therefore, when the current *world* is moved to the next step, our system asks a new entity. The number of worlds is equivalent to the number of entities that will be asked.

The formulae that would be used to reason about the expected utility derived from a strategy are explained below. As outlined in Sect. 2.2, the optimal strategy is determined by evaluating the expected utility ($EU$) of each of the generated strategies and selecting the one with the highest $EU$ value. As explained in Cheng (2005), a

[16] This is in contrast to the general approach provided in Sect. 2.2, where the final node in a chain is usually one where there is a full transfer of control back to the agent, who must then perform the decision.

strategy generation phase would begin with the simplest strategies (of length one) and then expand to longer strategies by adding an FTOC or a PTOC node to previously generated strategies. The strategy generation is then limited by bounding the maximal length of strategy. For the model of reasoning presented in this section, we limit the strategy generation based on the number of entities under consideration. If there are $k$ entities, $k!$ strategies are generated, among which we choose the one with the highest $EU$ value. Then, the overall $EU$ of strategy $s$ is computed by taking the sum of the $EU$ of all the leaf nodes in $s$.

In Cheng (2005), there are two types of nodes: an FTOC node, $fn$ and a PTOC node, $pn$. In the PTOC node, a question Q is asked an entity $e_i$. In addition, we introduce a SG node, $sg$, where a new strategy is generated in our model.

Considering the calculation for the expected utility of a strategy as the sum of the utilities of the leaf nodes in that strategy we then proceed to calculate separately the utility of (a) ending in a full transfer of control (b) ending in a strategy regeneration from a "no" response (c) going down a path of "silence"[17] to a final stage of strategy regeneration. Note as well that here the probability that a transfer of control is occurring is dependent on the probability that all the PTOC nodes prior to this one are silence and on the probability of the response associated with this node ("yes", "no" or silence).

Below are the equations to calculate $EU$ for the three cases having different a leaf node: "yes" , "no", and silence. If the response is "yes", the leaf node of the path from the initial PTOC node is an FTOC node. The $EU$ of an FTOC node $(fn_l)$ in the $j$th $world$ is computed as follows:

$$EU_j(fn_l) = \prod_{pn_{prev}} P_{e_{prev}}^{\{resp=Silence\}} \times P_{e_i}^{\{resp=Yes\}} \times \left(EQ_{e_i}^d - W(t_e - t_s) - BC_{fn_l}\right) \quad (9)$$

where $P_{e_i}^{\{resp=r\}}$ represents the probability that asking an entity $e_i$ the query will result in a particular response $r$; $BC_{fn_l}$ is the accumulated bother cost to entities resulting from all the transfers that agent has done up to the current transfer of control under consideration. Note that $t_e$ is the ending time of the FTOC (so where the arrow meets the FTOC square in Fig. 4), while $t_s$ is the starting time of the FTOC (so where the arrow heading into the FTOC square originates).

If the response is "no", the leaf node of the path from the initial PTOC node is an SG node. The $EU$ for each SG node in the $j$th world is calculated as follows:

$$EU_j(sg) = \prod_{pn_{prev}} P_{e_{prev}}^{\{resp=Silence\}} \times P_{e_i}^{\{resp=No\}} \times \left(EQ_{e_i}^d - W(t_e - t_s) - BC_{sg} - SGC\right) \quad (10)$$

[17] If nobody has been found to answer either "yes" or "no", we define this as a case of silence.

where $BC_{sg}$ is the accumulated bother cost to entities resulting from all the transfers that the agent has done up to the current transfer of control under consideration, and $SGC$ denotes the cost of generating a new strategy.

In case of a silence response, we put a virtual node $dfl$ ("default"), into the final world. The $EU$ of the virtual node $(dfl)$[18] in the final ($n$th) world is computed as follows:

$$EU_n(dfl) = \prod_{pn_{prev}} P_{e_{prev}}^{\{resp=Silence\}} \times P_{e_i}^{\{resp=Silence\}} \times \left(EQ_{e_i}^{d} - W(t_e - t_s) - BC_{sg} - SGC\right) \tag{11}$$

There are $n$ FTOC nodes, $n$ PTOC nodes, and one virtual node in the overall framework with $n$ worlds. We obtain the overall $EU$ of strategy $s$ by summing up $n$ $EU$ values for FTOC nodes, $n$ $EU$ values for SG nodes, and one $EU$ value for the virtual node as follows:

$$EU(s) = EU_n(dfl) + \sum_{j=1}^{n} \left(EU_j(fn_l) + EU_j(sg)\right) \tag{12}$$

where $n$ represents the number of worlds.

## 7 Addressing selection of experts for medical decision making

In this section we apply the streamlined hybrid transfer of control model from the previous section directly to the challenge of finding the right person, at the right time, to assist with the care of patients who are arriving at a hospital. We are considering the summary of real-life hospital scenarios outlined in Sect. 1, which emphasizes the value of modeling both quality of decision and cost of bothering, when interacting with experts. These are the scenarios that may benefit from the use of our model for reasoning about interaction in multiagent systems.

### 7.1 Determining the parameters

In fact, for a system to reason about which expert to interact with, in an effort to resolve patient care in emergency scenarios, it is important to be modeling the various medical experts and their areas of expertise. It may also be reasonable to be reasoning about the criticality of the task to be addressed. In these scenarios, the task is simply attending to a patient that has just arrived.

As explained earlier, in order for the first clinical assistants to make the best decisions about which experts to bring in, the proposal is to have our multiagent reasoning system running with parameters that model the medical experts and the patient. For the purpose of having these experts come to perform the decision making (i.e. to deliver

[18] The leaf node for the silence response is set to $sg$.

the required care to the patient). These experts are then the entities $\{e_1, e_2, \ldots, e_n\}$ that are considered in our reasoning about interaction.

We propose the addition of one new parameter as part of the user modeling for the bother cost, a *Lack-of-expertise factor*. This parameter is used to help to record the general level of expertise of each doctor (i.e. medical specialist), with respect to the kind of medical problem that the patient is exhibiting. Since the aim is to locate those who are specialists in the area of the patient's current problem, this factor can be used to greatly prefer those doctors who do indeed have strong expertise as required.

In order to obtain values for the various parameters in the bother cost calculation, we propose a simplification of the formulae outlined in Sect. 4 that relies on inferring the values of certain parameters solely on the basis of other parameter values. In this way, we limit the number of parameter values that need to be solicited, learned and updated, as part of the processing. In our Future Work section (Sect. 10.1), we provide further discussion on how to make the calculations more complex, with greater effort to learn various parameter values, over time.

As will be outlined below, we assume that a user's willingness is simply determined by their attentional state factor and their expertise level. We also assume that a user's probability of response is determined by their attentional state factor.

Some factors which affect bother cost in hospital settings are thus as follows.

- The difficulty of the query, $TOC\_Base\_Bother\_Cost$. In hospital settings with a streamlined model, this factor is fixed, since we are considering only one question to ask (whether the user can assist with the patient).
- $Attention\_State\_Factor$ reflects how busy the doctor (medical expert) is. A doctor currently without a patient would have a low attentional state value; a doctor currently attending to a patient would have a high attentional state value.
- The lack of expertise of the doctor, $Lack\_of\_Expertise\_Factor$. The expertise will then affect the unwillingness of the doctor. That is, as this factor increases, $User\_Unwillingness\_Factor$ increases.
- The user's unwillingness to interact with the system, *User_Unwillingness_Factor*. This is a measure of how receptive (or rather, unreceptive) the doctor is towards being TOC'ed, and how disrupted they are by interruptions. We currently present a simplification of this calculation. This factor is obtained by adding $Attention\_State\_Factor$ to $Lack\_of\_Expertise\_Factor$.
- The timings of the interruptions, $t(TOC)$, and the discount factor, $\beta$ $(0 < \beta < 1)$, which reduces the bother impact of past TOCs as time passes. We choose a relatively high $\beta$ because hospital settings are under time-critical situations where the time step is 'small'.

With the inclusion of these new parameters, we then propose adjusted formulae for modeling the bother to users, as follows:

$$User\_Unwillingness\_Factor = Attention\_State\_Factor + Lack\_of\_Expertise\_Factor \quad (13)$$

$$Init = User\_Unwillingness\_Factor \times Attention\_State\_Factor \times TOC\_Base\_Bother\_Cost \quad (14)$$

$$BSF(Bother\ So\ Far) = \sum_{toc \in PastTOC} TOC\_Base\_Bother\_Cost(toc) \times \beta^{t(toc)} \tag{15}$$

$$BotherCost(BC) = Init + BC\_Inc\_Fn(BSF, User\_Unwillingness) \tag{16}$$

Here are some suggestions for possible bother cost factor values:

- $[TOC\_Base\_Bother\_Cost]$ $Easy = 5,\ Medium = 10,\ Hard = 20$
- $[Attention\_State\_Factor]$ $Relaxed = 0.75,\ Neutral = 1,\ Busy = 1.25$
- [*Lack_of_Expertise_Factor*] $High$ (i.e., not very expert) $= 0.25,\ Medium = 0$, $Low$(i.e., very expert) $= -0.25$
- $[\beta]$ 0.90
- [*BC_Inc_Fn*] For Willing, $BC\_Inc\_Fn(x) = x^{0.75}$, for Neutral, $BC\_Inc\_Fn(x) = x^1$, for Unwilling, $BC\_Inc\_Fn(x) = x^{1.25}$.

Note that the $User\_Unwillingness\_Factor$ ends up producing the value of 2 when the user is Busy and the $Lack\_of\_Expertise\_Factor$ is High and a value of 0.5 when the user is Relaxed and the $Lack\_of\_Expertise\_Factor$ is Low. These values correspond to those suggested for Unwilling and Willing in Sect. 4.

We introduce another new parameter, *task criticality* (*TC*), to affect the reasoning about interaction. *TC* is used to enable the expected quality of a decision to be weighted more heavily in the overall calculation of expected utility, when the case at hand is very critical. This parameter may also be adjusted, dynamically. In general, *TC* represents that how critical the task is. In hospital settings, if the task criticality of a patient is higher than that of typical patients, it implies that the state of the patient having higher *TC* is more serious than other patients. When a patient has high task criticality, strong expertise is required because the patient may become much more serious if not treated intensively.

There are two characteristics of task criticality. First, the *TC* of a patient who is not treated increases as time passes. We consider different increasing rates for each *TC* level: high-level, medium-level, and low-level. In other words, high-level *TC* will increase faster than low-level *TC* as time passes. Second, the expected quality of a decision is weighted by the *TC* level and the $Lack\_of\_Expertise\_Factor$ as presented in Eq. 17. Table 2 shows weights for each case, which will be applied in example scenarios in Sect. 7.2.

$$EQ^d_{e_i} \rightarrow EQ^d_{e_i} + \left(Weight \times EQ^d_{e_i}\right) \tag{17}$$

If the *TC* of some patient is low, the patient does not have to consider the expertise of a doctor carefully. Thus, the expertise does not affect determining the expected quality of a decision. However, the *TC* of some patient is high, the patient should consider the expertise of a doctor seriously. Therefore, the expertise will affect determining the expected quality of a decision. In this case, when the *TC* is high, the expected quality of a decision needs to be adjusted with more weight. Also, when the *TC* is low, the expected quality of a decision needs to be adjusted with less weight.

**Table 2** Weights to determine the expected quality of a decision

| | Case 1 | Case 2 | Case 3 | Case 4 | Case 5 | Case 6 |
|---|---|---|---|---|---|---|
| Task criticality | High | High | Medium | Medium | Low | Low |
| Lack-of-expertise factor | Low | High | Low | High | Low | High |
| Weight | 10% | −10% | 5% | −5% | 0 % | 0% |

**Table 3** Level and increasing rate by the score of task criticality

| | [0, 10) | [10, 80) | [80, ∞) |
|---|---|---|---|
| Level of task criticality | Low | Medium | High |
| Increasing rate | 2% | 5% | 10% |

## 7.2 Example scenarios for reasoning about medical experts

In order to illustrate how effective choices are made to enable the coordination of medical professionals and the resolution of the decision making regarding patient care, we introduce some examples.

In our examples, we consider the case where there are four possible medical experts to approach. Each has differing expected quality of decision making, differing attentional state (e.g. attending to other patients currently or not), different inherent willingness to assist.

The model parameters used in our scenarios are as follows:

- $[TOC\_Base\_Bother\_Cost]$ 15
- $[Time\, discount\ factor\ \beta]$ 0.90
- $[initial\ EQ]$ 150
- $[Cost\, of\ Waiting,\ W(t)]\ t^{1.5}$
- $[the\ Number\ of\ Worlds\ (n)]$ 4
- $[SGC]$ 0 cost

Note that we assume that the cost caused by regenerating a strategy is simply zero.[19] We also assume that the expected quality of decision for all specialists begins with the same initial base value (which is then adjusted according to the user's expertise level as discussed earlier). The $TOC\_Base\_Bother\_Cost$ is set "somewhat high" (using the range of values listed in Sect. 4). This is because our one question has the purpose of getting a user to agree to carry out a decision.

As shown in Table 3, there are levels of task criticality and increasing rates for each level of task criticality. We assume that the patient's criticality will be assessed by the first clinical assistant attending to the patient (and it may be periodically updated, as the patient continues to be unattended).

[19] We discuss how this may be investigated more extensively as future work, in Sect. 10.1.

**Table 4** Probability of entity response to question $Q$ by attentional state factor

| Attentional state factor | Yes | No | Silence |
|---|---|---|---|
| Relaxed | 0.45 | 0.45 | 0.1 |
| Busy | 0.1 | 0.6 | 0.3 |

**Table 5** Elapsed time

| Attentional state factor | Lack-of-expertise factor | Yes | No | Silence |
|---|---|---|---|---|
| Relaxed | Low | 2 | 4 | 5 |
| Relaxed | High | 3 | 3 | 5 |
| Busy | Low | 3 | 3 | 5 |
| Busy | High | 4 | 2 | 5 |

**Table 6** Profiles of entities at that time when a patient arrived

| | $e_1$ | $e_2$ | $e_3$ | $e_4$ |
|---|---|---|---|---|
| Attentional state factor | Relaxed | Busy | Busy | Relaxed |
| Lack-of-expertise factor | Low | High | Low | High |
| Probability of response for yes | 0.45 | 0.1 | 0.1 | 0.45 |
| Probability of response for no | 0.45 | 0.6 | 0.6 | 0.45 |
| Probability of response for silence | 0.1 | 0.3 | 0.3 | 0.1 |
| Elapsed time for yes | 2 | 4 | 3 | 3 |
| Elapsed time for no | 4 | 2 | 3 | 3 |
| Elapsed time for silence | 5 | 5 | 5 | 5 |

Probability of response depends on the attentional state factor of each doctor. Table 4 represents a probability of response for "Yes", "No", and Silence, which includes two cases: relaxed and busy with respect to the attentional state factor.

The time by which a response to a question will be generated from the doctors will be referred to as the elapsed time. For simplification in our following examples, we have this elapsed time determined by the attentional state factor and the expertise level of the doctor, according to predefined values as provided in Table 5. The units of time are left unspecified. Note that we assume a fixed elapsed time for all the cases of "silence".

In our scenarios, we will divide task criticality into three levels: high, medium, and low level. Given the level of task criticality, we determine the expected quality of a decision by adding a weight represented in Table 2. The expected quality of a decision is dynamically changed by the change of the value of task criticality as time progresses. Table 6 shows profiles of available doctors in a hospital currently, for our sample scenarios below.

### *7.2.1 Scenario 1*

A patient has just arrived at the emergency room who is assessed as highly critical. The FCA tries to search for the right doctor for the current patient with the decision-support system. In this hospital, as in Table 6, there are four doctors, $e_1$, $e_2$, $e_3$, and $e_4$. Our system checks the profile of each doctor and begins finding out the optimal strategy by calculating an expected utility for each generated strategy. Since there are four doctors, we obtain 4! strategies. By evaluating each strategy, we obtain 24 expected utility values for each strategy. The greatest expected utility is $EU(s) = 198.39$ whose strategy chain is $e_1$-$e_3$-$e_4$-$e_2$.

The strategies that choose to ask $e_4$ first do not have high $EU$ values, even though the expert is not Busy and can attend to the patient. This is because the High Criticality of the patient has raised the weight of the $EQ$ value in the calculation. The maximal $EU$ of a strategy that asks $e_4$ first = 49.87. Likewise, strategies that select $e_2$ first have very low $EU$ values, as this expert is both Busy and with High $Lack\_of\_Expertise$. The values for the $EU$ of all the possible strategies are presented in Appendix B.

### *7.2.2 Scenario 2*

As in Scenario 1, a patient has just arrived at the emergency room, but this time has been assessed at medium criticality. Considering the same four doctors $e_1$, $e_2$, $e_3$ and $e_4$, and the expected utility values for each strategy outlined in Appendix B, the greatest expected utility is $EU(s) = 163.03$ whose strategy chain is $e_1$-$e_3$-$e_4$-$e_2$.

### *7.2.3 Scenario 3*

In this case the patient who has just arrived at the emergency room is assessed at low criticality (but still in need of specialized assistance). Considering the same four doctors once more, the greatest expected utility is $EU(s) = 128.20$ whose strategy chain is $e_1$-$e_4$-$e_3$-$e_2$.

## 7.3 Coordination

The model presented above for decision making in hospital scenarios is a first step in building on the model for bother cost and the model for reasoning about interaction with users, described at the beginning of this paper.

We recognize that effective decision making in hospital scenarios will also require a component to enable the hospital staff to be reasoning about handling multiple patients, with multiple experts, simultaneously. As a result, the challenge of coordinating the reasoning about interaction will need to be addressed. We leave for future work the development of an appropriate coordination mechanism that is also suitably attuned to the characteristics of dynamic, task critical decision making.

# **8 Simulations to demonstrate value of approach**

In this section, we offer three separate simulations to show the value of our proposed approach for reasoning about interaction with medical experts, sensitive to the cost of bother. These experiments show: (i) our proposed model compared to one where the cost of bother has not been modeled; (ii) our proposed model compared to one where

**Table 7** Calculation of user unwillingness factor

| Lack of expertise factor | Attention state factor | | |
|---|---|---|---|
| | Relaxed | Neutral | Busy |
| Low | 0.5 | 0.75 | 1 |
| Medium | 0.75 | 1 | 1.25 |
| High | 1 | 1.25 | 1.5 |

**Table 8** Probability of entity response to question $Q$ by user unwillingness factor

| User unwillingness factor | Yes (%) | No (%) | Silence (%) |
|---|---|---|---|
| $Willing$ | 60 | 20 | 20 |
| $MedWilling$ | 50 | 30 | 20 |
| $Neutral$ | 40 | 40 | 20 |
| $MedUnwilling$ | 30 | 50 | 20 |
| $Unwilling$ | 20 | 60 | 20 |

**Table 9** Probability of entity response to question (how quickly)

| User unwillingness factor | 1 Time unit (%) | 2 Time unit (%) | 3 Time unit (%) | 4 Time unit (%) |
|---|---|---|---|---|
| $Willing$ | 33 | 27 | 13 | 7 |
| $MedWilling$ | 27 | 23 | 17 | 13 |
| $Neutral$ | 20 | 20 | 20 | 20 |
| $MedUnwilling$ | 13 | 17 | 23 | 27 |
| $Unwilling$ | 7 | 13 | 27 | 33 |

strategy regeneration is not introduced and (iii) our proposed model compared to one where the task criticality does not result in a different weight of expected quality, compared to bother cost, in the calculation of expected utility.

In order to run these simulations, we made a few refinements to the model presented in Sect. 7. We allowed for an intermediate level for lack of expertise, called medium, which results in no changes to the relative weighting of expected quality compared to bother cost in Eq. 17. We also ultimately allowed for varying level of user unwillingness, reflecting the combination of attention state factor and lack of expertise factor values. This is displayed in Table 7. We also specified the probability of response at a finer grain of detail, through two new tables (Tables 8, 9).

We begin with an overview of the scenario and its parameters, followed by the simulation results, for each of the experiments.

## 8.1 Overview of the simulation parameters

Our validation measures performance of our model reflecting dynamic and time critical aspects. Our simulation used Matlab (R2010a) on a machine with the following

**Table 10** Profiles of patients for our simulation

| No. | Patient | Medical problem | Task criticality |
|---|---|---|---|
| 1 | $p_1$ | *Cardio* | 70 |
| 2 | $p_2$ | *Cardio* | 90 |
| 3 | $p_3$ | *Neuro* | 63 |
| 4 | $p_4$ | *Cardio* | 82 |
| 5 | $p_5$ | *Neuro* | 70 |

**Table 11** Profiles of entities for our simulation

| Entity | ASF | Specialized area | Number of patients |
|---|---|---|---|
| $e_1$ | *Relaxed* | *Cardio* | 7 |
| $e_2$ | *Relaxed* | *Cardio* | 100 |
| $e_3$ | *Relaxed* | *Neuro* | 15 |
| $e_4$ | *Relaxed* | *Neuro* | 120 |
| $e_5*$ | *Relaxed* | *Neuro* | 240 |
| $e_6*$ | *Relaxed* | *Cardio* | 98 |

settings: AMD athlon(tm) 64 X2 Dual, Core Processor 5600+, 2.91 GHz, and 3.25 GB of RAM. In the setting of our validation simulating hospital emergency scenarios, there are four entities on the entity list and five patients on the waiting list. Profiles of patients and entities are shown in Tables 10 and 11. Note that entities $e_5$ and $e_6$ on the entity profiles (see *) were not included during the simulation with four entities case.

Every patient has a task criticality for his/her specific medical problem, and the task criticality of each patient is changed dynamically as time progresses.

Our simulation considers all the patients in the emergency room, beginning with the most critical patient first and then sequentially processing the remaining patients on the waiting list, always processing the most critical patient first.

The first clinical assistant picks up the most serious patient from the waiting list. Then, the waiting list is updated by eliminating the selected patient which has been assessed as the most critical patient. Thus, the number of patients remaining becomes four.

Strategies are generated by following the process introduced in Sects. 6 and 7. In our validation, 4! strategies are generated since there are four entities available in our scenario. We evaluate the expected utility of each strategy and choose the one with the optimal utility.

We set values of parameters of each entity based on the profile of the current patient. We already know some information of each entity such as attentional state factor, specialized area, and the number of patients the entity has treated. Then, we set the following parameters: lack of expertise factor, probability of response for answer, and probability of response for how quickly the entity will respond, $bc\_inc\_fac$, and $Init$. Those parameter values affect the bother cost of each entity.

### 8.2 Simulations

We simulated a particular hospital scenario, described below and for each experiment, ran our simulation 100 times. In each case, we measured the number of problem patients that arose. These were patients whose task criticality rose to beyond 100. Such unattended patients could in a worst case simply die (a large problem) or at least would generate a greater cost for the hospital to bear (a less critical, but nevertheless significant problem). As will be shown, in each case our proposed model results in fewer problem patients, overall.

The entity will answer "Yes", "No" or "Silence" based on probability of response as shown in Table 8. With the current user unwillingness factor, we get information about when the entity will answer the question and which answer will be given. For example, if the user unwillingness factor is Willing, the probability to answer "Yes" is 60%. We use a uniform distribution. In other words, our simulation generates a random number between 0 and 10. If the random number is between 0 and 6, our simulation considers that the answer is "Yes". If the random number is between 6 and 8, the answer is considered as "No". Otherwise, we consider that the response from the entity is silence.

After obtaining the type of answer, we use another uniform distribution to simulate when the entity responded to the question. If the answer is silence, the response time becomes 5 units. Otherwise, we generate a random number between 0 and 80 and see which number is generated. If the entity is a $Willing$ person, the entity would give a response which may be either "Yes" or "No" in 1 unit time with the probability of 33% as shown in Table 9. Thus, if the random number generated from the uniform distribution is between 0 and 33, we consider that the entity responded in 1 unit time after being asked. Also, the entity has a probability of 27% to respond in 2 units, 13% in 3 unit time, and 7% in 4 unit time. Therefore, if the random number drawn from the uniform distribution between 0 and 80[20] is between 33 and 60, 2 units is given as the entity response time. If it is between 60 and 73, 3 units is given. Otherwise, 4 units is given to the entity as response time to the question.

The three sets of results are explored below.

### 8.2.1 *Time cost and bother cost*

This validation measures performance of our model reflecting dynamic and time critical aspects by comparing it with one that is missing the calculation of bother cost.

Every patient has a task criticality for his/her specific medical problem, and the task criticality of each patient is changed dynamically as time progresses. Our simulation first selects the patient whose task criticality is highest.

We obtain a strategy chain by calculating formulae (Eqs. 12, 16, 17) reflecting our model based on the patient's profile (medical problem and criticality). After choosing a entity in the chain, we ask him/her to treat the current patient and update the criticality of patients who have been treated by entities, as well as those remaining on the waiting list. If a patient has not been attended to (i.e. no doctor has replied "yes"), the task criticality of the patient increases as time passes. If the task criticality of the patient is increased over 100, we model this as a problem patient. When there are no

[20] We modeled that the probability of response for the case of silence as 20% for any type of user unwillingness factor. In other words, the probability of response for the case of "Yes" and "No" is always 80%. Thus, we arranged the range of the distribution to be between 0 and 80 for convenience in our calculation.

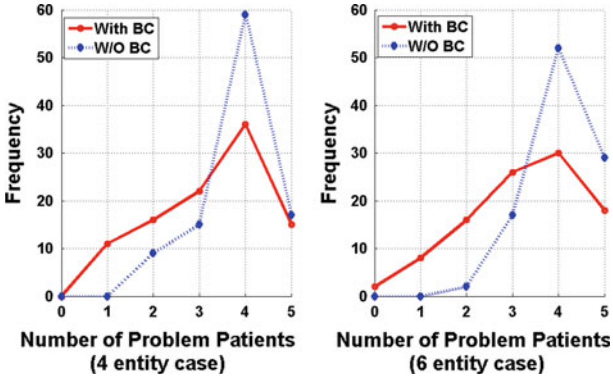


**Fig. 5** Our model with and without bother cost

more patients on the waiting list, we finally count the number of problem patients. By comparing the number of problem patients simulated by our model with bother cost and without bother cost, we can validate whether our model reflects dynamic and time critical domains effectively.

Figure 5 illustrates the distribution generated by our model with Bother Cost and without Bother Cost. The graph on the left represents the case of four entities and five patients and one on the right represents the case of six entities and five patients. The $x$-axis of each graph denotes the number of problem patients, and the $y$-axis the frequency of each value on $x$-axis after running our simulation 100 times. The solid line represents the version including Bother Cost, and the dotted line represents the version excluding Bother Cost. In Fig. 5, we can find the peak of the dotted line located in a higher position than the peak of the line at 4 on the $x$-axis and inclined to the right. This implies that there have been more problem patients during simulations with the version without Bother Cost (dotted line) than one with Bother Cost (solid line). In other words, the version calculating Bother Cost outperforms the one which does not calculate Bother Cost by comparing the number of problem patients on the graphs.

### *8.2.2 Strategy regeneration*

In this experiment, we compare the version with a SG node for strategy regeneration to the one without the SG node. As shown in Sect. 6, there is a SG node where a new strategy chain is generated if the response from the entity is "No" to reflect the aspect of real-time and dynamic environments. For the version excluding the SG node, we simply moved to the next world and asked the next entity instead of strategy regeneration.

Figure 6 illustrates the distribution generated by our model with strategy regeneration and without strategy regeneration. The graph on the left represents the case of four entities and five patients and one on the right represents the case of six entities and five patients. The $x$-axis of each graph denotes the number of problem patients, and the $y$-axis the frequency of each value on $x$-axis after running our simulation 100 times. The solid line represents the version including a SG node, and the dotted line represents the version excluding the SG node.

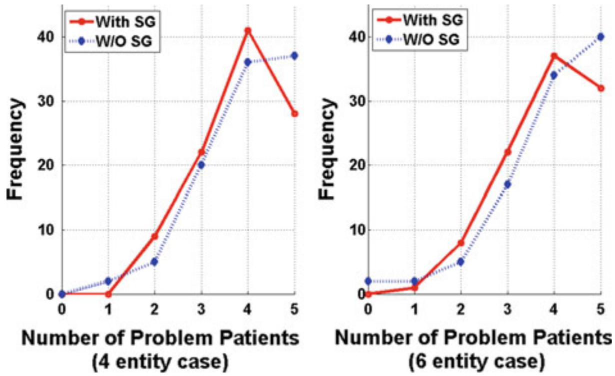


**Fig. 6** Our model with and without a SG node

In Fig. 6, we can find the peak of the solid line at 4 on the $x$-axis and inclined to the right. However, the peak of the dotted line is spotted at 5 on the $x$-axis. This implies that 4 problem patients are mostly found under the version with strategy regeneration but 5 problem patients under the version without strategy regeneration. In other words, the version including a SG node outperforms the one which does not regenerate a strategy chain.

### *8.2.3 Task criticality*

In this experiment, we compare the version with weights by task criticality of the patients to the one without weights. The expected quality of decision of each entity is determined by his/her lack of expertise factor as presented in Formula 10. In this section, we compare the version with weights and the one without weights. The version without weights implies that every entity has equal expected quality of decision.

Figure 7 illustrates the distribution generated by our model with strategy regeneration and without weights. The graph on the left represents the case of four entities and five patients and one on the right represents the case of six entities and five patients. The $x$-axis of each graph denotes the number of problem patients, and the $y$-axis the frequency of each value on $x$-axis after running our simulation 100 times. The solid line represents the version reflecting weights, and the dotted line represents the version excluding weights. In Fig. 7, we can find the peak of the dotted line located in a higher position than the peak of the line at 4 on the $x$-axis and inclined to the right. This implies that there have been more problem patients during simulations with the version without weights (dotted line) than one reflecting weights (solid line). In other words, the version with weights outperforms the one which does not reflect weights by comparing the number of problem patients on the graphs.

## 9 Relationship to other papers in this special issue

The aim of this special issue is to explore topics at the intersection of personalization and e-health. Our work is most closely related to that of Chittaro et al. (2011), which

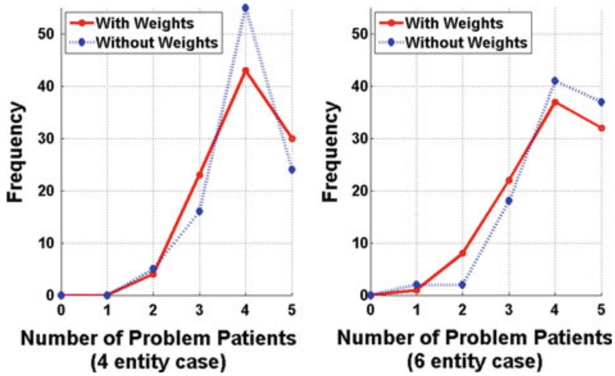


**Fig. 7** Our model with and without weights

is also focused on the hospital setting of emergency room decision making. Their approach is similar to ours, in its acknowledgement of the need for rapid, real-time decisions and its proposal to develop solutions that take into consideration the severity of the patient's current state (what we refer to as task criticality). Their focus is different, however, both in their exploration of the interface needs of users to support decision making and in their detailed modeling of a particular class of patients (disabled people).

These two distinct elements arise as well in the work of Lindgren (2011) in this volume, who propose differing decision making support to users with varying background and who distinguish typical and atypical patients whose care must be supported. In our work, the primary distinction between our medical experts is their expertise, together with their level of bother, while our patients are characterized by the core feature of criticality. Colineau and Paris (2011), like Lindgren (2011), explore how to design supportive interfaces for users, this time for the patients, and with a focus on offering tailored feedback. But they, too, continue to introduce a particular kind of user who is supported, namely an entire family.

Both Lindgren (2011) and Colineau and Paris (2011) also emphasize the value of teamwork in enacting effective care of patients. In particular, Colineau and Paris (2011) are able to demonstrate how setting a collective goal for a family resulted in considerable benefits. This proposal to have health decisions improved by engaging a team of individuals is aligned with our own approach which aims to introduce the most valuable experts into current health decisions for patients, to improve on the current state of that patient's care.

## 10 Contributions and future work

This research discusses the value of modeling the bother cost endured by users as part of an agent's reasoning about interaction. We first contribute to the field of user modeling by presenting methods not only to model the bother to a user but also to update this model, over time, in order to influence the problem solving activity of the agent. As discussed in Sect. 3, other researchers have considered modeling the bother cost to users arising from system-initiated interruption/interaction (e.g., Bauer et al. 2000; Horvitz and Apacible 2003). Our work extends that effort by proposing an integration of various factors into the modeling of bother cost and also by explicitly considering the problem of how to model the cost of bothering a user when there are multiple interactions initiated by various agents in a multiagent system. Our particular approach for reasoning about interaction in a multiagent setting goes beyond other adjustable autonomy efforts (Scerri et al. 2002) that do not attempt to model the cost of bothering users and do not provide for questions to be asked of users (interaction), focusing instead simply on asking users to assume the decision making. In all, we are offering a detailed framework for reasoning about interaction with users, based on specific proposals for the modeling of these users.

The particular approach to modeling bother cost presented in this paper coincides well with other research in user modeling, including a provision for stereotypical classification of users' base bother cost, attention state and willingness. Fleming (2004) discusses how values of variables in user models can be determined as a combination of modeling the specific user, the class to which that user belongs and the traits that are typical of all users. Although the context in that research is the modeling of a user's level of knowledge, the techniques presented may be extended to assist in the modeling of user willingness as well, namely to adjust weights on the contributions from each of the different models, over time, as more is learned about the specific user and his or her preferences.

Our research also provides insights into how to design interactive dialogues with users. It can be seen in part as an extension to the topic of determining when to initiate clarification dialogues (see van Beek and Cohen 1991), providing algorithms for systems to reason about interaction, based on a modeling both of the task at hand and the specific user. In particular, we focus on how to incorporate bother cost modeling when reasoning about whether to interact with a user. This contrasts with other research from the user modeling community, such as that of Ardissono et al. (1993), Carberry et al. (1999), Raskutti and Zukerman (1997), or Wu (1991), focused on modeling a user's plans in order to determine what to say, or that of Paris (1991), focused on how to tailor the content of the interaction, based on the user's level of expertise.

Beyond user modeling, our research relates well to efforts in designing mixed-initiative systems: ones where either the user or the system can take the initiative to direct the problem solving or begin a dialogue, as part of a collaborative effort (Hexmoor et al. 2003). One challenge in designing these systems is developing sound domain-independent procedures for systems to reason about when to take the initiative—the issue of providing a principled basis for the design of the system. In this paper, we have emphasized the value of having an explicit model of the bother to a user, as part of that principled decision making. We have, moreover, outlined how agents can account for interactions initiated by other agents, as they make their own decisions about interaction.

Some key comparisons with other efforts in designing mixed-initiative systems can be made. First, researchers like Barber et al. (2003) reinforce the need for a decision-theoretic approach to constructing these systems. They discuss the value of global strategies such as strong collaborative ties between the system and user. Our research, with its modeling of bother cost, would provide these systems with opportunities to avoid interaction, for instance in cases where the user is currently occupied with other tasks. Researchers such as Myers and Morley (2001, 2003) have focused more on developing strategies for users to specify conditions under which the system will suspend its processing and seek further input. In cases where they may be multiple agents all interacting with the same user, some kind of global coordination of these activities is needed. We have outlined a method using proxy agents to allow the possible bother to the user to be considered by each of the individual agents.

Our research is also relevant to the field of designing adjustable autonomy multi-agent systems. In contrast to the approach developed in E-Elves (Scerri et al. 2002), we propose that agents select the best user to transfer control to based not only on the expected decision quality, probability of response of the user and waiting cost, but also on other factors such as the user's willingness to interact and the user's attention state. The bother cost model introduced here can be viewed as a useful starting point for other researchers in the field wanting to explore the concept of bother cost and our framework provides a focal point for testing new bother cost models. We also extend the concept of transfer of control strategy, developed in the E-Elves project (Scerri et al. 2002) to allow agents to also reason about generating queries to users and then we discuss how individual agents can coordinate their proposed strategies by communicating with proxy agents concerning bother costs.

Other researchers, such as Schreckenghost et al. (2002), have also discussed the use of proxy agents in multi-user multiagent environments. That work is focused however on coordinating the completion of tasks. But the authors discuss the value of considering, as future work, how to address interruptions to users and agents. Our research begins to resolve this problem, by proposing methods for reasoning about whether to interact.

One element of our current research that would benefit from further exploration in the future is our proposed use of a base bother cost for each interaction with the user. Our current proposal is to have a scale of possible values, with decision making interruptions carrying a higher base of bother than interactions that simply elicit preferences. The work of various researchers focused more on measuring in greater detail the actual bother endured by users that comes with varying kinds of interruptions may in fact provide some insights into how the base bother cost value may be independently set. The work of Bailey and Iqbal (2008) may be particulary valuable here. This research suggests that different tasks provided to users may in fact have distinct levels of difficulty, as evidenced by their tracking in the change of pupil sizes of these users.

Moreover, we are interested in continuing to investigate methods for appropriately setting the Attentional State Factor that influences the calculation of the cost of bother to the user. Of relevance here is the work of various researchers who have developed models that explore how to set the state of interruptibility of a user. Included is the work of Fogarty et al. (2004b) which suggests that information provided by users can be used as the basis of an initial model, the work of Horvitz and Apacible (2003) which indicates that the application that is currently in focus will influence the modeling of the user's interruptability and the work of Iqbal and Bailey (2005) which also confirms that the kind of task which is interrupted when the user is approached will be an important factor.

Another interesting path for future work is to examine additional uses for the proxy agents, introduced to assist in the dissemination of information about the user to the various agents in the system. Since the user's reply always goes through the proxy agent, the proxy can actually build a better model of its user. This would allow the proxy agent to make better decisions for its user in the future (if, for instance, the proxy agent also takes on the role of a personal assistant agent). Interesting future work can be done exploring the way that proxy agents can facilitate user modeling, which in turn assists with the proper setting of various model parameter values. Another valuable path for future work would be to investigate the merit of allowing proxy agents to make use of deeper models of their users to improve the management of requests for transfers-of-control. For example, if the proxy agent already knows how the user would respond to a query, it can just reply back to the TOC requesting agent, without relaying the question to its user. This results in no bother to the user, and a faster response time to the requesting agent, achieving an effect similar to caching. There may be some value as well to having all the queries go through the proxy agent, in order to ensure that the queries are presented in a consistent manner to the user (e.g., present the query in a style suitable for the user, to make it easier for the user to respond).

## 10.1 Future work for the application to hospital decision making

We have also projected our proposed model for reasoning about interaction with users, sensitive to bother, into the hospital decision making scenario. In so doing, we have provided a framework that is sensitive to the dynamic nature of the environment and its need to update variables. We have also acknowledged the importance of modeling the criticality of the tasks at hand, to adjust the reasoning about which experts to consult (and the balance of the expected quality of decision, compared to the cost of bother, for these users).

As we extended our original model of bother cost to be useful for dynamic settings such as hospital decision making, it became apparent that several of the parameters within the formulae have specific relationships between each other. Another thread for future research is to explore more carefully what these interrelationships are, in order to exploit this during the phase where these parameters are learned and set.

The model presented in Sects. 6 and 7, of use in dynamic, time critical settings such as hospitals, has in fact been constructed to employ various default values for parameters. For example, we propose always waiting a fixed time when there is silence following a request for assistance from a medical expert. We also assume in our example scenarios that the probability that any medical expert will respond instead of reverting to silence is the same, for all of the experts. For future work, one valuable thread will be to work on acquiring the values for the parameters being used in our calculations. We can imagine two primary directions, here.

The first is to set aside our current simplified assumptions, included to enable the calculations to be done quite quickly (and to experiment with our proposed model). So, for example, user unwillingness in future investigations may not be simply a function of attentional state and expertise level. It could instead be assessed, for each separate medical specialist (learning over time just how willing or unwilling this person inherently is).

The second is to investigate the most effective methods for actually learning how to set the parameter values that are used within the formulae. One possible starting point for this work is the research of Kapoor and Horvitz (2008) which advocates the use of active learning, to involve the user in the process of progressively determining appropriate parameter values for interruptions.

One particular direction that we are currently exploring as well is how to introduce a richer modeling of the experts and their possible willingness to interact, as part of the reasoning about interaction. While our current formulation of user unwillingness depends on the attentional state factor of the expert (whether that person is currently attending to another patient or not) and on the level of expertise (whether that person feels that they could be of some assistance), we recognize that it is also important to be modeling more deeply the current task that the expert is involved with. For example, an expert who has just begun to attend to a very critical patient will not only have a high attentional state factor but also what we would term a high stress level, which would make them inherently more unwilling to assist, whereas an expert who is about to complete their current task with a patient who is in a less critical state may in fact be fairly willing to help, even though the state is still one where the expert is busy. New formulae to reflect these considerations are currently being developed.

In addition, we are exploring further use of the task criticality factor that currently influences the weight of expected quality of decision, compared to the cost of bothering an expert. For some very critical tasks, it becomes more important to ensure that the level of expertise of the attending doctor is appropriate, even if there is a wait for that expert, which also affects the state of the patient. A related question is whether the calculations required to determine who is best to ask may be best curtailed, in order to bring some expert in to attend to the patient, in critical scenarios. One can imagine for instance opting for less optimal solutions in those cases and this is another avenue for future study.

It is important to note other limiting assumptions currently within the model presented for hospital settings, which also need to be explored more extensively, for future work. We include here preliminary ideas for the tasks of (i) limiting the length of strategies being generated, in order to introduce regeneration, with novel parameter values; (ii) setting the non-expertise level of the medical specialists and (iii) reasoning about the probability of response from the specialists. For the first item, we believe it will be promising to simply cut off the number of experts who are included in a chain depending on the criticality of the task. So, very critical tasks should result in more frequent strategy regeneration. For the second item, it should be possible to try to track more carefully the level of expertise of the specialists relative to the specific condition of the patients. As these specialists elect to assist more patients, their expertise should increase (though their expected quality of decision making should in part increase from greater knowledge and in part decrease from additional tiredness). For the third item, we believe that the probability of response of a specialist can be modeled more deeply in terms of the number of patients they have been seeing and more specifically in terms of both the time when they began addressing their current patient (if currently occupied) and the severity of that patient. We leave for future work the development of more detailed models to determine how best to set this parameter value.

We note as well that we are currently assuming a zero cost for strategy regeneration in our examples. To proceed with updated parameter values is of some definite merit. Yet, regeneration will also consume some time, at the very least. We leave open for future work more careful consideration of the value of the strategy regeneration cost.

Currently, we perform strategy regeneration when there is a “No” response, as an indication that our parameter modeling may need to be refreshed. For future work, we can explore the value of also introducing strategy regeneration after a certain chain of “Silence” responses. While simply waiting for a “Yes” reply is desirable, under certain circumstances opting for a refresh of the calculation may be beneficial.

It is also important to note that for the specific application of hospital decision making, our aim in fact was to resolve the question of who should be asked, what, when, in order to coordinate the use of personnel most effectively. While our current model is restricted to simply asking each potential medical expert whether they can come to assist with a patient, one may imagine extending this to incorporate other kinds of information gathering, towards the improvement of the immediate decision to be made. In addition, one could also imagine a more detailed communication with each expert, as part of the decision making. Currently we ask the question of whether the expert can come to assist with a current patient. If we are able to succinctly convey as well the level of criticality of that patient (or perhaps the current challenge to be addressed), this may lead to differing kinds of responses from the experts and more effective modeling. We leave this consideration for future work.

We also currently assume that distinct users are asked within one chain (and that chains are kept short). It may be useful to reason about possibly returning to ask a previous user, especially if that user’s attentional state has changed from “busy” to “relaxed”. In addition, when the number of users becomes fairly high, we may consider techniques for reducing the number of strategies that are generated, for example to limit consideration only to those entities with high initial expected quality of decision and probability of response.

As such, a valuable thread for future research is to incorporate research on delivering sensor readings and enabling effective communication networks within hospital settings. The integration of workflow modeling, sensing and networking is in fact the overall aim of the hSITE project (Plant 2008) currently being conducted by a group of researchers across Canada. In particular, carefully placed sensors could also serve the purpose of providing values for parameters that are useful to reason about as part of our model for determining whether to bother a particular medical expert in hospital scenarios. We view this as a promising topic for future research. At the moment, the hSITE project is refining proposed models for fusion of data from multiple sensors, in order to then offer a significant approach to achieve the sensing.

Once the team of researchers focused on proposing the most effective implementation and integration of sensors within the hospital setting has provided a clear basis for the collection of user and expert data, we would then be able to leverage this in the continued validation of our model, towards possible extensions as well. Our work is being used currently as part of the overall modeling of the workflow within the hospital setting—to whom the information should be flowing and when. How data is actually obtained, recorded and updated is seen as a companion question that will be integrated with the models that are being developed, as will be ongoing proposals for the effective networking of hospital communications.

Another group of researchers from the hSITE project is also currently conducting a detailed analysis of the workflow needs and challenges of healthcare professionals within hospital settings. Once this study is complete, we will then be able to fine tune our model to match the requirements that this research project will specify. Enlisting a group of medical professionals as users to further evaluate the effectiveness of our proposed bother cost model might be another direction for future research, at that point in time. Further studies with medical professionals could serve to quantify in greater detail the impact of our model for reasoning about bother with users in real hospital settings.

A final direction for future research will also be to move beyond our current focus on which medical expert should attend to each specific patient, to explore the handling of multiple patients simultaneously through effective coordination possibly using proxy agents for the communication management, or reasoning about a set of resources that all need to be brought to bear in order to care for each current patient.

## 11 Summary

This paper has presented a framework for modeling bother cost to users in collaborative problem solving environments, detailing how an agent can incorporate bother cost into reasoning about whether to interact with a user. We elaborate in particular on how to effectively model the bother cost endured by a user when it is possible for that user to be bothered by other agents in the system. This requires a method for agents to share information and adjust their decision-making process, based on this new information. This research therefore provides a valuable framework for anyone designing a system of multiple agents reasoning about multiple users, allowing for agents to adjust their autonomy and interact with users during the processing.

More specifically, our research has shown that it is possible to develop an effective model that is motivated by achieving appropriate user modeling, in order to provide the basis for reasoning about interaction in hospital settings, to improve the decision making. As such, it offers contributions which are at the heart of this special issue of the UMUAI journal.

**Acknowledgments** The authors gratefully acknowledge the support of the Natural Sciences and Engineering Research Council (NSERC) of Canada and in particular their support of the Strategic Networks of Research hSITE project. Thanks as well to Diane Doran, R.N. and Ivana Matic of the Lawrence S. Bloomberg Faculty of Nursing for their discussions regarding hospital scenarios.

## Appendix A

Example to illustrate multiagent coordinating of bother

To give a better understanding of how the coordinating of bother operates, we will present an example scenario. The model parameters used in this example are the following (where the earlier section on the bother cost model describes these model parameters):

[$TOC\_Base\_Bother\_Cost$] 10 *for all TOCs*
[$User\_Unwillingness\_Factor$] $0.5(Willing)$, $1(Neutral)$, $2(Unwilling)$
[*Time discount factor* $\beta$] 0.90
[$Attention\_State\_Factor$] 1 *for all users*
[*Number of agents in system*] 5
With the basic user data being:[21]

| | Willingness type | EQ decision class1 | EQ decision class2 |
|---|---|---|---|
| $User_1$ | $Willing$ | 60 | 77 |
| $User_2$ | $Neutral$ | 85 | 85 |
| $User_3$ | $Unwilling$ | 100 | 50 |

And the following scenario events:
[$Event_1$] $Agent_1$ needs a decision of class 1 at time step 1.
[$Event_2$] $Agent_5$ needs a decision of class 1 at time step 2.
[$Event_3$] $Agent_4$ needs a decision of class 2 at time step 10.

On [$Event_1$], $Agent_1$ does the following:

- Computes the following bother cost for the different users:
  - $BC(User_1) = Init + BC\_Inc\_Fn(BSF, User\_Unwillingness\_Factor) = User\_Unwillingness\_Factor \times Attention\_State\_Factor \times TOC\_Base\_Bother\_Cost + 0 = 0.5 \times 1 \times 10 = 5$
  - $BC(User_2) = 1 \times 1 \times 10 = 10$
  - $BC(User_3) = 2 \times 1 \times 10 = 20$
  - Note: Assume that the system just started, so there has been no bother before, so $BSF = 0$. Recall from the Bother Cost Model section that $BC\_Inc\_Fn(BSF, User\_Unwillingness\_Factor) = BSF^{UserWillingnessAdjustment}$, where $UserWillingnessAdjustment = 0.75$ (*Willing*), *UserWillingnessAdjustment* = 1 (*Neutral*), and *UserWillingness Adjustment* = 1.25 (*Unwilling*).
- Computes the utility of transferring control to each user:
  - $Utility(User_1) = EQ^{D_1}_{User_1} - BC(User_1) = 60 - 5 = 55$
  - $Utility(User_2) = 85 - 10 = 75$
  - $Utility(User_3) = 100 - 20 = 80$
- $Agent_1$ selects $User_3$ for the TOC, and the process proceeds as outlined in the Coordination by Broadcast section.

On [$Event_2$], $Agent_5$ does the following:

- Computes the following bother cost for the different users (based on its knowledge of $User_3$ having being interrupted):

[21] To acknowledge that different users have different capabilities in handling different types of decisions, we have allowed for users to have different $EQ$ values for different decision classes. This makes the example more realistic (and interesting).

- $BSF = 0$, $BC(User_1) = Init + BC\_Inc\_Fn(BSF, User\_Unwillingness\_Factor)$ $= 5 + 0 = 5$
- $BSF = 0$, $BC(User_2) = 10 + 0 = 10$
- $BSF = \sum_{toc \in PastTOC} TOC\_Base\_Bother\_Cost(toc) \times \beta^{time\_elapsed(toc)}$ $= 10 \times 0.9^{2-1} = 9$, $BC(User_3) = Init + BC\_Inc\_Fn(BSF, User\_Unwillingness\_Factor) = 20 + 9^{1.25} = 35.56$

– Computes the utility of transferring control to each user:

- $Utility(User_1) = EQ^{D_1}_{User_1} - BC(User_1) = 60 - 5 = 55$
- $Utility(User_2) = 85 - 10 = 75$
- $Utility(User_3) = 100 - 35.56 = 64.44$

– $Agent_5$ selects $User_2$ for the TOC, and the process proceeds as outlined in the Coordination by Broadcast section.

On $[Event_3]$, $Agent_4$ does the following:

– Computes the following bother cost for the different users (based on its knowledge of $User_2$ and $User_3$ having being interrupted):

- $BSF = 0$, $BC(User_1) = 5 + 0 = 5$
- $BSF = 10 \times 0.9^{10-2} = 4.30$, $BC(User_2) = 10 + 4.30 = 14.30$
- $BSF = 10 \times 0.9^{10-1} = 3.87$, $BC(User_3) = 20 + 3.87^{1.25} = 25.43$

– Computes the utility of TOC'ing to each user:

- $Utility(User_1) = 77 - 5 = 72$
- $Utility(User_2) = 85 - 14.30 = 70.7$
- $Utility(User_3) = 50 - 25.43 = 24.57$

– $Agent_4$ selects $User_1$ for the TOC, and the process proceeds as outlined in the Coordination by Broadcast section.

Altogether, after the three events, $TotalUtility = 80 + 75 + 72 = 227$.

## Appendix B

EU values for Scenario 1

In this Appendix, we show the values obtained for the $EU$ of each of the 24 different strategies that could be considered in Scenario 1 of Sect. 7.2. These are displayed in Table 12.

**Table 12** Expected utility of strategies at the time the patient arrives

| No. | Expected utility ($EU$) | Strategy chain |
|---|---|---|
| 1 | 189.055932 | $e_1$-$e_2$-$e_3$-$e_4$ |
| 2 | 186.222260 | $e_1$-$e_2$-$e_4$-$e_3$ |
| 3 | 197.846134 | $e_1$-$e_3$-$e_2$-$e_4$ |
| 4 | 198.391167 | $e_1$-$e_3$-$e_4$-$e_2$ |
| 5 | 189.251631 | $e_1$-$e_4$-$e_3$-$e_2$ |

**Table 12** continued

| No. | Expected utility ($EU$) | Strategy chain |
|---|---|---|
| 6 | 188.493741 | $e_1$-$e_4$-$e_2$-$e_3$ |
| 7 | 49.190011 | $e_2$-$e_1$-$e_3$-$e_4$ |
| 8 | 46.356339 | $e_2$-$e_1$-$e_4$-$e_3$ |
| 9 | 44.739695 | $e_2$-$e_3$-$e_1$-$e_4$ |
| 10 | 32.315057 | $e_2$-$e_3$-$e_4$-$e_1$ |
| 11 | 4.896448 | $e_2$-$e_4$-$e_3$-$e_1$ |
| 12 | 5.150315 | $e_2$-$e_4$-$e_1$-$e_3$ |
| 13 | 138.701167 | $e_3$-$e_2$-$e_1$-$e_4$ |
| 14 | 126.276528 | $e_3$-$e_2$-$e_4$-$e_1$ |
| 15 | 177.340219 | $e_3$-$e_1$-$e_2$-$e_4$ |
| 16 | 177.885253 | $e_3$-$e_1$-$e_4$-$e_2$ |
| 17 | 136.679229 | $e_3$-$e_4$-$e_1$-$e_2$ |
| 18 | 133.046656 | $e_3$-$e_4$-$e_2$-$e_1$ |
| 19 | 35.955365 | $e_4$-$e_2$-$e_3$-$e_1$ |
| 20 | 36.209232 | $e_4$-$e_2$-$e_1$-$e_3$ |
| 21 | 44.745566 | $e_4$-$e_3$-$e_2$-$e_1$ |
| 22 | 48.378139 | $e_4$-$e_3$-$e_1$-$e_2$ |
| 23 | 49.867505 | $e_4$-$e_1$-$e_3$-$e_2$ |
| 24 | 49.109615 | $e_4$-$e_1$-$e_2$-$e_3$ |

## References


AAAI: Call for papers. Technical report SS-99-06. In: AAAI Spring Symposium on Agents with Adjustable Autonomy, Stanford, CA, USA (1999)

Altmann, E., Trafton, J.: Task interruption: Resumption lag and the role of cues. In: Proceedings of the 26th Annual Conference of the Cognitive Science Society (CogSci 2004), Chicago, IL, USA, pp. 42–47 (2004)

Ardissono, L., Lombardo, A., Sestero, D.: A flexible approach to cooperative response generation in information-seeking dialogues. In: Proceedings of the 31st Annual Meeting of the Association of Computational Linguistics, Columbus, OH, USA, pp. 274–276 (1993)

Bailey, B.P., Iqbal, S.T.: Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. ACM Trans. Comput. Hum. Interact. **14**(4), 1–28 (2008)

Bailey, B., Konstan, J., Carlis, J.: The effects of interruptions on task performance, annoyance, and anxiety in the user interface. In: Proceedings of Human–Computer Interaction—INTERACT'01, Tokyo, Japan, pp. 593–601 (2001)

Barber, K., Gamba, I., Martin, C.: Representing and analyzing adaptive decision making frameworks. In: Hexmoor, H., Castelfranchi, C., Falcone, R. (eds.) Agent Autonomy, pp. 243–280. Kluwer, Norwell (2003)

Bauer, M., Dengler, D., Paul, G.: Instructible agents for web mining. In: Proceedings of International Conference on Intelligent User Interfaces, New Orleans, LA, USA, pp. 21–28 (2000)

Carberry, S., Chu-Carroll, J., Elzer, S.: Constructing and utilizing a model of user preferences in collaborative consultation dialogues. Comput. Intell. **5**(3), 185–217 (1999)

Chen, D., Vertegaal, R.: Using mental load for managing interruptions in physiologically attentive user interfaces. In: CHI '04 Extended Abstracts on Human Factors in Computing Systems, Vienna, Austria, pp. 1513–1516. ACM Press, New York (2004)

Cheng, M.: A hybrid transfer of control approach to designing adjustable autonomy multiagent systems. Master of Mathematics Thesis, University of Waterloo, Waterloo, ON (2005)
Cheng, M., Cohen, R.: A hybrid transfer of control model for adjustable autonomy multiagent systems. In: Proceedings of AAMAS'05—Fourth International Joint Conference on Autonomous Agents and Multiagent Systems 2005, Utrecht, Netherlands, pp. 1149–1150 (2005)
Chittaro, L., Carchietti, E., Marco, L.D., Zampa, A.: Personalised emergency medical assistance for disabled people. User Model. User-Adapt. Interact. **21** (this issue) (2011)
Colineau, N., Paris, C.: Motivating reflection about health within the family: the use of goal setting and tailored feedback. User Model. User-Adapt. Interact. **21** (this issue) (2011)
Fleming, M.: Reasoning about interaction in mixed-initiative artificial intelligence systems. Ph.D. Thesis, University of Waterloo (2003)
Fleming, M.: The use of increasingly specific user models in the design of mixed-initiative systems. In: Proceedings of the Seventeenth Canadian Conference on Artificial Intelligence. University of Western Ontario, London, ON, Canada, pp. 434–438 (2004)
Fleming, M., Cohen, R.: A decision procedure for autonomous agents to reason about interaction with humans. In: Proceedings of the AAAI 2004 Spring Symposium on Interaction Between Humans and Autonomous Systems Over Extended Operation, San Jose, CA, USA, pp. 81–86 (2004)
Fogarty, J., Hudson, S., Atkeson, C., Avrahami, D.: Predicting human interruptibility with sensors. In: Proceedings of ACM Transactions on Computer–Human Interaction, pp. 257–264, ACM Press, New York (2004a)
Fogarty, J., Hudson, S.E., Lai, J.: Examining the robustness of sensor-based statistical models of human interruptibility. In: CHI '04: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vienna, Austria, pp. 207–214 (2004b)
Hexmoor, H., Falcone, R., Castelfranchi, C. (eds.): Agent Autonomy. Kluwer, Norwell (2003)
Horvitz, E., Apacible, J.: Learning and reasoning about interruption. In: Proceedings of the 5th International Conference on Multimodal Interfaces (ICMI'03), Vancouver, BC, Canada, pp. 20–27. ACM Press, New York (2003)
Horvitz, E., Koch, P., Kadie, C., Jacobs, A.: Coordinate: probabilistic forecasting of presence and availability. In: Proceedings of the 18th Conference in Uncertainty in Artificial Intelligence (UAI'02), Edmonton, AB, Canada, pp. 224–233. Morgan Kaufmann, San Fransisco (2002)
Iqbal, S.T., Bailey, B.P.: Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. In: CHI '05: CHI '05 Extended Abstracts on Human Factors in Computing Systems, Portland, OR, USA, pp. 1489–1492 (2005)
Kapoor, A., Horvitz, E.: Experience sampling for building predictive user models: a comparative study. In: CHI '08: Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems, Florence, Italy, pp. 657–666 (2008)
Lindgren, H.: Towards personalized decision support in the dementia domain based on clinical practice guidelines. User Model. User-Adapt. Interact. **21** (this issue) (2011)
Maheswaran, R., Tambe, M., Varakantham, P., Myers, K.: Adjustable autonomy challenges in personal assistant agents: a position paper. In: Agents and Computational Autonomy, Melbourne, Australia, pp. 187–194 (2003)
McCrickard, D., Catrambone, R., Chewar, C., Stasko, J.: Establishing tradeoffs that leverage attention for utility: empirically evaluating information display in notification systems. Int. J. Hum. Comput. Stud. **58**(5), 547–582 (2003)
Myers, K., Morley, D.: Directing agent communities: an initial framework. In: Proceedings of the IJCAI-01 Workshop Autonomy, Delegation and Control: Interacting with Autonomous Agents, Seattle, WA, USA (2001)
Myers, K., Morley, D.: Policy-based agent directability. In: Hexmoor, H., Castelfranchi, C., Falcone, R. (eds.) Agent Autonomy, Kluwer, Norwell (2003)
Paris, C.: Responding to a user's domain knowledge: incorporating the user's level of expertise into a generation system. Comput. Intell. **7**(2), 71–93 (1991)
Plant, D.: hSITE: healthcare support through information technology enhancements. NSERC Strategic Research Network Proposal (2008)
Raskutti, B., Zukerman, I.: Generating queries and replies during information-seeking interactions. Int. J. Hum. Comput. Stud. **47**(6), 689–734 (1997)
Scerri, P., Pynadath, D., Tambe, M.: Why the elf acted autonomously: towards a theory of adjustable autonomy. In: Proceedings of AAMAS'02, Bologna, Italy, pp. 857–864. ACM Press, New York (2002)

Schreckenghost, D., Martin, C., Bonasso, P., Kortenkamp, D., Milam, T., Thronesbery, C.: Supporting group interaction among humans and autonomous agents. In: Proceedings of AAAI02 Workshop on Autonomy, Delegation and Control: From Inter-Agent to Groups, Edmonton, AB, Canada, pp. 361–369. AAAI Press, New York (2002)

van Beek, P., Cohen, R.: Resolving plan ambiguity for cooperative response generation. In: Proceedings of the 12th International Joint Conference on Artificial Intelligence, Sydney, NSW, Australia, pp. 938–944 (1991)

Vertegaal, R.: Attentive user interfaces. In: Special Issue on Attentive User Interfaces, Communications of the ACM, vol. 46. pp. 30–33. Kluwer, Dordrecht (2003)

Walji, M., Brixey, J., Johnson-Throop, K., Zhang, J.: A theoretical framework to understand and engineer persuasive interruptions. In: Proceedings of the 26th Annual Meeting of the Cognitive Science Society (CogSci 2004), Chicago, IL, USA (2004)

Wu, D.: Active acquisition of user models: implications for decision-theoretic dialog planning and plan recognition. User Model. User-Adapt. Interact. **1**(3–4), 149–172 (1991)

## Author Biographies

**Robin Cohen** is a Professor of Computer Science at the University of Waterloo in Waterloo, ON, Canada. She received her B.A. Honours Mathematics from McGill University and her M.Sc. and Ph.D. in Computer Science from the University of Toronto, under the supervision of Ray Perrault. Her research interests include both user modeling and multiagent systems and she has served as the graduate supervisor of the other three co-authors. She is presently a part of the Theme 1 team of the hSITE Strategic Networks of Research project funded by NSERC, offering support to healthcare through information technology enhancements.

**Hyunggu Jung** is a Master's candidate and a research assistant at the Stanford Center for Biomedical Informatics Research at Stanford University. He received his B.S. in Computer Science and a minor in Business Economics from Korea Advanced Institute of Science and Technology (KAIST) in 2008, and his M.Math. degree from the University of Waterloo in 2010, under the supervision of Prof. Robin Cohen. His research interests include user modeling and multiagent systems in the field of artificial intelligence, as well as biomedical decision support to assist health-care workers in the medical domain. He was also part of the Theme 1 team of the hSITE Strategic Networks of Research project.

**Michael W. Fleming** is an Associate Professor of Computer Science at the University of New Brunswick. He received his B.Sc. in Mathematics and Computer Science from Mount Allison University and his M.Math. and Ph.D. degrees in Computer Science from the University of Waterloo. His research interests within the field of artificial intelligence include automated negotiation, preference elicitation, decision theory, user modeling, and mixed-initiative interaction.

**Michael Y. K. Cheng** received his Masters of Mathematics in Computer Science degree from the University of Waterloo in 2005 under the supervision of Prof. Robin Cohen. His research work was in the area of adjustable autonomy in multiagent systems. He is now currently a developer at Microsoft in Redmond, Washington.