A Reusable Framework for Health Counseling Dialogue Systems based on a Behavioral Medicine Ontology

Timothy Bickmore  
Assistant Professor, College of Computer and Information Science  
Northeastern University  
Boston, Massachusetts

Daniel Schulman  
College of Computer and Information Science  
Northeastern University  
Boston, Massachusetts

Candace Sidner  
Research Professor, Department of Computer Science  
Worcester Polytechnic Institute  
Worcester, Massachusetts

Corresponding author:
Timothy W. Bickmore  
Northeastern University  
Computer Science, WVH202  
Boston, MA 02115  
bickmore@ccs.neu.edu  
Phone: 617-373-5477
**Abstract**

Automated approaches to promoting health behavior change, such as exercise, diet, and medication adherence promotion, have the potential for significant positive impact on society. We describe a theory-driven computational model of dialogue that simulates a human health counselor who is helping his or her patients to change via a series of conversations over time. Applications built using this model can be used to change the health behavior of patients and consumers at low cost over a wide range of media including the web and the phone. The model is implemented using an OWL ontology of health behavior change concepts and a public standard task modeling language (ANSI/CEA-2018). We demonstrate the power of modeling dialogue using an ontology and task model by showing how an exercise promotion system developed in the framework was re-purposed for diet promotion with 98% reuse of the abstract models. Evaluations of these two systems are presented, demonstrating high levels of fidelity to best practices in health behavior change counseling.

**Keywords**

Dialogue system, behavioral informatics, consumer informatics, motivational interviewing, transtheoretical model, natural language processing.
1. Introduction

Poor lifestyle health behaviors, such as lack of physical activity and unhealthy dietary habits, are among the leading causes of death and chronic disease in the United States [1], yet a large proportion of Americans fail to meet guidelines published by U.S. public health authorities. For example, participation in moderate amounts of physical activity has important health benefits, including beneficial effects on risk factors for disease, disability, and mortality [2-6], yet a substantial proportion of the U.S. adult population remain underactive or sedentary [2]. Fruit and vegetable consumption plays a protective role in a large number of epithelial cancers [7, 8], and is associated with reduced risk for heart disease, stroke, and hypertension [9-15], yet only 23% of American adults meet the (year 2000) recommendation to consume five or more servings of fruit and vegetables per day [16, 17].

The “gold standard” for health behavior change remains one-on-one, face-to-face interaction with an expert human counselor, who has at his or her disposal a range of theory-based interventions that can be deployed based on the needs of the patient and the conversational context [18]. Unfortunately, one-on-one interventions only have the potential to reach a very small portion of the population, resulting in very low impact (efficacy multiplied by recruitment rate[19]). To reach a larger audience, a wide range of media-based and automated health behavior change interventions have been developed over the last two decades. Of these, the ones that come closest to the “gold standard” are those which autonomously communicate with patients using natural language dialogue, emulating the behavior of an expert counselor as closely as possible (The appendix provides sample transcripts of such dialogues for exercise promotion.)

Many health counseling dialogue systems have now been developed and evaluated in randomized trials. The design of these systems is typically based on one or more theoretical frameworks from behavioral medicine, such as the Transtheoretical Model (also known as the “stages of change” model) [20], Social Cognitive Theory [21], and Motivational Interviewing [22]. They have been designed to promote a range of health behaviors including diet, physical activity, smoking cessation, medication adherence, and chronic disease self-care regimens, and most have been shown to be effective on at least one outcome measure, compared to standard-of-care or non-intervention control conditions [18].

1.1. The Need for Reusable Computerized Behavioral Interventions

Although many health counseling dialogue systems have been shown to be efficacious, they are still being developed with little concern for reusability. In addition to representing a waste of money and resources due to duplications of effort by researchers, these issues impact the ability of these systems to be disseminated to health care practitioners. Adapting a computerized behavioral intervention developed for a research study for large-scale deployment by an institution or public health department typically requires significant modifications to the original system, and these modifications often represent an insurmountable barrier to dissemination.
These issues have recently become a concern of funding agencies who have been supporting the development of computerized behavioral interventions and are now interested in seeing these systems disseminated [23, 24].

1.2. How Ontologies Support Extensibility, Reuse, and Dissemination

An ontology is a taxonomic description of the concepts in an application domain and the relationships among them [25]. Computational ontologies have three primary uses in knowledge-intensive software systems: (1) they provide a formalism that can facilitate clarification and description of the fundamental concepts in a domain through consensus of experts; (2) they can facilitate interchange of information among diverse systems by describing, at various levels of detail, the kinds of data entities that can be exchanged, independent of the particular names the entities are given in each system; and (3) they can promote reuse of software components through the development of ontologies of the components themselves (descriptions of what the components do) and the data entities they operate on.

Ontologies can be used in several ways in a computerized behavioral intervention to promote reuse and interoperability, including the following (Figure 1):

![Figure 1. Ontologies in Health Counseling Dialogue Systems](image)

A **Theory Model** contains knowledge of the behavioral medicine theory the intervention is based on, including relationships among health behavior change theories, constructs, and counseling techniques (as in [26]). This knowledge can be readily reused across all computerized interventions that are based on the theories represented.
A Behavior Model contains knowledge of how health behavior change theories are applied to a specific health behavior, such as exercise promotion or smoking cessation. This includes general knowledge, such as whether the intervention is acquisition or cessation oriented, and whether the behavior can be shaped (incrementally modified) or not. This can also include very specific information such as homework exercises or tips that can be recommended to patients, or utterances or phrases the dialogue system uses in its conversations (e.g., that the behavior is called “walking”, that the units of measure are called “steps”, etc).

A Protocol Model contains knowledge about a particular behavior change intervention, including fixed parameters such as duration of the intervention and expected number of patient contacts per week, and ultimate behavioral criteria such as achieving “30 minutes a day of moderate physical activity on most days of the week” [27-29].

A User Model contains knowledge about a particular patient, including fixed “tailoring” parameters that affect the intervention and messages (e.g., gender, age, ethnicity) [30], information about the patient’s medical condition (e.g., from an EHR or PHR), construct measures (e.g., self-efficacy, stage of change), and information that is dynamically updated over the course of an intervention (e.g., daily behavioral goals).

An External Data Model describes data inputs and outputs from the system. This includes data that can be used by the system, such as medical data from an EHR or PHR, or measurement data from devices such as pedometers (for exercise promotion) or glucometers (for diabetes self-care promotion). This also includes data that can be output by the system for use by clinicians, including measurements and patient status information to the patient’s EHR or PHR, detailed transcripts or summary reports of individual counseling sessions (e.g., a psychiatrist might want to know exactly what the patient said in a given session), and alerts describing emergent conditions to be broadcast in a timely manner to appropriate clinicians. Describing these inputs and outputs in an ontology promotes interoperability and reuse.

Finally, a Task Model contains knowledge about how the system enacts the intervention, taking all of the above information into account. While the other knowledge representations are conceptual in nature, knowledge in the Task Model is procedural in nature, specifying sequences of steps the system should perform. Much of the knowledge in the Task Model is also specific to a particular mode of interaction (specifying, for example, the format of daily or weekly counseling conversations), while knowledge in the other models can be applicable to a wider range of computerized behavioral interventions, such as tailored print or web pages [31] in addition to dialogue. However, Task Models can be specified in ways that promote reuse of their components, through appropriate parameterization and hierarchical decomposition of tasks and sub-tasks.

In summary, explicit representation of concept and task knowledge in standard formalisms facilitates reuse by identifying and exposing all the ways a computerized behavioral intervention can be parameterized (e.g., tailoring parameters, behaviors, etc.), what its inputs and outputs are (to facilitate interoperability), and how its components (task models and the data they depend on)
can be extracted and recombined into new systems. This information can also be used to validate that an intervention has fidelity to a particular theoretical framework (as in [26]), or to provide a clinical summary of the theories, constructs and counseling techniques used in a given automated counseling session with a patient.

1.3. Overview
We have undertaken the development of a health behavior change counseling dialogue system that models the theory-driven therapeutic planning processes of a human health advisor during a counseling session, and is designed to be reusable. To address issues of reuse, we have built the system on public standard formalisms for computerized ontologies (OWL [32] and RDF [33]) and hierarchical task models (ANSI/CEA-2018 [34]), and have used these formalisms to explicitly represent knowledge in the ontologies described above. In the rest of this paper we review related work, then describe the behavioral medicine concept ontologies and task modeling language formalism we are using. We focus our discussion on the design of the Task Model and how its hierarchical design and parameterization promote reuse. We then describe two health counseling dialogue systems created using these representations to demonstrate reuse. We finally present results from formative evaluation studies before concluding.

2. Related Work
Bickmore and Giorgino provide a review of work in health education and behavior change dialogue systems developed to date [18]. Most systems have been deployed as Interactive Voice Response systems over the phone, although a few have been developed using animated Embodied Conversational Agent interfaces. The majority of fielded systems have been designed for system-initiated conversation using state machines or hierarchical transition networks, implemented with ad hoc representations of persistent knowledge and dialogue structure. Application domains span exercise, diet, and smoking cessation promotion, medication adherence, and chronic disease self-care management education and promotion (e.g., for coronary heart disease and diabetes mellitus). The majority of the approximately two dozen systems evaluated in clinical trials have been shown to be effective.

Beveridge and Fox describe a clinician-facing, speech-based dialogue system for breast cancer screening that uses an ontology of concepts in breast cancer together with a patient’s medical record to drive the behavior of the system. However, this behavior is very narrow in scope, used only by the natural language processing algorithms to allow clinicians to volunteer information or to state information using words and phrases not explicitly in the speech grammar.

Lenert, et al, describe an ontology of concepts from behavioral medicine, and use it, together with a task model in GLIF3 [35], to validate aspects of an exercise and diet intervention for adolescents [26]. Lenert’s ontology is one of the first developed for behavioral medicine—describing health behavior change theories, constructs, protocols and their inter-relationships—and we have developed our Theory and Protocol ontologies as an extension of his basic framework. However, his task model does not contain nearly enough detail to drive an automated
intervention, only representing protocols to the “module” level, each potentially abstracting many turns of counseling dialogue. In addition, GLIF3 is both inappropriate and lacks sufficient representational power for dialogue planning. The GLIF languages were developed to encode clinical guidelines for physicians to follow, in which there is a single best sequence of actions for clinicians take in a given situation. This is fundamentally different from dialogue planning in which goals (system, patient, and mutual) must be explicitly represented and reasoned about, and task fragments are dynamically selected and added to the current task tree at each turn in the conversation.

3. Ontology of Behavior Change Counseling Concepts

Our ontology is developed in Web Ontology Language (OWL) [32] using Protégé [36]. We developed our ontology based on a review of concepts in behavioral medicine (e.g., [37]), prior work in computerized interventions by others[18], requirements for counseling dialogue systems we have developed (e.g., [38, 39]), and discussions with experts in behavioral medicine and computerized interventions. We also reviewed a number of “upper level” ontologies for suitability, but decided that these representations were largely inappropriate for representing psychological concepts. We do build upon two existing ontologies: the behavioral medicine ontology from Lenert, et al [26], and the OWL-time ontology [40].

3.1. Theory and Protocol Models

Following Lenert, we model theories in behavioral medicine as collections of constructs, where a construct can be any concept or entity the theory is concerned with, from a psychological construct, such as self-efficacy, to a counseling technique, such as “problem solving” or “positive reinforcement”. For a counseling dialogue system, a “session” is a single conversation between a patient and the system, while an “intervention” is an intervention activity (e.g., a counseling strategy) deployed over one or more sessions that concerns one or more constructs. A “protocol” represents a complete, deployable intervention that includes specifications of the criteria behavior, duration and contact frequency, and all theories, constructs, and interventions that might be used.

3.2. Behavior Model

We extend Lenert’s ontology with a behavior model that specifies behavior-specific information that allows domain-independent counseling strategies to be developed. For example, phrases that refer to different aspects of a target behavior (“exercise” vs. “diet”, “walk” vs. “eat”, “steps” vs. “servings”, etc) are represented in the ontology and used as parameters for behavior-independent task models, such as dialogue to negotiate a goal, or to report on progress relative to a goal. The behavior model also lists declarative knowledge such as barriers to change (e.g., discomfort and boredom for exercise), homework exercises that can help motivate patients to change, and other information specific to different counseling techniques. The behavior model is used throughout the course of a conversation to select applicable task model fragments. For example, the system may assign “homework” to a patient (see lines 2-45 to 2-54 in the appendix); information in the
A behavior model is used to determine which homework assignment should be used based upon the choices that are appropriate at certain stages of behavior change.

### 3.3. User Model

We also define a user model, that specifies information about a specified patient, ranging from their name, where they are in the intervention (how long they have been in the intervention, the number of sessions they have had, behavioral goals they have committed to, etc), tailoring parameters (e.g., gender), time-indexed assessments of psychological construct (e.g., stage of change, motivation), and other information that may be required by a dialogue system (e.g., information about social chat topics, such as weather conditions).

### 4. Task Model for Health Behavior Change Counseling

For a health counseling dialogue system, the Task Model describes the structure of a counseling session by describing each of the atomic steps in the dialogue (system or patient utterances) and the relationships among them. To model counseling dialogue in an extensible and reusable manner, we use a standard declarative task representation language (CEA-2018 [34, 41]) for representing dialogue fragments, and have developed our own dialogue planner (DTask [42]) that uses these fragments to enact a dialogue with a patient.

This design was motivated by our previous experience in building systems that required authoring large quantities of dialogue [43, 44]. In these systems, we modeled dialogue using hierarchical transition networks, typically with agent utterances as states in a network and patient utterances as transitions. This formalism had the advantage of being simple to understand, particularly for clinicians and behavioral scientists without backgrounds in computer science or computational linguistics. However, we observed over the course of building several systems designed for health counseling, that reuse and generalizability of the dialogue between applications was often poor. In addition, while the therapeutic dialogue of these systems was based on well-established counseling theory and practice, this underlying knowledge was not explicitly modeled in the structure of the dialogues. Our current approach has made it possible to create more general and reusable dialogue models. At the same time, parts of the task library can be used to codify more domain knowledge into the declarative dialogue model representation.

#### 4.1. Task Model Representation Language: CEA-2018

Following Shared Plans theory [45, 46], we treat dialogue as a collaboration in which participants coordinate their action towards achieving a shared goal. The intentional structure of dialogue is modeled as hierarchical task decomposition: a library of recipes (goal decompositions), which may be used to achieve the overall goal of the collaboration.

Our dialogue recipes are specified in a declarative language that was recently approved as an ANSI Consumer Electronics Association standard: CEA-2018 [41]. CEA-2018 provides a number of abstractions that are useful in designing reusable task models including:
- Task (“applicability”) preconditions, enabling many recipes to be developed for accomplishing a given goal, with conflict resolution performed at runtime.
- Partial ordering of recipe steps, with complete ordering as default.
- Optional recipe steps.
- Declaration of input and output parameters for a given recipe, providing a locally scoped parameter passing mechanism.

CEA-2018 includes ECMAScript [47] for performing algorithmic tasks, data representation, and providing logical expressions for applicability preconditions.

To more easily describe dialogue actions, we extended the CEA-2018 specification with a declarative specification of dialogue based on the notion of an “adjacency pair” which is a pair of adjacent utterances that are logically related [48]. A turn of dialogue is specified as an adjacency pair template: an agent utterance and a multiple choice list of patient responses, which comprise a primitive action in the goal description. An adjacency pair template achieves a particular sub-task in the recipe model; there may be an arbitrary number of templates, as well as recipes, which can achieve a goal. Furthermore, each template automatically includes a patient response for clarification of the agent utterance. Clarifications either reproduce the same agent utterance, or an alternative utterance, if available. The surface form of agent and patient utterances are natural language utterances, but the template may include variables that are to be filled in when the utterance is produced (“template-based text generation” [49]). Utterance text can also be annotated with XML tags to specify, for example, prosodic or nonverbal behavior to be used in its performance by the system [50].

4.2. Task Model Interpreter: DTask
DTask is a hierarchical task decomposition-based dialogue planner, inspired by the COLLAGEN collaborative dialogue system [51]. It is designed to model and execute system-directed dialogue, where patient input is performed via selection from a multiple choice menu of possible utterances, dynamically updated at each turn of the dialogue. DTask uses a CEA-2018 recipe library to plan a task decomposition for each segment of a dialogue.

DTask provides support for loading and saving data and metadata represented in Resource Description Framework (RDF) [33] at runtime. RDF is a general-purpose knowledge representation format, distinguished by the use of a Universal Resource Identifier (URI) to identify each resource or individual being described, independent of a particular RDF document or knowledge store. Several ontology languages, such as the Web Ontology Language (OWL) are available for describing the structure of a RDF knowledge store, and for enabling inference upon them. A large ecosystem of software, both open-source and closed-source, exists for working with RDF, including editors, storage engines, and query and inference engines.

Arbitrary RDF documents can be loaded into the DTask knowledge store, either from files or a URL using an ECMAScript-based API. This function is currently used to load parts of the
ontology (Theory, Protocol, User and Behavior Models) into the runtime dialogue system and to save updated parts of the ontology (User Model) at the completion of a counseling session.

DTask can also generate an RDF description of a loaded set of recipes, including a unique identifier (URI) for each goal, goal decomposition, and dialogue turn, and a description of the goal-subgoal relationships. These descriptions can be annotated with arbitrary additional metadata, such as the behavior change theory a given recipe derives from. DTask can optionally use this metadata to filter a set of recipes, only including those that meet a specified set of criteria, for example only loading those models that adhere to a particular theoretical model. These models are useful in design and reuse by providing a visualization of the goals and recipes in a given dialogue system (Figure 2).

During the course of a dialogue, DTask maintains a representation of a “plan tree”, recording the recipes that are actively being used and their goal-subgoal task decomposition relationships (as in [51]). This representation is also maintained in RDF and can be exported at the completion of a dialogue to provide a complete description of the counseling session, including not only a complete transcript of the conversation (as in the Appendix) but the rationale for why the system used any particular utterance.

Figure 2. Fragment of the Task Model Hierarchy for Negotiating Goals
4.3. Health Counseling Task Model Design

Our current counseling Task Model is based on the transtheoretical ("stages of change") model [20], social cognitive theory, behavioral theory, and motivational interviewing. The transtheoretical model posits that individuals go through a set of distinct stages as they change, from Precontemplation (no intention to change) to Maintenance (well established habit), and that individuals need different interventions based on the stage they are in. We initially focused our effort on those counseling techniques relevant for exercise (walking) promotion, and developed models to address all stages of change, including Precontemplation, an area most automated interventions do not focus on.

Our approach to Precontemplation is drawn from Motivational Interviewing, a brief, directive, client-centered counseling method for enhancing intrinsic motivation to change by helping clients explore and resolve ambivalence [22]. The method includes a number of conversational strategies for eliciting “change talk” from clients in order to increase their motivation to and confidence in changing their behavior.

Our approach to the later stages of change involves long-term and short-term goal setting negotiation, positive reinforcement when goals are met (behavioral theory) and “problem solving” to overcome barriers when goals are not met (social cognitive theory). We also provide stage-appropriate homework assignments, tips and educational information throughout every intervention.

In addition to drawing from the counseling literature, we videotaped several counseling sessions between a clinical psychologist specializing in health behavior change and his patients. Transcripts of these interactions were used as the basis for many aspects of our system, from the overall conversational structure to specific recommendations and homework assignments. This psychologist also provided a significant amount of input to and feedback on our designs.

4.4. Task Model Implementation

Our goals in modeling the task structure of health behavior change counseling sessions were to maintain as much fidelity as possible to counseling practice while creating abstractions that promoted reusability.

The system’s knowledge of health behavior change counseling is represented in the models outlined in Figure 1, loaded into the system from RDF files and represented in the runtime system as ECMAscript objects.

4.4.1. High-level Structure of a Counseling Dialogue

Figure 3 shows a fragment of the task models which represent the top-level steps in a typical brief counseling dialogue. These steps include (line numbers reference the sample dialogues in the appendix):
Health Counseling Dialogue Systems 12

**Figure 3. Task model pseudocode showing the high-level structure of a counseling dialogue.**

1. opening the conversation with a greeting (2-1 to 2-2, and 3-1 to 3-2),
2. conducting ritualized social dialogue, including small talk (e.g., about the weather) and inquiries into the patient’s general emotional and physical state (e.g., “how are you?”) (2-3 to 2-12, and 3-3 to 3-16),
3. reviewing previously assigned behavioral tasks and goals (3-17 to 3-26),
4. assessing the patient’s current state in the intervention, including stage of change and attitude toward the behavior (2-13 to 2-18, and 3-27 to 3-32),
5. conducting counseling based on Motivational Interviewing (2-19 to 2-37, and 3-33 to 3-37),
6. negotiating new behavioral goals and tasks (2-38 to 2-54),
7. preparing to close the conversation [52] (2-55 to 2-58), and
8. delivering a farewell (2-59 to 2-60).

We note two particular features of this recipe. First, while some portions (e.g., the opening, social dialogue, pre-closing and closing) are applicable to dialogue in a wide variety of domains, much of the recipe is specific to counseling dialogue. Second, while we do not claim that this recipe represents the only possible decomposition of a counseling dialogue, we argue that this recipe produces a more highly coherent counseling dialogue than many other possible recipes, due to the particular ordering of subtasks. For example, reviewing previous tasks should be done near the beginning of a dialogue, as this represents a continuation of a previous conversation. Both reviewing tasks and assessing the patient’s current attitudes should be done before assigning new tasks, as both steps provide information that can be used to more effectively select new counseling tasks.
4.4.2. Structure of Motivational Interviewing Dialogue

In Motivational Interviewing, a counselor typically uses open-ended questions to elicit statements from the client relevant to their motivation or confidence to engage in the target behavior. The counselor can then respond with tactics designed to elicit further statements, such as requesting more information (e.g., “Tell me more…”) or reflection (i.e., restating the client’s utterance). Lines 2-21 to 2-37 of the appendix give an example of this type of dialogue. In our system these tactics are modeled in a reusable manner, so that new target behaviors, and new topics of discussion relevant to those behaviors, can be easily added. We follow the abstraction strategies discussed above, and create abstracted high-level recipes that do not contain information about any particular behavior or topic. A key challenge in this approach is defining an appropriate semantics for statements made by the patient during counseling.

Miller and Rollnick [22] discuss the tactics that a counselor can use to respond to two different types of patient statements: "change talk" indicates motivation for behavior change and/or confidence to change, while "resistance" indicates a lack of motivation and/or confidence. They also present 4 categories of meaningful statements, all of which can be positively valenced as change talk, or negatively valenced as resistance:

1) Positive or negative implications of the status quo.
2) Positive or negative implications of the proposed behavior change.
3) Statements of positive or negative outlook towards change.
4) Statements of intention or lack of intention to change.

The first two categories generally indicate motivation, or lack of motivation for change, while the latter two indicate confidence, or lack of confidence.

We have adapted this taxonomy to provide a simple domain-specific semantics for patient statements during the Motivational Interviewing portions of the dialogue. A patient statement is characterized by a tuple of three values:

1) valence: change talk or resistance
2) category: one of the 4 categories discussed above
3) content: the specific topic being discussed

Figure 4 gives an example of the use of these semantics in an adjacency pair template. This example shows an adjacency pair which can be used if the agent chooses to respond to a patient utterance about stress with a reflection tactic. There are currently over 400 such adjacency pairs specified for the “Reflection” tasks and for similar tasks representing other possible tactics a counselor may use. While this quantity of dialogue represents a significant authoring effort, such adjacency pairs are straightforward to write and can be extended by domain experts with little computer science knowledge or knowledge of the overall dialogue model. Conversely, there is a much smaller set of higher-level recipes which determine the appropriate tactic to use in order to
respond to a particular patient utterance. These high-level recipes are more complex and do require computer science expertise to create. However, the high-level recipes are independent of the particular topics discussed, and therefore these two portions of the dialogue model can be modified and extended independently of each other.

<table>
<thead>
<tr>
<th>task</th>
<th>Reflection</th>
</tr>
</thead>
<tbody>
<tr>
<td>input parameters</td>
<td>behavior, talk_valence, talk_category, talk_content</td>
</tr>
<tr>
<td>outputs</td>
<td>valence, category, content</td>
</tr>
<tr>
<td>adjacency pair</td>
<td></td>
</tr>
<tr>
<td>precondition: applicable when</td>
<td>behavior='exercise' and talk_valence='change_talk' and talk_category='status_quo_implication' and talk_content='stress'</td>
</tr>
<tr>
<td>agent utterance</td>
<td>You sometimes feel like you have too much stress.</td>
</tr>
<tr>
<td>user utterance</td>
<td>If I did exercise, it might be better.</td>
</tr>
<tr>
<td>valence</td>
<td>‘change_talk’</td>
</tr>
<tr>
<td>category</td>
<td>‘change_implication’</td>
</tr>
<tr>
<td>content</td>
<td>‘stress’</td>
</tr>
</tbody>
</table>

Figure 4. Example recipe for motivational interviewing

To summarize, the treatment of Motivational Interviewing dialogue given here illustrates how the careful specification of a domain-specific semantics allows the separation and abstraction of high-level recipes from low-level dialogue. This abstraction enables the creation of complex dialogue models which can be extended on multiple levels: with new target behaviors and topics of discussion, and with new higher-level tactics for counseling.

4.4.3. Structure of Behavioral and Social Cognitive Counseling Dialogue

Once a patient is past the Precontemplation stage, the agent begins negotiating long-term and short-term explicit behavioral goals (e.g., number of steps walked). The overall objective is to gradually move a patient from their current level of performance up to the criterion level. Goals are negotiated by giving patients some latitude in selecting their short-term (day-to-day) goals to account for fluxuations in their motivation, self-efficacy or personal situation (e.g., going away for a weekend).

Once a goal is established, patient performance relative to their goal is reviewed in subsequent counseling sessions. If they have met their goal, the agent provides positive reinforcement (“great job!”). We have observed (as have other researchers [53]) that patients do acquire a sense
of accountability to an automated counseling agent that motivates them to meet their goals. If goals are not met, the agent engages the patient in a problem-solving discussion to identify a specific barrier or obstacle to meeting their goals and makes suggestions for how the patient might overcome it.

4.5. Task Model Design for Reuse

The use of a hierarchical task decomposition as a model of dialogue permits the reuse of dialogue structures at different levels of abstraction. For example, we observed that several points in the dialogue required the agent to negotiate a numerical quantity with the patient, such as a long-term goal (e.g., an average number of steps to walk after several weeks) or a daily goal (e.g., a number of steps to walk before the next session). Although there may be many differences between the two dialogue fragments, the high-level steps are the same: the agent makes an initial recommendation, and depending on the patient’s answer, may respond with a higher or lower daily goal, or (in the case of strong resistance), a discussion of possible obstacles. Figure 5 shows a portion of a high-level recipe for negotiation, and an example of a dialogue turn that works with it. To extend the system to perform negotiation on long-term diet goals (for example), requires only adding a new adjacency pair for the “Recommend” task (and other similar tasks) which is applicable only for that behavior.

Figure 5. Example pseudocode for a high-level recipe and a low-level dialogue specification
5. Evaluation

We have developed two behavioral interventions using the above framework: the first to promote physical activity (walking) and a second to promote fruit and vegetable consumption. In this section we describe separate, formative evaluation studies of each system. In addition, we report on the amount of reuse between two behavioral interventions.

5.1 Evaluation of Exercise Promotion Intervention

We conducted a mixed-method, formative evaluation of the dialogue system described above in the domain of exercise promotion. Study participants interacted with the system via an animated Embodied Conversational Agent interface (Figure 6), making their contributions to the dialogue via a touch screen monitor [42]. Three evaluation methods were used in parallel: 1) evaluation of dialogue transcripts by a health behavior change expert; 2) self-report assessments of each dialogue by participants following completion of an entire counseling session; and 3) self-report assessments of every dialogue utterance by retrospective review (playback) of each dialogue by participants.

5.1.1 Participants and Protocol

Eight participants, aged 21-41 (mean 30.4), 75% female, were recruited via an online advertisement: three were in preparation, three in action and two in maintenance relative to the stages of change for minimum recommended levels of physical activity [20].

Each participant conducted an initial conversation with the agent then filled out a summative evaluation questionnaire. They were then told to pretend that they had come back the next day to conduct a follow-up conversation with the counselor. Immediately following this second conversation, the participant filled out another summative evaluation questionnaire then conducted a retrospective review of the entire second conversation. In this review, they watched
a computer-controlled replay of their conversation that stopped after each turn and asked a series of questions. They then repeated this protocol for a third and final conversation.

5.1.2 Measures

An expert reviewer trained in Motivational Interviewing, but who was not involved in the design of the intervention, provided summative evaluations of text transcripts of nine conversations (three each from the participants in Preparation stage). She used the global ratings of Empathy and Motivational Interviewing Spirit (each on 7 point scales) from a standard evaluation instrument used to assess human counselor fidelity to Motivational Interviewing principles and practice [54].

The summative self-report measures by participants comprised 19 scale questions (range 1-7), assessing overall satisfaction (3 items, Cronbach’s alpha 0.88), agent empathy (3 items, alpha 0.82), Motivational Interviewing “spirit” (8 items, alpha 0.86), and relationship with the animated counselor (3 items, alpha 0.83).

During the retrospective review, following every agent utterance, participants were asked to rate the appropriateness of the utterance, the naturalness of the utterance, and their trust in the agent, all on 7 point scales. Following every replay of a participant contribution to the dialogue, they were asked to rate how well their choice reflected what they actually wanted to say on a 7 point expressivity scale.

5.1.3 Results

Table 1 presents descriptive statistics of the study results, with retrospective results averaged over entire conversations. Expert ratings of agent empathy and Motivational Interviewing spirit were surprisingly high, and the expert provided insightful justification for her scores. The first conversation is primarily structured to provide patients with an orientation to the agent and intervention and to perform a stage-of-change assessment. The expert felt that this conversation did not elicit much information from the study participants (also reducing empathic opportunities), resulting in its relatively low score. By the third conversation, however, the expert felt that the conversation was being tailored to both the participant’s current needs as well as to things talked about in the prior sessions, resulting in higher scores.
### Table 1. Exercise Intervention Evaluation Ratings (mean (SD))

All 7-point Scales with 7.0 Best

<table>
<thead>
<tr>
<th></th>
<th>Dialog 1</th>
<th>Dialog 2</th>
<th>Dialog 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Expert Summative</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agent Empathy</td>
<td>5.00 (0.00)</td>
<td>5.33 (1.15)</td>
<td>6.00 (1.00)</td>
</tr>
<tr>
<td>MI Spirit</td>
<td>4.33 (0.58)</td>
<td>5.00 (1.00)</td>
<td>6.00 (1.00)</td>
</tr>
<tr>
<td><strong>User Summative</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction</td>
<td>5.33 (1.38)</td>
<td>5.57 (1.07)</td>
<td>4.92 (1.43)</td>
</tr>
<tr>
<td>Agent Empathy</td>
<td>5.04 (1.34)</td>
<td>5.38 (1.03)</td>
<td>4.92 (1.43)</td>
</tr>
<tr>
<td>MI Spirit</td>
<td>4.84 (1.14)</td>
<td>5.43 (0.99)</td>
<td>4.55 (1.37)</td>
</tr>
<tr>
<td>Relationship</td>
<td>4.71 (1.29)</td>
<td>5.24 (0.76)</td>
<td>4.88 (1.18)</td>
</tr>
<tr>
<td><strong>User Retrospective</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriateness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naturalness</td>
<td>Not</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trust</td>
<td>measured</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expressivity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.40 (0.54)</td>
<td>6.35 (0.49)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.35 (0.67)</td>
<td>6.00 (0.73)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.13 (0.91)</td>
<td>6.01 (0.76)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.23 (0.66)</td>
<td>6.02 (0.79)</td>
<td></td>
</tr>
</tbody>
</table>

Overall ratings of agent appropriateness and participant expressivity in the retrospective review were very high, with less than 3% of agent utterances and less than 4% of participant utterances rated below the midpoint of the scales. However, this methodology did enable us to identify parts of the dialogue that had an especially negative impact on participants. For example, in one sequence, the agent gave appropriate empathic feedback to a participant for not feeling well, then immediately launched into social chat by saying “So, I hope your day is going well.”, which did not impress the participant (turn 7 in Figure 7).
5.2 Evaluation of Reuse
We extended the DTASK task models to promote fruit and vegetable consumption to demonstrate that the abstractions used reliably lend themselves to reuse. These extensions required an additional 4 tasks (98% reuse), 3 recipes (98% reuse), and 961 agent utterances (14% reuse), totaling an additional 24,096 lines of source code (22% reuse). This demonstrates that the high level task abstractions (recipes and goals encoding general counseling strategies) are highly reusable, and most of the effort in porting the system to a new health behavior involves authoring new agent utterances at the leaves of the task hierarchy. Extending the system to diet promotion took 9% of the calendar time and 4% of the person hours required to develop the initial exercise promotion system.

5.3 Evaluation of Fruit and Vegetable Promotion Intervention
We conducted a mixed-method, formative evaluation of the fruit and vegetable promotion dialogue system, using the same protocol and measures used in the exercise promotion system evaluation described in section 5.1 above.

5.3.1 Participants
Nine participants, aged 22-68 (mean 41), 78% male, were recruited via an online advertisement. With respect to the behavioral criteria of 5 or more half-cup servings of fruit and vegetables on five or more days of the week [17], two were in pre-contemplation, two were in contemplation, three were in preparation, one was in action and two were in maintenance.

5.3.2 Results
Table 2 presents descriptive statistics of the User Summative and Retrospective and Expert Summative results. These results are very close to those for the exercise promotion system, indicating that participants were as satisfied with the fruit and vegetable promotion system and
had as positive attitudes towards the automated counselor as they did with the exercise promotion system.

<table>
<thead>
<tr>
<th>Expert Summative</th>
<th>Dialog 1</th>
<th>Dialog 2</th>
<th>Dialog 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent Empathy</td>
<td>5.33 (0.58)</td>
<td>6.00 (0.00)</td>
<td>5.50 (0.71)</td>
</tr>
<tr>
<td>MI Spirit</td>
<td>4.00 (0.00)</td>
<td>5.00 (1.00)</td>
<td>6.00 (1.41)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>User Summative</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfaction</td>
<td>5.15 (1.19)</td>
<td>5.04 (1.53)</td>
<td></td>
</tr>
<tr>
<td>Agent Empathy</td>
<td>Not</td>
<td>4.22 (1.49)</td>
<td>4.79 (1.83)</td>
</tr>
<tr>
<td>MI Spirit</td>
<td>measured</td>
<td>4.65 (0.76)</td>
<td>4.73 (1.00)</td>
</tr>
<tr>
<td>Relationship</td>
<td>4.48 (0.80)</td>
<td>4.58 (1.32)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>User Retrospective</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Appropriateness</td>
<td>6.02 (0.55)</td>
<td>6.17 (0.76)</td>
<td></td>
</tr>
<tr>
<td>Naturalness</td>
<td>Not</td>
<td>6.02 (0.62)</td>
<td>5.97 (0.88)</td>
</tr>
<tr>
<td>Trust</td>
<td>measured</td>
<td>5.64 (0.91)</td>
<td>5.62 (0.88)</td>
</tr>
<tr>
<td>Expressivity</td>
<td>5.90 (0.59)</td>
<td>6.27 (0.66)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Fruit and Vegetable Intervention Evaluation Ratings (mean (SD))

6. Conclusion and Future Directions

We have demonstrated that health counseling dialogue can be modeled using representations that promote reuse, that study participants and expert health counselors are largely satisfied with the results and their fidelity to health behavior change theory, and that the representations do indeed lend themselves to rapid reuse.

The two behavioral interventions we implemented were intentionally selected to be similar at the most abstract levels; both are acquisition interventions, and both are shapeable behaviors, in which the frequency of the desired behavior is incrementally increased once a patient’s motivation to change has been established. Extending our models to a very different health behavior change problem, such as smoking cessation, would require much more extensive modifications to the models. Even for those problems, though, much could be reused; the overall structure of a counseling session, the structure of motivational interviewing dialogue, and many smaller dialogue fragments (such as greetings and social chat) would likely remain unchanged.

The exercise and fruit and vegetable interventions we have developed will be evaluated in a randomized 4-arm, 60-day trial to compare each intervention to a combined intervention in...
which both exercise and diet modifications are promoted, and with a non-intervention control group.

The DTTask dialogue engine, the LiteBody web-based Embodied Conversational Agent, and the ontology and task models for the exercise and fruit and vegetable behavioral interventions will all be released open source in 2010 [55]. We hope that these resources will, at a minimum, promote further discussion and development and reusable tools for building computerized behavioral interventions. The ultimate evaluation of this work will be the extent to which these tools are reused by other researchers and the extent to which they facilitate dissemination. Through the contributions of many researchers in an open source framework, we could eventually deploy systems that can be tailored on hundreds of variables rather than the handful most systems use today [31].

The continued development of ontologies for behavioral medicine is an important activity in its own right, as it can help to drive the many disparate theoretical camps to consensus on what works, or at least provide a framework in which theories can be more precisely compared and evaluated. As Lenert noted, the development of such formalisms helps achieve many goals, such as making the linkages between protocols and theories more explicit, and aiding in the evaluation of protocols and protocol components [26]. By extending these formalisms to the point at which they can actually drive the behavior of a computerized intervention, the subjectivity in mapping theories from behavioral medicine to interventions can be greatly reduced, improving our ability to compare and evaluate theories. We also hope that this work serves to promote collaboration within and across the fields of medical informatics, behavioral medicine, computational linguistics, and AI.

Acknowledgements

Thanks to Ken Baclawski, Dolphy Fernandez, Jeffrey Migneault, Zoe Chang, Tom Brown and Jenna Zaffini for their help in developing the system, to Juan Fernandez for running the evaluation study, and to Amaura Kemmerer for providing expert evaluation. This work was supported by NIH National Library of Medicine grant R21LM008995.

References


[29] National Center for Chronic Disease Prevention & Health Promotion. BRFSS trends data nationwide grouped by age, for no leisure time physical activity. In: Centers for Disease Control and Prevention; 2005.
[32] OWL Web Ontology Language Overview. Available at http://www.w3.org/TR/owl-features/
[40] OWL-Time Home page. Available at http://www.isi.edu/~hobbs/owl-time.html
[47] ECMAScript. Available at http://www.ecmascript.org/
Appendix - Transcripts of Sample Counseling Sessions

The following is a partial transcript of a sequence of 3 exercise promotion dialogues with a single patient, gathered during the evaluation study. These transcripts were among those evaluated by the expert counselor.

**First Conversation**

1-1.A: To talk to me, just touch one of the options on your right.
1-2.U: OK
1-5.A: Great! I am excited about working with you to improve your health.
1-6.U: Me too.
1-7.A: I am going to be your health advisor. My job is to help you set health goals and overcome obstacles, and I will also be giving you lots of tips.
1-8.U: OK.
1-9.A: You need to keep in mind that I am just a computer character with limited capabilities, so I hope you will understand if there are some things I cannot help you with.
1-10.U: OK.
1-11.A: So, how do you feel about talking to a computer character?
1-13.A: That is good to hear.
1-14.A: So, I would like to chat with you every day over the next month to see how you are doing.
1-15.U: OK.
1-16.A: Each day I will ask you some questions about your health.
1-17.U: OK.
1-18.A: I hope it does not become too repetitive, but it is important that I understand how your feelings about health change over time.
1-20.A: Do you currently exercise at least five days per week, for thirty minutes or more?
1-21.U: No
1-22.A: During the next thirty days, do you plan to begin exercising five or more days a week, for at least thirty minutes each time?
1-23.U: Yes!
1-24.A: Cool. We'll talk more about exercise the next time we chat.
1-25.U: OK
1-27.A: So, when do you want to get together next?
1-28.U: Tomorrow
1-29.A: Okay, take care.

**Second Conversation**

2-1.A: Hi Elissa. Nice to see you again!
2-2.U: Hi, Karen. Nice to see you too.
2-3.A: How are you?
2-4.U: I'm doing okay.
2-5.A: Good.
2-6.A: So, Do you have anything fun planned for the weekend?
2-7.U: No, not really. And you?
2-8.A: I don’t have any special plans.
2-9.A: So, how's the weather out there today?
2-10.U: It's pretty good.
2-11.A: Great! I hope you can get outside and enjoy the weather while doing some nice, refreshing exercise.
2-12.U: Ok
2-13.A: So, Elissa, I would like to ask you a few questions about exercise.
2-15.A: How motivated are you to exercise, on a scale from 1 to 10?
2-16.U: 4 or 5
2-17.A: How confident are you that you can exercise, on a scale from 1 to 10?
2-18.U: 2 or 3
2-19.A: How are you feeling about your progress with exercise?
2-20.U: I think I'm having problems.
2-21.A: So when I asked you how important exercise is to you, you gave it about 5 out of 10.
2-22.A: What do you think is important about exercise that made you say 5 instead of zero?
2-23.U: If I exercised, things would be better.
2-24.A: Can you tell me what you think would be better if you started an exercise program?
2-25.U: Maybe I would lose weight.
2-26.A: So if you started an exercise program, you would lose weight.
2-27.U: I could also improve my health.
2-28.A: So if you exercised more, your health would be better.
2-29.U: Yes.
2-30.A: Okay. Let me just summarize what we've been talking about.
2-31.U: OK.
2-32.A: If you did start exercising, it might improve your health.
2-33.U: Yes.
2-34.A: If you started exercising, you could lose weight.
2-35.U: I might.
2-36.A: Okay. Thank you for telling me your thoughts about exercise.
2-38.A: So starting today, we will make a goal to get a little more exercise every day.
2-39.U: Okay
2-40.A: So, before the next time we talk, would you be willing to try and do some walking?
2-41.U: How much should I do?
2-42.A: You don't have to walk fast or for a long distance. The idea right now is just to get moving a little.
2-43.U: Okay, I can do that.
2-44.A: I think we are going to do some great work together.
2-45.A: I am going to assign you some exercise homework to do. Just something simple to help you start thinking about different things you can do to help you achieve your exercise goals.
2-46.U: Sure
2-47.A: When you don't want to exercise, think about what you would normally do. If that action is anything but exercise, try substituting a new action in place of the old one.
2-49.A: If you feel tired, you could try going for a walk. It can be energizing and very refreshing.
2-51.A: So Elissa, how about trying this for next time?
2-52.U: OK
2-53.A: Great, I think this will help you explore exercise in a different way and will help you to increase your step count a little faster.
2-54.U: Okay.
2-55.A: Great! We've talked about everything for today.
2-56.A: So, I hope we can see each other soon. When can we chat again?
2-57.U: Tomorrow
2-58.A: Okay then, I will be here. Just come by any time.
2-60.U: Bye Karen.

3-1.A: Hi Elissa. Nice to see you again!
3-2.U: Hi, Karen. Nice to see you too.
3-3.A: How is it going?
3-4.U: Good. How are you?
3-5.A: Same as usual.
3-6.A: So, How was is your day going? I hope it is going well.
3-7.U: Sorry, I didn't quite catch that.
3-8.A: So, How was is your day going? I hope it is going well.
3-9.U: Yes!
3-10.A: Really? What was it like?
3-11.U: Something fun...
3-13.A: So, how's the weather out there today?
3-15.A: Great! I hope you can get outside and enjoy the weather while doing some nice, refreshing exercise.
3-16.U: Ok
3-17.A: Last time I suggested that you think about substituting healthy actions for unhealthy ones. Did you do this?
3-18.U: Yes
3-19.A: Was this assignment useful in thinking about exercise in your life?
3-20.U: Yes
3-21.A: Good, you are making real progress!
3-22.A: Last time, we talked about trying to do a little walking.
3-23.U: Yes
3-24.A: Did you have a chance to go walking since our last conversation?
3-25.U: No
3-26.A: Sorry to hear that. I'm sure you'll manage it next time.
3.27. A: OK, Elissa, can we talk a little about exercise?
3.28. U: Sure.
3.29. A: How motivated are you to exercise, on a scale from 1 to 10?
3.30. U: 4 or 5
3.31. A: How confident are you that you can exercise, on a scale from 1 to 10?
3.32. U: 2 or 3
3.33. A: Are you feeling good about exercise?
3.34. U: I'd like to talk about it.
3.35. A: Is there anything that worries you about your exercise or starting to exercise?
3.36. U: It's easier if I don't exercise.
3.37. A: Tell me more. What do you like about your current lifestyle, without a regular exercise program?