

Embodied Agents for Long-Term Discourse

A dissertation presented
by

Daniel Schulman

to the Faculty of the Graduate School
of the College of Computer and Information Science
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Northeastern University
Boston, Massachusetts

January, 2013

Abstract

An Embodied Conversation Agent (ECA) is a computer interface designed to simulate human face-to-face conversation with its users, through the production of synthesized or prerecorded speech combined with an humanoid embodiment: a representation, either physical or visual, of a body capable of producing some portion of the nonverbal behaviors associated with speech, such as mouth movements, eye movements, head movements, hand gestures, facial expressions, and body posture. A key research problem in the design and implementation of realistic ECAs is generating the range of verbal and nonverbal behavior present in human conversation with appropriate frequency, timing, and quality.

ECAs have been used in a variety of applications, motivated by their potential to leverage the affordances of face-to-face conversation to build trust and engagement with users, and their learnability given their use of universally understood communicative cues. Many applications, including education and counseling, are examples of long-term interaction; where an effective agent must have many conversations, over a long period of time, while building rapport with its users. However, prior work on realistic conversational behavior for ECAs has focused heavily on single conversations, isolated from any larger context. To the extent that human conversational behavior is not fixed and unchanging across multiple conversations with the same conversation partner, this approach risks producing ECAs with behavior that becomes increasingly unrealistic in long-term interaction.

In this thesis, I present an approach to designing ECAs with realistic verbal

and nonverbal behavior in long-term interaction. Based on a longitudinal corpus of health behavior change counseling dialogue, containing multiple conversations between several counselor-client dyads, I construct a series of statistical models demonstrating systematic changes in human conversational behavior across multiple conversations; these changes are predicted both by the interaction history of a dyad, and by the strength or quality of their interpersonal relationship. Based on these findings, I present a model and implementation of verbal and nonverbal behavior generation for ECAs which reproduces some of the observed behavior patterns. Finally, I present a longitudinal randomized controlled evaluation study demonstrating that the resulting model of behavior generation, implemented in an ECA that acts as a virtual health behavior change counselor, produces measurable improvements in user-agent interpersonal bond in long-term interaction.

Acknowledgments

Contents

Abstract	i
Acknowledgments	iii
Contents	v
List of Figures	xiii
1 Introduction	1
2 Background and Related Work	7
2.1 Social Reactions and Human-like Behavior in Computer Inter- faces	7
2.2 Human Behavior in Long-Term Interaction	9
2.2.1 Comparison of Friends to Strangers	9
2.2.2 Greetings and Farewells	11
2.2.3 Indicators of Interpersonal Relationship	12
2.2.4 Coordination and Rapport	12
2.2.5 Politeness and Familiarity	15
2.3 ECAs Designed for Long-Term Interaction	16
2.3.1 Agents with Computational Models of Affect and Rela- tionship	17
2.3.2 Agents with Behavior Designed to Promote Long-Term Interaction	18

2.3.3	Agents with Varying Linguistic Style	19
2.4	Conclusions	19
3	The Exercise Counseling Corpus	23
3.1	Procedure	25
3.2	Measures	27
3.2.1	Trait Measures	28
3.2.2	Therapeutic Alliance	28
3.2.3	Trust	29
3.2.4	Task Outcome Measures	30
3.2.5	Cognitive Inventory	31
3.3	The Corpus	32
3.4	Descriptive Summary of Measures	33
3.4.1	Trait Measures	33
3.4.2	Therapeutic Alliance	34
3.4.3	Task Outcome Measures	36
3.4.4	Behavioral Measures	38
3.4.5	Cognitive Inventory	38
3.5	Correlations Between the Measures	39
3.5.1	Therapeutic Alliance	40
3.5.2	Therapeutic Alliance and Conversation Duration . .	43
3.5.3	Decisional Balance	46
3.5.4	Stage of Change	52
3.6	Discussion	56
3.6.1	The Size and Scope of the Corpus	57
3.6.2	Characterizing the Corpus by Self-Report Assessments	58
4	Speaking Rates and Coordination	61
4.1	Related Work	61
4.2	A Model of Articulation Rates	62
4.2.1	Results	63

4.3	Discussion	67
5	Posture and Discourse Structure	69
5.1	Background and Related Work	70
5.2	Coding	70
5.2.1	Coding of Topic Shifts	71
5.2.2	Coding of Posture Shifts	71
5.2.3	Results	72
5.3	A Model of Posture Shifts	72
5.4	Discussion	76
6	Openings and Reopenings	79
6.1	Coding	80
6.1.1	Outcome Variables	80
6.1.2	Development of a Coding Manual	82
6.1.3	Description	82
6.1.4	Interrater Reliability	84
6.1.5	Descriptive Results	85
6.2	Models of Behavior in Conversation Openings	85
6.2.1	Gaze	90
6.2.2	Mouth	90
6.2.3	Nodding	93
6.2.4	Speech	93
6.2.5	Self-Adaptors	97
6.2.6	Hand Gesture	97
6.2.7	Eyebrow Movement	97
6.3	Discussion	100
7	Rhythm: Behavior Generation for Long-Term Interaction	105
7.1	Introduction	105
7.2	Background and Related Work	107

7.3	Implementing Longitudinal Behavior Changes	108
7.4	Architecture	110
7.4.1	Analysis	111
7.4.2	Behavior Generation	113
7.4.3	Behavior Adjustment	114
7.4.4	Behavior Filtering	115
7.5	Examples	115
8	Evaluation	119
8.1	An ECA for Long-Term Interaction	121
8.2	A Preliminary Study of the Effect of Behavior Changes . . .	122
8.2.1	Apparatus and Measures	123
8.2.2	Procedure	124
8.2.3	Participants	124
8.2.4	Results	124
8.2.5	Discussion	125
8.3	A Long-Term Interaction Scenario	125
8.3.1	Behavior Change Intervention	126
8.3.2	Interaction Schedule	129
8.4	Methods	132
8.4.1	Measures	133
8.4.2	Procedure	135
8.5	Hypotheses	137
8.6	Results	139
8.6.1	Screening and Intake	139
8.6.2	Participants and Demographics	140
8.6.3	Baseline Assessments	141
8.6.4	System Usage	142
8.6.5	Therapeutic Alliance	144
8.6.6	Perceived Behavior and Immediacy	146

8.6.7	Behavioral Realism (Gestalt)	148
8.6.8	Behavioral Realism (Specific Behaviors)	150
8.6.9	Behavioral Realism (Combined Behaviors)	157
8.6.10	Attitudes toward Physical Activity	160
8.6.11	Self-Reported Walking	161
8.7	Discussion	166
8.7.1	Agent Behavior and User Engagement	167
8.7.2	Exaggerated Behavior and Realism	168
8.7.3	Engagement and Behavior Change	169
9	Conclusions	171
9.1	Research Contributions	172
9.1.1	A Methodology for the Study of Conversational Behavior in Long-Term Interaction	173
9.1.2	Patterns of Conversational Behavior in Long-Term Interaction	174
9.1.3	An ECA with Realistic Behavior in Long-Term Interaction	175
9.2	Future Research: Toward Rich Models of Behavior in Long-Term Interaction	175
9.3	Future Research: Toward Rich Models of Conversational Behavior in Health Behavior Change	179
9.4	Future Research: Toward an Integrated Model of Verbal and Nonverbal Behavior	180
9.5	Closing	180
	Bibliography	183
A	Experimental Protocol for the Collection of the Exercise Counseling Corpus	207
A.1	Initial Session	207

A.2	Subsequent Sessions	212
B	Coding Manual for Discourse Structure	235
B.1	Introduction	235
B.2	Identifying Segment Boundaries	235
B.3	Examples	237
B.4	General Suggestions	237
B.5	Difficult or Special Cases	237
B.5.1	Greetings	237
B.5.2	Topic Shifts over Multiple Phrases	238
B.5.3	Subtopics	238
B.5.4	Lengthy Narrative	239
B.5.5	Vague or Open-Ended Questions	239
C	Coding Manual for Posture Shifts	241
C.1	Introduction	241
C.2	Identifying Posture Shifts	242
C.3	Coding Features of Shifts	242
C.3.1	Timing	242
C.3.2	Movement Pattern	243
C.3.3	Direction	243
C.3.4	Energy	244
C.3.5	Co-occurrence	244
C.4	General Suggestions	245
C.5	Difficult or Special Cases	245
C.5.1	Small Movements	245
C.5.2	Repetitive Movement	246
C.5.3	Co-occurring Actions	246
C.5.4	Single versus Multiple Shifts	247
C.5.5	Conversational Openings	247
C.6	Changes from Earlier Drafts	247

D Coding Manual for Nonverbal Behavior in Openings	249
D.1 Introduction	249
D.2 Materials and Coding Task	250
D.3 Visibility	250
D.4 Nonverbal Behavior	251
D.4.1 Gaze Direction	252
D.4.2 Eyebrows	252
D.4.3 Head Movement	253
D.4.4 Mouth Shape	254
D.4.5 Hand Gesture	255
D.4.6 Shoulder Raising	258
D.4.7 Arm Openness	258
E Experimental Protocol for the Evaluation Study	261
E.1 Experiment Sessions	261
E.1.1 Intake Session	261
E.1.2 Weekly Sessions	262
E.1.3 Post-Intervention Sessions	263
E.2 Email Messaging	264
F Sample Conversation Transcripts for the Evaluation Study	299
F.1 Initial Conversations	299
F.1.1 Initial Motivational Interviewing	300
F.1.2 Initial Goal Setting	304
F.2 Routine Conversations	307
F.2.1 Motivational Interviewing	308
F.2.2 Motivational Interviewing, and Goal Setting	310
F.2.3 Baseline Goal Setting	314
F.2.4 Incremental Goal Setting	316
F.3 Final Sessions	318
F.3.1 Final Motivational Interviewing	318

F.3.2	Final Goal Setting	323
-------	------------------------------	-----

List of Figures

2.1	Relative importance of different components of rapport.	13
3.1	A counselor-client conversation in the Exercise Counseling Corpus.	26
3.2	A participant completing a retrospective review of a conversation.	30
3.3	Durations of conversations, shown separately by dyad.	33
3.4	Personality assessments of clients at the time of intake.	34
3.5	Attachment style assessment of clients at the time of intake. . .	35
3.6	Therapeutic alliance, self-reported by clients and counselor. . . .	36
3.7	Stage of Change for physical activity, self-reported by clients. .	37
3.8	Decisional Balance for physical activity, self-reported by clients.	38
3.9	Correlations between overall therapeutic alliance assessments. .	41
3.10	Correlations between therapeutic alliance goal assessments. . . .	42
3.11	Correlations between therapeutic alliance task assessments. . . .	42
3.12	Correlations between therapeutic alliance bond assessments. . .	43
3.13	Correlations between conversation durations and client overall therapeutic alliance assessments.	45
3.14	Correlations between conversation durations and counselor overall therapeutic alliance assessments.	45
3.15	Correlations between assessments of decisional balance.	47
3.16	Correlations between client assessments of therapeutic alliance and decisional balance pros.	47
3.17	Correlations between counselor assessments of therapeutic alliance and client assessments of decisional balance pros.	49

3.18	Correlations between client assessments of therapeutic alliance and decisional balance cons.	49
3.19	Correlations between counselor assessments of therapeutic alliance and client assessments of decisional balance cons.	51
3.20	Correlations between stage of change assessments and decisional balance pros.	53
3.21	Correlations between stage of change assessments and decisional balance cons.	53
3.22	Correlations between stage of change assessments and client assessments of therapeutic alliance.	55
3.23	Correlations between stage of change assessments and counselor assessments of therapeutic alliance.	55
4.1	A model of the normalized articulation rates of words.	64
4.2	Predicated changes in articulation rate.	66
5.1	A model of the occurrence of posture shifts.	74
5.2	Predicted changes in rate of posture shifts for a client.	75
5.3	Predicted changes in rate of posture shifts for the counselor.	75
6.1	Model A of the proportion of time the counselor or client perform a behavior.	87
6.2	Models B, C, and D of the proportion of time the counselor or client perform a behavior.	88
6.3	Model A of the number of times the counselor or client gazes away during speech.	89
6.4	Gaze-aways during speech, by session.	90
6.5	Occurrence of non-neutral mouth positions, by session.	93
6.6	Nodding when not speaking, by session and alliance.	95
6.7	Proportion of time speaking, by session and role.	95
6.8	Occurrence of self-adaptors, by session and role.	97

6.9	Occurrence of hand gestures, by session and role.	100
7.1	The SAIBA framework for multimodal behavior generation . . .	106
7.2	The Rhythm behavior generation pipeline.	112
7.3	Sample behavior generation for a first conversation with high therapeutic alliance	116
7.4	Sample behavior generation for a fifth conversation with low therapeutic alliance	116
8.1	The Litebody web-based ECA.	122
8.2	A fragment of a sample user-agent conversation, showing Motivational Interviewing-inspired content.	127
8.3	A fragment of a sample user-agent conversation, showing goal-setting content.	128
8.4	Decisional Balance at intake.	142
8.5	Proportion of participants completing a weekly conversation, by week and by condition.	143
8.6	Working Alliance Inventory (Short Revised) scores, by week and study condition.	144
8.7	Self-Report of Immediacy Behavior (SRIB) scores, by week and by study condition.	147
8.8	Gestalt behavioral realism scores, by study condition and by week.	148
8.9	The proportion of participants reporting the agent gazes at them “too little” or “too much,” by study condition and week.	149
8.10	The proportion of participants reporting the agent shifts her posture “too little” or “too much,” by study condition and week.	151
8.11	The proportion of participants reporting that the agent smiles “too little” or “too much,” by study condition and week.	153
8.12	The proportion of participants reporting that the agent nods “too little” or “too much,” by study condition and week.	155

8.13	The proportion of participants reporting that the agent's speech was "too fast" or "too slow," by study condition and week. . . .	156
8.14	Distribution of behavior items rated as realistic, by condition and week.	158
8.15	The proportion of participants who report negative, positive, or no progress on the Stages of Change.	160
8.16	Self-reported days of planned brisk walking per week, aggregated by study condition.	162
8.17	Negotiated goals for days of walking per week, aggregated by study condition.	163
8.18	The difference between self-reported days of walking and negotiated goals per week, aggregated by study condition.	165

CHAPTER 1

Introduction

Consider the problem of designing a computer interface for an application such as promoting positive health behavior changes (e.g., increased physical activity or improved diet) in its users. Such behavior change interventions typically take months or longer to have sustained effectiveness [133], and a computer-based intervention will not be effective unless users are willing to voluntarily interact with it over long time periods. An interface should promote strong user engagement, both to encourage continued usage, and to more directly improve outcomes, where long-term engagement) of a user with a computer interface is defined (following [15]) as the degree of involvement — including continuity, regularity, and depth of interaction — that a user chooses to have with a system over time.

The development of computer-based interventions for the promotion of positive behaviors, and for other behavioral health concerns, has become a large and growing area of research [138]. Major ongoing public health problems, such as obesity [59], have a behavioral component, and improvements in efficacy of interventions that target such behavioral health concerns — even modest improvements — can potentially have a large real-world impact.

The topic of this dissertation, broadly, is the design of computer interfaces for applications that require long-term, voluntary interaction, and strong user engagement. More specifically, I focus on promoting long-term engagement by simulating some of the behaviors humans might use in long-term face-to-face

conversational interaction — modeling, for example, a computer-based health intervention on a human health counselor — and I focus on a particular type of interface that I argue is well-suited for such applications: an Embodied Conversation Agent. While health behavior change intervention serves as the primary example within this thesis, other applications may have similar features, including the broader category of counseling applications (e.g., psychotherapy), other applications related to long-term health (e.g., chronic condition monitoring and management), education and tutoring, and interactive entertainment.

An Embodied Conversational Agent (ECA) is a computer interface designed to interact with its users via simulated face-to-face conversation [37]. They are a subset of the larger category of conversational agents or dialogue systems: interfaces that interact with users through some form of conversation, typically natural language, whether spoken or textual. Compared to other dialogue systems, ECAs are distinguished by their ability to communicate using nonverbal behaviors as well as verbal, including hand gestures, head movements, eye gaze, body posture, and facial expressions. An ECA must have an embodiment, or a representation, whether physical (e.g., a robot) or virtual (e.g., an animated character). A typical ECA has an embodiment with an anthropomorphic appearance, and communicates using synthesized or prerecorded speech, along with synchronized nonverbal behavior produced using this embodiment.

An ECA can potentially take advantage of richer communication channels than dialogue systems that lack nonverbal behavior. In face-to-face conversation between humans, nonverbal behavior has numerous functions, such as communication of emotions and affect [110], turn-taking and other conversation management [1], establishing and maintaining mutual knowledge (grounding) [114], and communicating and negotiating relationship status. A key research problem in the design of an ECA is the generation of behavior, both verbal and nonverbal, that appears realistic and human-like to users, and

makes effective and appropriate use of this rich communication channel.

ECAs have been studied in a wide variety of applications: as virtual counselors for health behavior change [13], as virtual patients for doctors or counselors [87], as virtual tutors [140, 84], as tour guides [16], as participants in interactive dramas designed for entertainment [106, 44] or for education [7, 109], and for long-term social engagement and companionship [172, 127].

Many of these applications are examples of long-term, voluntary interaction, in which an ECA may be required to have multiple, perhaps many, conversations with a single user. For instance, a virtual counselor to aid with smoking cessation may have hundreds of conversations with a smoker over months or years. An intelligent tutoring system may benefit from engaging its users in multiple sessions of tutoring, interspersed with homework or other forms of learning. An interactive drama can be more complex when characters have multiple interactions with a user.

My goal is to extend previous work on the generation of realistic, human-like verbal and nonverbal behavior in ECAs into the domain of long-term interaction and what I will call *multi-conversation discourse*: interaction consisting of multiple conversations between the same participants (human or agent), over some period of time, all intended to accomplish parts of the same task. I attempt to examine, model, and reproduce ways that verbal and nonverbal behavior may change across multiple conversations. This may include both systematic changes over time — as participants develop a mutual interaction history — and changes which are associated with changes and development of the interpersonal relationship between conversation partners.

It is useful to distinguish two different types of changes which may be studied: *systematic* changes, which are those changes in verbal and nonverbal behavior that would occur for all or most dyads given a similar interaction history and interpersonal relationship, and the development of *idiosyncratic* behavior by a dyad over time. This idiosyncratic behavior may be most easily understood as mutual learning: Conversation partners, during repeated inter-

action, learn about each other, and this may lead to changes in verbal and nonverbal behavior over time. They may engage in grounding [46], building common ground, or mutual knowledge. Their behavior may change as a result of this learning, from aligning their lexical choices and other linguistic choices with each other [125] to developing a “language” built of references to their shared relationship history [54]. As an initial investigation, my focus is on systematic rather than idiosyncratic changes, as being the most easily amenable to reproduction in conversation with an ECA.

The majority of prior work on the generation of verbal and nonverbal behavior in ECAs is based on the examination of single conversations, isolated from a larger context, typically involving a previously unacquainted dyad. An ECA based solely on this work will exhibit similar behavior on its first conversation with a user as on its second conversation, or its tenth, or hundredth. In this work I will attempt to answer research questions this raises: What systematic changes occur in verbal and nonverbal behavior as human conversation partners engage in multiple interactions? And can we, by modeling and implementing these dynamic changes in an ECA, create conversational interfaces that appear more human-like and realistic, and are more engaging and efficacious in long-term interaction?

A predominant methodology for the design of ECA behavior is to collect examples of human behavior in which one or more behaviors of interest are observed; wherein the resulting data can be analyzed to produce a model that may be implemented in an ECA. I develop an extension of this methodology in which ECA behavior is designed based on a corpus containing examples of *multi-conversation discourse* between participants, in the context of a developing interpersonal relationship.

Thesis Statement When humans interact in multi-conversation discourse, their verbal and nonverbal behavior changes in systematic ways, both over time and in the context of changes in their interpersonal relationship. Mod-

eling these behaviors and implementing them within an Embodied Conversational Agent designed for long-term interaction will increase user engagement with the agent, and will benefit the user’s perceptions of the agent as realistic and human-like.

In the remainder of this dissertation, following a review of related work, I describe the “Exercise Counseling Corpus”: a longitudinal video corpus of face-to-face conversation designed for the investigation of verbal and nonverbal behavior in long-term interaction. Based on this corpus, I construct a series of models that show systematic changes in verbal and nonverbal behavior across multiple conversations. I describe an implementation of these models in an Embodied Conversational Agent. Finally, this implementation is evaluated in a longitudinal randomized controlled trial: I test its effect on user-agent interpersonal bond — a construct analogous to client-counselor interpersonal bond in interaction with a human counselor, where it is a mediator of positive outcomes [104] — and demonstrate that this implementation produces significant improvements in interpersonal bond relative to a baseline implementation which does not incorporate changes over time.

CHAPTER 2

Background and Related Work

2.1 Social Reactions and Human-like Behavior in Computer Interfaces

The Computers as Social Actors theory predicts that social effects which occur in human-human interpersonal interaction can also occur in human-computer interaction [137, 117]. Nass and colleagues propose that these effects are due to unconscious and automatic reactions to cues produced by the computer which resemble the social cues that normally function to manage human-human interpersonal interaction [119]. They introduced a methodology in which effects predicted to occur in human-human interaction are replicated when one of the participants is replaced by a computer. For example, users were shown to give more positive evaluations of a computer system when the computer requested an evaluation of itself, compared to when the evaluation was requested by a different computer [118]. This resembles effects observed in human-human interaction; people apply “politeness strategies” to avoid offending their interaction partner.

Since verbal and nonverbal behavior in conversations carry a multitude of potential social cues, it is expected that an ECA can cause social effects in its users similar to those caused by a human conversation partner, to the extent that the ECA’s behavior resembles human behavior. Several researchers have

demonstrated such effects: Hoffmann et al. demonstrated that a politeness response (similar to that cited above) can occur when a user is asked to evaluate an ECA by the agent itself [77]. Rickenberg and Reeves showed that an ECA that appeared to monitor users during a task caused increased anxiety and decreased task performance, as would be expected if users were monitored by a human [141]. Conversational agents have been demonstrated to elicit effects of perceived ethnicity [99, 113, 11], gender [117, 11], and personality [117] similar to effects observed in human-human interaction.

In some cases, an ECA has been shown to produce more pronounced social responses than a corresponding non-embodied interface. Krämer found that users reported greater willingness to interact with a TV/VCR system using natural language, and were more likely to use polite forms of address, if the system presented an embodied interface rather than audio-based or text-based interfaces [91]. Hone reported that empathic messages may be more effective at reducing user frustration when delivered by an embodied agent than a non-embodied interface [78].

However, overall the evidence for a differential effect of an ECA compared to non-embodied interfaces is mixed. Dehn and van Mulken conducted a systematic review (which predates the studies cited here) of the effects of embodiment on user attitudes and behaviors across a variety of application domains. They argued that the available evidence did not support a conclusion that the presence of an embodiment had an effect on user behavior, and that the evidence suggested that the effect of an embodiment on user attitudes was dependent on the specific embodiment and the specific application domain [49]. Similarly, Yee et al. conducted a meta-analysis and demonstrated that the average effect size of embodiment was small, particularly when examining behavioral rather than subjective outcomes [173].

These results suggest that an ECA that displays human-like verbal and nonverbal behavior will elicit user responses similar to responses toward the same behavior in a human conversation partner. However, in general it may

not be safe to assume that a particular ECA will elicit a particular response, for at least three reasons. First, the effects of an embodiment may be small or inconsistent, as argued above. Second, an ECA is unlikely to reproduce human conversational behavior perfectly, and thus may produce attenuated or nonexistent social effects. Finally, some social responses to an agent may be moderated by users' beliefs about the agent and the nature of its agency, even in the absence of any observable differences in behavior.

This last point is demonstrated by several experiments which have explored the effects of manipulating user beliefs about an agent, by informing users that an agent is either autonomous or controlled by a computer, or a human-controlled avatar — with no observable differences in behavior. Shechtman reported that participants were more likely to respond to behavior by a computer that impacted their interpersonal goals (e.g., displaying frustration if the computer appeared to aggressively promote its own decisions) if they believed it to be a human [150, 151]. Bailenson et al. reported that participants who interacted with a virtual character in a virtual reality setting maintained a smaller interpersonal distance if they believed the character to be a human-controlled avatar [8].

2.2 Human Behavior in Long-Term Interaction

2.2.1 Comparison of Friends to Strangers

A simple approach to investigating behavior in long-term interaction is to compare conversations between friends — participants with a long-term interpersonal relationship — to conversations between strangers or acquaintances, who have little or no prior history of interaction or interpersonal relationship. Any observed differences in conversational behavior may be an indication of changes that take place over multiple interactions, or of behaviors that are as-

sociated with differences in interpersonal relationship other than interaction history.

Studies that take this approach have typically been cross-sectional, comparing, for example, pairs of friends to *different* pairs of strangers. This leads to a caveat in the findings reported in this section, as cross-sectional studies have limited ability to distinguish differences between participants (or pairs of participants, here) from changes over time. In this case, an observed difference in conversational behavior between friends and strangers could be explained as either a behavior that changes as people become friends, or a difference that predicts whether people are likely to become friends and continue interacting, or a combination of the two.

Planalp and Benson demonstrated that observers were able to discriminate (with 79% accuracy) audiotaped conversations between friends from conversations between acquaintances. When asked what cues they used, observers reported that friends referred more often to mutual knowledge, showed higher content intimacy, sounded more relaxed, interrupted each other more often, and had more equal distribution of floor time [129]. In a second study, this set of cues was adapted into a coding system. References to mutual knowledge and references to continuity (past and future conversations) had the strongest predictive power [128]. As this dissertation is not primarily focused on the content of conversation, these features are not further examined in this dissertation; however, the use of both mutual knowledge and references to continuity were explored in related work by Bickmore [13], which is discussed in further detail below (Section 2.3.2). Incorporating these findings with the work presented in this dissertation represents a future area of research.

Cassell et al. compared direction-giving dialogues between friends and between strangers [40]. Videotapes of these dialogues were transcribed, and coded for dialogue acts (using a subset of DAMSL [47]), gaze behavior, and head nods. Strangers used more explicit acknowledgments than friends when giving or receiving information. Strangers also used more nonverbal behavior

related to coordination: head nods and mutual gaze were more likely to occur during acknowledgments.

2.2.2 Greetings and Farewells

Greetings and farewells feature highly complex and ritualized nonverbal behavior, which is often specific to a particular culture and social situation [58]. Other behavior may be more universal: in most cultures recognition and friendliness are signaled with an “eyebrow flash”: a rapid raising and lowering of the eyebrows [55]. This behavior would be more likely to occur on the second and subsequent interactions of a multi-conversation discourse.

In investigating changes in behavior that occur across multiple conversations, the beginnings and ends of individual conversations may be useful to examine, as these portions of conversations often include statements that explicitly reference past and future interactions. For example, statements of continuity are often included in greetings (e.g., “Good to see you again”), and in preclosings [145] and farewells (e.g., “See you next week.”).

Riggio et al. examined nonverbal behavior during role-played greetings [142]. Participants were more likely to use handshakes, hugs, and taps on the shoulder or arm when greeting a role-played friend, compared to a role-played acquaintance. Greetings between “friends” were judged by observers as more intimate.

O’Leary and Gallois examined the last ten turns of conversations, and compared friends to strangers [122]. The end sections of conversations were distinguished by characteristic patterns of verbal behavior: information statements followed by summary statements (which the authors interpreted as ending the current conversational topic), and continuity and well-wishing statements (interpreted as signals that the relationship will continue beyond the conversation). End sections were also marked by nonverbal behavior, including mutual smiling and looking away from each other more often than in other

sections. Friends tended to look away more often than strangers, performed more grooming behaviors (e.g., brushing hair with fingers), and nodded less often.

2.2.3 Indicators of Interpersonal Relationship

Separate from the core research questions raised here — how verbal and non-verbal behaviors change over multiple conversations, as interpersonal relationship changes — a closely related set of research questions ask which features of verbal and nonverbal behaviors are indicators of various dimension of interpersonal relationship. Perhaps the most studied dimension is *immediacy* (described as intimacy, warmth, or closeness), which has been associated with a wide variety of signals including [5]: closer proximity, direct body orientation, forward leaning, touch, smiling, head nods, increased and more expressive gesture, relaxed and open body posture, eye contact and gaze, increased vocal expressiveness and variability, use of verbal backchannels (e.g., “mm-hmm”), and movement synchrony and congruence.

Burgoon and colleagues have studied nonverbal indicators of a large number of possible dimensions of interpersonal relationship, identifying at least 8 dimensions (composure, formality, dominance, equality, task-social orientation, and intimacy, with subdimensions of affection, similarity/depth, and receptivity/trust) [26, 27]. The associations between nonverbal behavior and these dimensions is complex, with each dimension typically being conveyed by a number of different cues [28].

2.2.4 Coordination and Rapport

In general, the examinations of relational indicators discussed above do not explicitly consider change over time, although some dimensions of relationship change in fairly predictable ways — for example, intimacy should generally be higher in previously acquainted dyads.

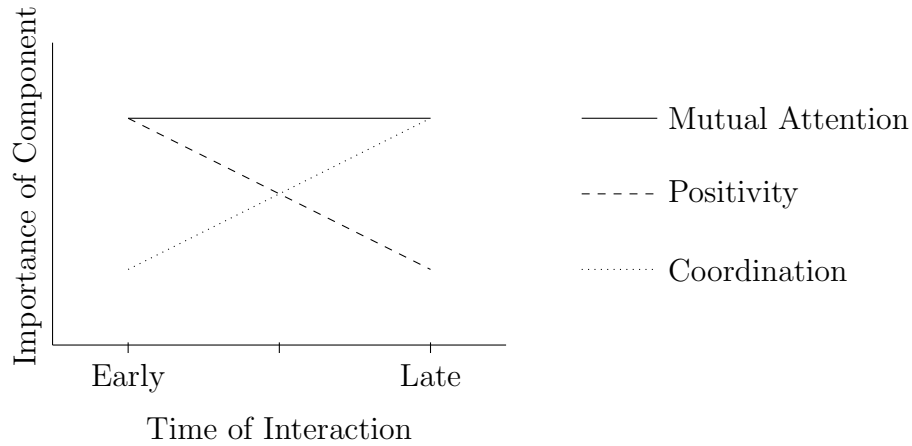


Figure 2.1: Relative importance of different components of rapport (from Tickle-Degnen and Rosenthal, 1990 [162])

A notable exception is the study of *rapport*. Tickle-Degnen and Rosenthal propose a model of rapport that deepens over time [161, 162], and consists of three components: 1. *mutual attentiveness*, or perceived interest by an interlocutor; 2. *positivity*, or mutual friendliness and caring; and 3. *coordination* in interaction. The relative importance of these components is predicted to vary throughout the course of a relationship, with coordination increasing and positivity decreasing (Figure 2.1).

Tickle-Degnen and Rosenthal also argue that each component of rapport can be observed via its correlates in nonverbal behavior. However, these correlates may be context-dependent, and difficult to analyze in terms of discrete behaviors. A meta-analysis [162] did indicate that a participants' evaluation of their partner's level of positivity was positively associated with the partners' nonverbal behaviors including forward trunk leaning, smiling, nodding, direct body orientation, and uncrossed arms.

In a small longitudinal study (6 dyads, 8 sessions each) of speech therapists, Tickle-Degnen and Gavett [160] show some support for a three-stage model of the development of rapport. In the first two stages (taken to be the first session, and session 2–4, respectively), observer ratings of positivity were positively correlated with the therapist's self-reported rapport with the client.

However, in the third stage (sessions 5–8), positivity was negatively correlated with rapport.

Several researchers have examined changes in the coordination component of rapport. Interactional synchrony — similarity in the rhythm and timing of behaviors — tends to increase within a single conversation, in cases where participants and/or observers judge the conversation as having proceeded smoothly and positively [169, 34]. However, interactions between strangers are often characterized by “tight” coordination — highly symmetric, regular, predictable patterns of interaction — whereas dyads with greater familiarity show “looser” coordination [33].

Some prior work has examined “nonconscious mimicry”, in which a person mimics or matches their conversation partners’ behavior without reporting awareness of this mimicry. This work has not explicitly examined change across multiple conversations, but in some cases has examined associations between nonconscious mimicry and various relational variables. Chartrand and Bargh presented a series of studies exploring the “chameleon effect”, or non-conscious mimicry of conversational behavior including facial expressions and posture [45]. They demonstrated that participants would mimic the behavior of confederates who varied their facial expressions and mannerisms (e.g., rubbing their face or tapping a foot): participants were observed, for example, rubbing their face significantly more times per minute when interacting with a confederate who also did so, and did not report awareness of the confederate’s behavior. A second study demonstrated a causal link between mimicry and liking: participants reported more liking of a confederate who mimicked their behavior compared to a confederate who did not use mimicry, and reported greater “smoothness” in the interaction with that confederate, again without reporting awareness of the confederate’s behavior. A third study demonstrated an association between mimicry and the cognitive aspects of empathy: participants high in perspective-taking showed more nonconscious mimicry of a confederate.

Similarly, Chartrand and Lakin [95] presented a pair of studies which demonstrated a causal link between a desire for affiliation and mimicry. Participants given a goal to increase affiliation, either through explicit instruction or through subliminal priming, displayed more mimicry of a confederate's face-touching behavior. In a second study, participants who were given an affiliation goal through subliminal priming and manipulated to perceive failure to build affiliation with a confederate showed more mimicry in a subsequent interaction, relative to participants who perceived they had succeeded in building affiliation.

2.2.5 Politeness and Familiarity

Brown and Levinson's theory of politeness [23] argues that speakers in conversation will modify the linguistic style of their speech acts, based on relational variables including the power of the hearer over the speaker, and the social distance (closeness) between the speaker and hearer. Speech acts in which the speaker is attempting to request something of the listener (e.g., "Go for a walk tomorrow.") can be made more polite by "hedging" (e.g., "Go for a walk tomorrow, maybe, if you want to.") or rephrasing a request to make it more indirect and ambiguous (e.g., "Tomorrow would be a good day to go for a walk.") The theory predicts that speakers will be more polite when addressing a listener with low familiarity or high power to the speaker. Since familiarity between conversants would generally increase over time, this theory predicts that politeness would systematically decrease across multiple conversations. Brown and Levinson focused on the linguistic content of speech acts; there has been less work examining nonverbal features and para-verbal features (e.g., prosody) in the context of politeness.

LaPlante and Ambady showed that observers rated speech as more polite when speakers were asked to use a positive tone of voice [97], but did not explore the features of speech that were identified as positive tone of voice.

Ambady et al. demonstrated that speakers delivering negative content (“bad news”) were rated by observers as more polite when observers viewed video-only recordings, compared to audio-only [4]. The authors argue that these results indicate that politeness is signaled partially by nonverbal cues, but they did not examine the specific cues involved.

Trees and Manusov examined the effect of nonverbal behavior on politeness strategies in female friendship dyads [164]. Participants either viewed a video of an interaction in which a speaker delivered a message with mitigating nonverbal behavior, a similar video with aggravating nonverbal behavior, or (as a control condition) read a transcript of a similar message. Mitigating behaviors included a pleasant facial expression (associated with affection or liking [24]), direct body orientation (associated with rapport [144]), touch and close proximity (associated with immediacy [5]), and a set of behaviors associated with submissiveness (lower power or influence) [25]: raised eyebrows, tense and closed posture, small gestures, and a softer voice. Conversely, aggravating behaviors included an unpleasant facial expression, non-direct body orientation, lack of touch, low proximity, lowered eyebrows, a relaxed and open posture, large gestures, and a loud voice. Aggravating behaviors caused participants to rate a message as less polite relative to a transcript, while mitigating behaviors caused speech to be rated as more polite, although with a smaller effect.

2.3 ECAs Designed for Long-Term Interaction

A number of researchers have explored the design of ECAs that interact with users over multiple conversations, ranging from a handful of interactions to hundreds of interactions spanning months or years.

2.3.1 Agents with Computational Models of Affect and Relationship

One approach to creating lifelike behavior in a virtual agent is through the creation of a computational model of affect (e.g., [66, 61]), which can produce an emotional response to an agent's interaction with a user, an environment, or another virtual agent. Such a response can influence an agent's verbal or nonverbal behavior. For example, certain facial expressions are characteristic of basic emotions [57], and an agent's affect model may produce these facial expressions. A model of affect may also include a concept of interpersonal relationship, which can produce varying behavior as relationships change.

Prendinger and Ishizuka present an agent that maintains a model of its social relationships with other agents [130], including dimensions of attitude [123] and a monotonically increasing dimension of familiarity [23]. This model is used to choose from a pre-defined set of responses to other agents. In the BASIC (believable, adaptive, socially intelligent character) model [143], an agent maintains memory of relationships with other agents, containing a record of emotional responses.

Kasap et al. [85, 102] discuss the design of "Eva," a virtual tutoring agent designed for repeated interaction. Eva maintains a relationship model for each user (based on Argyle's 2-dimensional model of friendliness and dominance [6]) as part of a larger affect model. Relationship influences behavior indirectly, by biasing the agent's mood, and is updated based on the emotional content of events during a session.

Cassell and Bickmore [39] discuss "Rea," an embodied conversational agent that maintains a dimensional model (depth of familiarity, breadth of familiarity, and solidarity) of its relationship with a user during a conversation. This model is updated based on the discourse history, with depth of familiarity dependent on the intimacy level of topics that have been discussed. This agent differs from others discussed here in that Rea actively tries to improve the

user-agent relationship by selecting conversational moves with an appropriate level of face threat.

2.3.2 Agents with Behavior Designed to Promote Long-Term Interaction

Bickmore introduced and explored the concept of “relational agents”: computer agents designed to form long-term social-emotional relationships with their users [13]. He created an ECA that used various conversational behavior intended to promote an interpersonal relationship with the user, including small talk [147], humor, empathetic messages, and reciprocal self-disclosure [3]. Bickmore’s relational agent implementation was focused on identifying particular messages which, when delivered appropriately during conversation, would promote a strong user-agent relationship.

Bickmore and colleagues have also explored the design space of ECA behaviors intended to support long-term interaction and engagement, using the novel methodology of a “virtual laboratory” in which an agent has daily interactions with a stable pool of participants for an indefinite period of time [15]. The agent’s behavior can be modified to easily test the effect of different interaction strategies, and self-report and behavioral data on participant engagement and outcomes can be collected. This system has been used to explore the effect of variability in the agent’s dialogue, and the effect of including a first-person or third-person backstory for the agent. To date, this group of studies does not directly examine differences in nonverbal behavior, but they do illustrate that small differences in agent behavior (e.g., lexical variability) can produce measurable differences in user engagement.

Gockley et. al. describe Valerie, an agent designed as a “Roboceptionist” who could give information to visitors [63]. To provide interest for repeat interactions, Valerie uses story fragments which are changed weekly; visitors must interact with her regularly in order to hear the full storyline. Later work

added additional time-varying behavior, such as daily mood variation [64].

The Companions Project [154, 127, 172] seeks to provide a “lifestyle companion” for senior citizens, capable of engaging users in intermittent, long-term interaction. The Companions will support activities such as discussing the users’ likes and dislikes, and discussing personal items such as photographs. The project focuses on fundamental research in dialogue management, such as the challenge of providing sufficient and appropriately relevant content for long-term interaction. Similarly, several researchers have explored the use of agents able to create episodic memories and refer to them in later conversation [152, 32, 85, 102].

2.3.3 Agents with Varying Linguistic Style

Walker et. al. describe a framework for generating various speech acts with appropriate degrees of politeness [168], based on Brown and Levison’s theory of politeness (Section 2.2.5).

Over the course of a dialogue, interlocutors tend to align their linguistic choices with each other on many levels [125]. Isard et al. describe a natural language generation system which can display linguistic alignment by choosing the surface forms of utterances based on an n-gram model dialogue history, among other factors such as personality [82]. Buschmeier et al. describe “SPUD prime,” a microplanner that generates natural language with alignment, based on a model of priming [29]. While alignment effects may plausibly occur across multiple conversations, to my knowledge all prior work on generation with alignment has been in the context of a single interaction.

2.4 Conclusions

There is extensive prior work on the production of realistic verbal and non-verbal behavior in Embodied Conversational Agents, and there are several examples of prior work exploring the use of ECAs in long-term interaction. I

find no prior work that combines these concerns and explores the production of realistic verbal and nonverbal behavior for ECAs in long-term interaction. Similarly, prior work on the effect of behavioral realism in ECAs does not adequately examine the context of long-term interaction.

Prior work on human behavior in face-to-face conversation is the primary source of information for efforts to produce realistic ECA behavior. Within this, there are several examples of prior work that explores associations between verbal and nonverbal behavior and various relational variables; to the extent that these variables are expected to change over the course of multiple face-to-face conversations, this work suggests ways that verbal and nonverbal behavior might change over time. I find, for the purposes of this dissertation, two major gaps in prior work: First, there are few examples of longitudinal studies of face-to-face conversation, exploring how behavior changes from one conversation to the next; the majority of prior work that examines change over time does so in a cross-sectional manner. Second, there are few examples of prior work that examines a causal effect of changes in verbal and nonverbal behavior. Causal inference is inherently difficult when studying human behavior in face-to-face conversation, as the most robust way to examine causal effects is through experimental manipulation, and attempts to manipulate a chosen set of behaviors (e.g., through the use of trained confederates [45]) can easily result in unintended changes in behavior as well.

This dissertation attempts to address these gaps in the research literature: I present a longitudinal observational study of human behavior in face-to-face conversation, providing information on human behavior in long-term interaction. I describe an approach and proof-of-concept implementation for producing ECA behavior based on the results of this study, addressing a lack of prior work on the production of realistic ECA verbal and nonverbal behavior in long-term interaction. Finally, I present a longitudinal randomized controlled trial of an ECA based on this work, addressing a lack of prior work on the effects of realistic ECA behavior in long-term interaction, as well as exploring

the causal effects of verbal and nonverbal behavior in long-term interaction.

CHAPTER 3

The Exercise Counseling Corpus

My approach to generating realistic and human-like behavior in an ECA follows previous work, suitably extended to the context of long-term interaction:

1. Collect a corpus of human-human dyadic face-to-face interaction, which is appropriate in the sense that it contains examples of human behavior closely analogous in several ways to the agent behaviors we are interested in generating including: a) long-term interaction, b) the development of interpersonal relationships, and c) a topic or goal of conversation relevant to an application domain of interest for ECAs.
2. Manually annotate various verbal and nonverbal behaviors in the corpus.
3. Create statistical models of the occurrence and features of these behaviors.
4. From these models, derive stochastic rules for the generation of behaviors, which can be implemented in a conversational agent.

In this chapter, I present the Exercise Counseling Corpus, a corpus of dyadic conversations designed for the study of verbal and nonverbal behavior over multiple conversations. The corpus consists of videotapes of conversations between a behavior change counselor (an exercise trainer) and clients, with up to six weekly conversations per dyad. It also includes a number of self-report

assessments, from both the counselor and client, the most significant (for the purposes of this dissertation) being self-reported measures of the counselor-client interpersonal relationship at each conversation.

The Exercise Counseling Corpus is a longitudinal corpus, containing multiple weekly conversations between each counselor-client dyad. A longitudinal corpus allows us to examine changes over time, and to separate them from differences between individuals. In contrast, a cross-sectional corpus — containing, for example, conversations between strangers and conversations between (different) previously acquainted dyads — would let us compare dyads with differing amounts of prior interaction, but would give us less ability to examine how differences in behavior appear and change over time.

Behavior change counseling was chosen as a promising real-world application for conversational agents (e.g. [13]), and an example of an application in which maintaining realistic and engaging behavior in long-term interaction with an agent is important for positive outcomes. Behavior change typically requires long-term interaction and multiple interactions in order to have long-term effectiveness [133]. In most cases, a user’s participation in behavior change counseling will be at least partly voluntary¹, so it is critical that his or her engagement and desire to continue participating be maintained.

Behavior change counseling is also an example of an application in which differences in the strength and quality of participants’ interpersonal relationship are both predictive of outcomes, and may be predictive of differences in conversational behavior. The counselor-client interpersonal relationship is well-studied, and a strong rapport between counselor and client is associated with positive outcomes [80, 104].² Well-defined constructs for describing and

¹Clients in the corpus collection study were compensated for their time (Section 3.1) and their participation was therefore not completely voluntary. Their engagement and desire to continue may have been influenced by this compensation, as may the counselor’s motivation to build and maintain engagement. However, I note that the compensation was not large and was partly balanced by requirements for clients to complete experimental tasks aside from participation in counseling (e.g., questionnaires and other assessments).

²The evidence that rapport and therapeutic alliance are predictive of positive outcomes is strongest when applied to psychotherapy, and this association is less well-studied as

assessing the counselor-client relationship have been developed, and extensively validated measures are available (see Section 3.2.2). There has been research interest in verbal and nonverbal behavior that may be related to the development of this relationship (e.g., [160]).

3.1 Procedure

A single counselor (a female graduate student who is a certified exercise trainer) completed all conversations. The counselor attempted to encourage each client to increase his or her daily activity. She was instructed to suggest brisk walking as an activity, as this was readily assessed with a pedometer. However, the counselor was allowed to suggest alternate or additional activities as she felt appropriate. It was suggested that sessions be brief (10–15 minutes), but otherwise allow the counselor to conduct the sessions as dictated by her training and experience. The full instructions given to the counselor are included in Appendix A.

Clients were recruited through flyers and classified ads. To ensure that the conversations represented a meaningful task, potential clients who were already exercising regularly were excluded (assessed by a Stage of Change questionnaire [103] at the time of recruitment). Clients were asked to complete six sessions at approximately one week intervals. They were told that they would be talking about physical activity, and that the counselor would try to encourage them to increase their daily physical activity, but otherwise were not informed of the content of the conversation. Clients were paid \$10 per session; payment was not contingent on completing all six sessions.

Both the clients and the counselor were informed that the conversations would be videotaped, and that the resulting videos would be examined afterwards. They were not told what specific behaviors would be examined, and were not informed of either specific hypotheses or general research questions.

applied to physical activity promotion specifically and to health behavior change in general.



Figure 3.1: A counselor-client conversation in the Exercise Counseling Corpus.

No attempt was made to disguise or obscure the recording equipment, and cameras and microphones were clearly visible to all participants.

All conversations were conducted in a small laboratory room. Two identical standard office chairs were placed facing each other, with a small, approximately knee-high table between them. The wheels of the chairs were locked so they could not easily be rolled, in order to keep them positioned in front of the cameras, but the chairs could be rotated freely. The client was seated first in every session, and was always seated in the chair facing the door.

Figure 3.1 shows an example frame of video from the corpus. The counselor is visible in the upper right and bottom left quadrants, while the client is visible in the upper left and bottom left (facing away from the camera, in the bottom left). The bottom right quadrant was blank in all recordings, as three cameras were used, which were then combined into a single recording using a

4-input video mixer.

Videotaping began when the counselor entered the room and sat in the facing chair, and continued until the conversation completed and the counselor left the room. The conversation was ended, in all sessions, by the participants, and the participants were not prompted to end the conversation in any way. The client remained in the room following the conversation in order to complete various assessments (Section 3.2), and the counselor and client had no interaction outside of the conversation.

3.2 Measures

The exercise counseling corpus is intended to be suitable for examining systematic changes over time in conversational behavior, including behavior that varies simply with the amount of prior interaction, and behavior that is also associated with changes in the attitude of the participants toward their interpersonal relationship, and the task of the conversation. These attitudes will likely vary over time, but cannot be assumed to have a systematic or simple linear association with the amount of prior interaction.

At various points in the collection of the corpus (Table 3.1), a number of measures of these attitudes and other possible predictors of conversational behavior were collected, including trait measures of the participants (Section 3.2.1), self-report measures of interpersonal relationship (Sections 3.2.2 and 3.2.3), and both self-report and behavioral measures of the task outcome (Section 3.2.4). While several measures were collected at every session, some were collected only on alternating sessions (Stage of Change and Decisional Balance); this choice was made in order to reduce the overall length and effort required for clients' participation.

Assessment	Session					
	1	2	3	4	5	6
Personality (BFI)	✓					
Attachment (ECR-R)	✓					
Stage of Change	✓					
Decisional Balance	✓					
Steps Walked (pedometer)		✓	✓	✓	✓	✓
Before Conversation						
After Conversation						
Therapeutic Alliance (WAI-SR)	✓	✓	✓	✓	✓	✓
Thought-Listing	✓	✓	✓	✓	✓	✓
Stage of Change		✓		✓		✓
Decisional Balance		✓		✓		✓
Trust	✓	✓	✓	✓	✓	✓

Table 3.1: Schedule of Assessments

3.2.1 Trait Measures

Both the counselor and all clients completed assessments of trait measures prior to the first conversation. As an overall assessment of personality, both completed the Big Five Inventory (BFI) [83], based on the five-factor model of personality [53].

Adult attachment style has been shown to be predictive of nonverbal behavior in conversation, particularly in close relationships [165, 98]. To assess attachment style, the revised Experiences in Close Relationships (ECR-R) questionnaire [60] was used.

3.2.2 Therapeutic Alliance

Therapeutic alliance, also called working alliance, is a model of interpersonal relationship specific to counselor-client relationships. Bordin [20, 19] developed a concept of therapeutic alliance which includes three parts: *goal*, or agreement between counselor and client on the desired overall outcome of counseling; *task*, or agreement on the actions taken during counseling; and the interpersonal *bond* or rapport and trust between counselor and client. The

goal, task, and bond components may be assessed individually or combined as an overall measure of the strength of alliance.

Several reliable and validated instruments to assess therapeutic alliance have been developed, including the Working Alliance Inventory (WAI) [79] and the Penn Helping Alliance Questionnaire (HAQ) [101]. These and similar instruments have been widely used in prior research, and overall, a strong therapeutic alliance has been shown to be predictive of positive outcomes in counseling [80, 104].

In the exercise counseling corpus, therapeutic alliance was assessed separately by the counselor and client immediately following each conversation. Both completed the short revised Working Alliance Inventory (WAI-SR) [75], an abbreviated (12-item) version of the original (36-item) Working Alliance Inventory. In order to make the assessment appropriate for the more general domain of counseling rather than psychotherapy alone, I modified the WAI-SR slightly, replacing the words “therapy” and “therapist” with “counseling” and “counselor.” The modified questionnaires are included in Appendix A.

3.2.3 Trust

While therapeutic alliance, as measured by the WAI-SR and similar assessments, is an integrative concept of interpersonal relationship, I am also interested in assessing momentary changes in interpersonal relationship, during a conversation, and identifying specific moments at which the relationship may have changed.

Following each conversation, the client watched a videotape of the conversation, while indicating (by pointing to a paper rating scale) points at which his or her trust in the counselor increased or decreased. Each of these replays was itself videotaped. Figure 3.2 shows an example frame of video from a trust assessment. The client is visible in the upper left quadrant, the replay of the conversation is visible in the lower left quadrant, and the trust instrument



Figure 3.2: A participant in the Exercise Counseling Corpus completing a retrospective review of a conversation to assess moment-to-moment trust.

is visible in the upper right quadrant. The bottom right quadrant was blank in all recordings, as three cameras were used, which were then combined into a single recording using a 4-input video mixer.

The continuous trust measure has not yet been analyzed due to time limitations, and is not included in the following discussion or elsewhere in this dissertation. It may be examined in future work, to attempt to identify behavior associated with particular points at which counselor-client trust changes.

3.2.4 Task Outcome Measures

The stated purpose of the conversations, for both the counselor and the client, was to discuss (and intervene on) the clients' daily physical activity. The success or failure of this shared task may potentially be associated with differences in conversational behavior. I assessed both the clients' behavior (the amount

of daily physical activity performed) and their attitudes toward physical activity.

All clients were issued pedometers (Omron HJ-720ITC) at the time of intake, and asked to wear them as much as possible. The pedometers recorded an estimated count of total steps walked per day, as a measure of overall physical activity. The recorded step data was made available to the counselor prior to each conversation after the initial session.

Attitudes toward physical activity were assessed based on constructs drawn from the Transtheoretical Model of Behavior Change [132, 134, 133], a widely-used theoretical model of health behavior change. Decisional balance [121] assesses the subjective importance a client places on the positive implications (pros) and negative implications (cons) of a hypothetical behavior change. Stage of Change [103] is one of: Precontemplation (not actively considering a change); Contemplation (considering a change, but not intending or committed to any action); Preparation (intending to take action); Action (in the process of changing behavior); and Maintenance (in which a client has changed his or her behavior and is working to prevent relapse). Both were assessed at intake, and following the second, fourth, and final conversations.

3.2.5 Cognitive Inventory

The thought-listing procedure [31, 30] is a semi-structured self-report instrument for obtaining participants' cognitive responses to a stimulus. Participants are given a sheet of paper divided into boxes, but otherwise blank, and asked to write whatever thoughts they have about the stimulus as quickly as possible, with one thought per box.

Both the clients and counselor completed (separately) a thought-listing following each conversation. They were given a two minute time limit, and asked to write any thoughts related to the conversation, physical activity, or their conversational partner (the counselor or the client).

3.3 The Corpus

Six clients (5 female), were recruited. Of this group, five completed all six weekly sessions, while the remaining client chose to withdraw from the study following the second session, without citing a specific event or cause. Including all clients, 32 sessions were conducted, with a total of 500.25 minutes (approximately 8.3 hours) of videotaped conversations, an equal duration of videotaped conversation replays (collected for the trust assessment), and 101,493 words of spoken dialogue.

The duration of individual conversations varied widely, ranging from 7.8 to 24.6 minutes (mean 15.6, SD 4.2). Much of this variation was between dyads, with the mean conversation duration per-dyad ranging from 12.5 to 19.6 minutes. Conversations tended to be longer in earlier sessions: the mean duration of an initial conversation was 17.9 minutes, dropping to 13.7 minutes for a final conversation.

I examined this trend with a random-slope mixed-effect linear regression model. The duration in minutes y_{ij} for session i and dyad j is:

$$\begin{aligned}
 y_{ij} &= \beta_0 + u_{0j} + \beta_1 t_i + u_{1j} t_i + \epsilon_{ij} \\
 u_{0j} &\overset{iid}{\sim} N(0, \sigma_0^2) \\
 u_{1j} &\overset{iid}{\sim} N(0, \sigma_1^2) \\
 \epsilon_{ij} &\overset{iid}{\sim} N(0, \sigma_\epsilon^2)
 \end{aligned} \tag{3.1}$$

where t_i is the number of previous conversations (zero at the first session).

Table 3.2 shows the result of a restricted maximum likelihood fit of this model, using the lme4 package [10] in R [136]. Confidence intervals were constructed from a parametric bootstrap with 500 replications. There is a significant linear trend toward shorter conversations in later sessions, at a rate estimated at slightly over one minute per session ($\beta_1 = -1.2$, 95% CI [-1.8, -0.5]), and this trend is consistent across different dyads: while there is substantial variability between dyads in the mean duration of a conversation ($\sigma_0^2 = 2.7$, 95% CI [0.0, 4.8]), there is very little variability in the rate

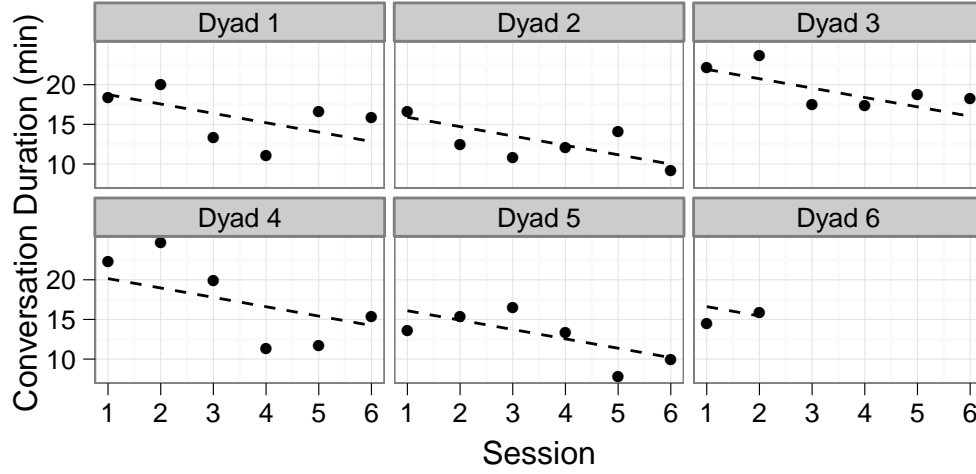


Figure 3.3: Durations of conversations, shown separately by dyad. Dashed lines indicate a linear trend fit by a random-slope mixed effect linear regression model.

Parameter		Estimate	95% CI
Fixed Effects	Intercept β_0	18.2	[15.1, 21.1]
	Sessions β_1	-1.2	[-1.8, -0.6]
Random Effects (dyads)	Intercept σ_0^2	2.7	[0.0, 4.8]
	Sessions σ_1^2	0.0	[0.0, 0.8]
Residual σ_ϵ^2		2.8	[2.0, 3.6]

Table 3.2: Random-slope regression model fit to a linear trend on conversation durations. Confidence intervals are estimated from a parametric bootstrap (500 replicates).

of decrease ($\sigma_1^2 < 0.001$, 95% CI [0.0, 0.8]). Note, however, that the estimation of between-dyad variability in this model is poor, indicated by the wide confidence intervals, which include zero.

3.4 Descriptive Summary of Measures

3.4.1 Trait Measures

The personality assessments of the clients were compared to population norms [157]. All clients were within one standard deviation of the estimated popu-

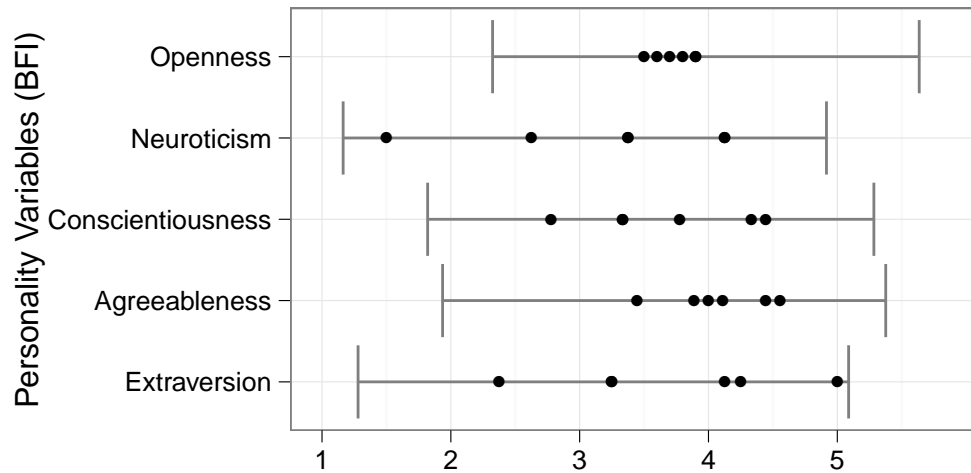


Figure 3.4: Personality assessments of clients at the time of intake. The overlaid intervals indicate population norms: one standard deviation above or below the population mean, as reported in [157].

lation mean on all dimensions of the Big 5 personality model (Figure 3.4).

Assessments of attachment style were also compared to population norms (Figure 3.5)^{footnote}The illustrated population means are reported at <http://internal.psychology.illinois.edu/~rcfrale/measure/ecrr.htm>. Here, the clients' scores were overdispersed, showing high variance relative to the larger population: 5 of 6 clients were between one and two standard deviations of the estimated population mean. However, the mean of the clients' scores was near the estimated population mean.

3.4.2 Therapeutic Alliance

The clients reported high therapeutic alliance overall: the mean WAI-SR score was 4.33 (SD 0.57) on a 1–5 scale. The counselor reported weaker and more variable therapeutic alliance (mean 3.29, SD 0.88). Much of the variability in the counselor's report was between different dyads, with the counselor's per-dyad mean WAI-SR score ranging from 2.08 to 4.33. The clients had less between-dyads variability, ranging from 3.76 to 4.78.

Both the clients and the counselor reported increased therapeutic alliance

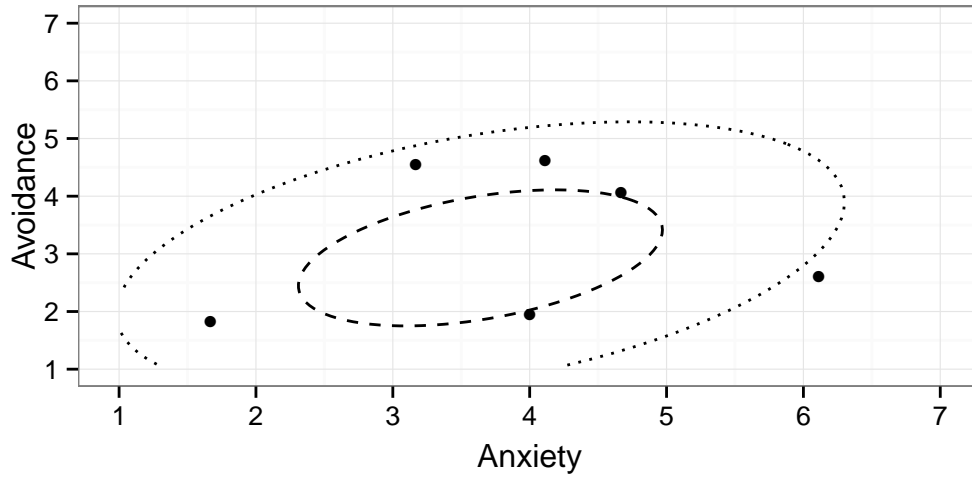


Figure 3.5: Attachment style assessment of clients at the time of intake. The overlaid ellipses indicate population norms: one and two standard deviations away from the population mean.

in later sessions. The mean WAI-SR score reported by a client was 3.71 for a first session (SD 0.47), increasing to 4.68 for a final session (SD 0.38). Similarly, the mean score reported by the counselor was 2.60 in a first session (SD 0.60) and increased to 3.83 in a final session (SD 0.86).

I examined this trend by fitting random-slope mixed-effect linear regression models to the counselor's and clients' WAI-SR scores, separately. Except for the outcome variable, these models are identical to the model used to examine conversation durations (3.1).

Table 3.3 and Table 3.4 give the results of restricted maximum likelihood fits of this model, using the lme4 package [10] in R [136]. Confidence intervals were constructed from a parametric bootstrap with 500 replications. For both the clients and the counselor, there is a significant linear trend toward higher reported therapeutic alliance in later conversations. The slope of this trend is similar for the clients ($\beta_1 = 0.17$, 95% CI [0.10, 0.23]) and the counselor ($\beta_1 = 0.15$, 95% CI [0.03, 0.27]). While there is substantial variability between dyads in the mean reported therapeutic alliance, particularly for the counselor ($\sigma_0^2 = 0.62$, 95% CI [0.00, 1.12]), the rate of increase has less variability for

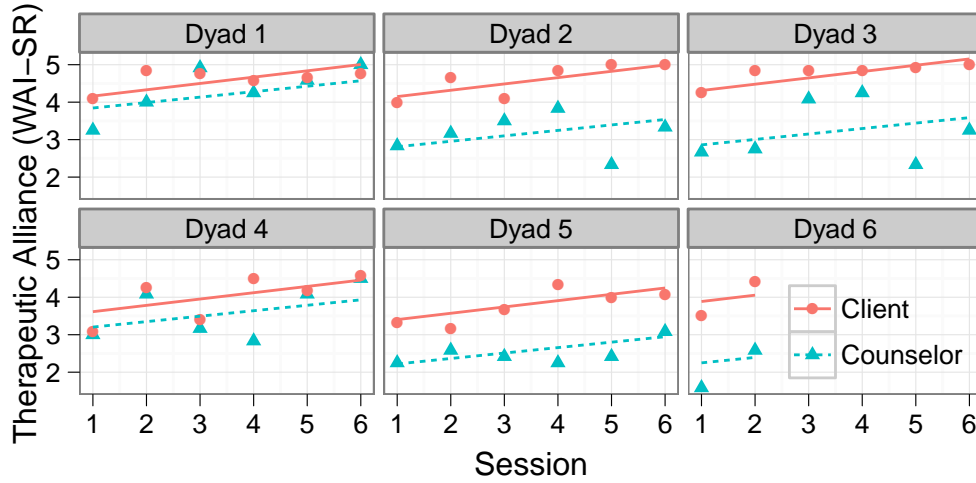


Figure 3.6: Therapeutic alliance, self-reported by clients and counselor, shown separately by dyad. The lines indicate linear trends fit by random-slope mixed-effect regression models.

Parameter			Estimate	95% CI
Fixed Effects	Intercept	β_0	3.92	[3.58, 4.28]
	Sessions	β_1	0.17	[0.10, 0.23]
Random Effects (dyads)	Intercept	σ_0^2	0.38	[0.00, 0.68]
	Sessions	σ_1^2	0.00	[0.00, 0.10]
Residual		σ_ϵ^2	0.32	[0.22, 0.40]

Table 3.3: Random-slope regression model fit to a linear trend on clients' reported therapeutic alliance (WAI-SR). Confidence intervals are estimated from a parametric bootstrap (500 replicates).

both clients ($\sigma_1^2 = 0.00$, 95% CI [0.00, 0.10]) and for counselor ($\sigma_1^2 = 0.06$, 95% CI [0.00, 0.18]). However, the estimation of between-dyad variability in these models is poor, with wide confidence intervals.

3.4.3 Task Outcome Measures

Figure 3.7 shows the clients' self-reported Stage of Change for the uptake of regular physical activity, as assessed at intake and after every other session. All but one began in the contemplation (C) or preparation (P) stages, and all but one advanced one or more stages by the end of their participation in the

Parameter		Estimate	95% CI
Fixed Effects	Intercept β_0	2.86	[2.24, 3.52]
	Sessions β_1	0.15	[0.03, 0.27]
Random Effects (dyads)	Intercept σ_0^2	0.62	[0.00, 1.12]
	Sessions σ_1^2	0.06	[0.00, 0.18]
Residual σ_ϵ^2		0.58	[0.40, 0.73]

Table 3.4: Random-slope regression model fit to a linear trend on the counselor's reported therapeutic alliance (WAI-SR). Confidence intervals are estimated from a parametric bootstrap (500 replicates).

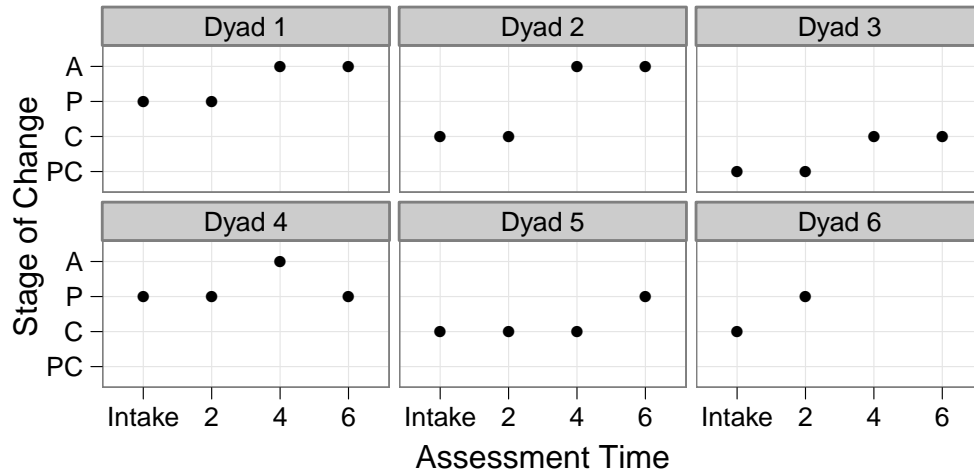


Figure 3.7: Stage of Change for physical activity, self-reported by clients at intake, and following the second, fourth, and last sessions. PC=precontemplation, C=contemplation, P=preparation, A=action.

study.

Figure 3.8 shows the clients' self-reported decisional balance, also assessed at intake and after every other session. While there is no clear and consistent trend of change across time, I note that with a single exception (dyad 5), all clients reported placing consistently higher importance on pros than cons, and that this difference either remained the same or increased across sessions.

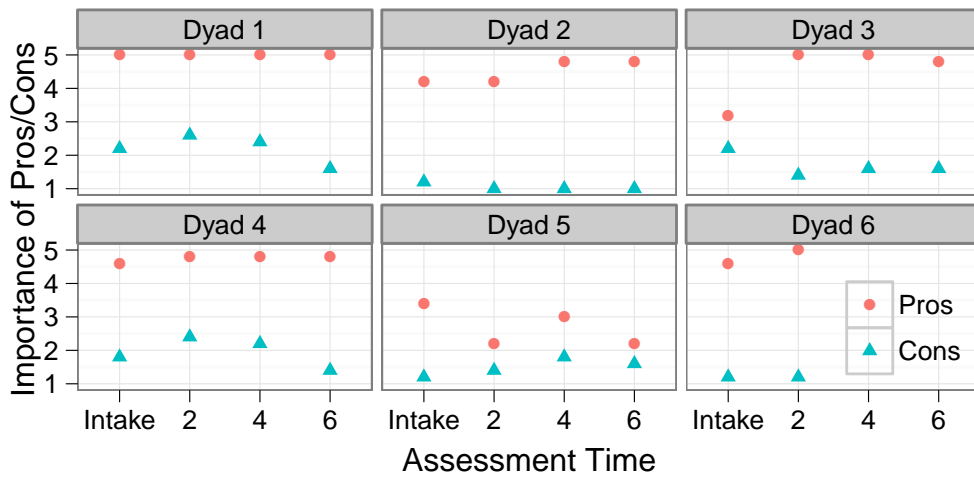


Figure 3.8: Decisional Balance for physical activity, self-reported by clients at intake, and following the second, fourth, and last sessions.

3.4.4 Behavioral Measures

Due to a technical problem, the recorded pedometer data was not accessible following the study, and is not analyzed here.

3.4.5 Cognitive Inventory

A total of 440 “thoughts” were reported in the thought-listing tasks following each conversation, with a mean of 6.9 per participant per session (SD 2.9). The counselor tended to report more thoughts per session (mean 9.0, SD 2.6) than the clients (mean 4.8, SD 1.2).

Participants frequently reported thoughts relating to the content of the conversations, and the counselor often reported thoughts about the quality of interaction (e.g., “good flow of conversation”). However, with only two exceptions, (“doesn’t always look me in the eye,” and “[the counselor] was very nice and had good eye contact”), no participants reported any thoughts directly relating to any of the specific verbal or nonverbal behaviors studied here. No other thoughts were reported that appeared to be conscious assessments of specific verbal and nonverbal behaviors.

Further analyses of the cognitive inventories was not performed due to time constraints. Future work may explore using this information to separate the conversations in the corpus along various dimensions (e.g., into “successful” and “unsuccessful” conversations), in order to examine conversational behavior associated with these dimensions.

3.5 Correlations Between the Measures

As an exploratory analysis, I computed pairwise correlations between several of the measures assessed during the collection of the corpus. Most measures were assessed several times per dyad: either every session (in the case of therapeutic alliance) or every other session (in the case of the task outcome measures). Ordinary Pearson correlation is not an appropriate measure of association, as it is based on the assumption that observations at each instance of measurement are independent [18], whereas here observations are grouped by dyad. Common alternatives to Pearson correlation, such as Spearman rank correlation, relax other assumptions (e.g. linearity) but also assume independence of observations and are similarly inappropriate.

Depending on the research question of interest, there are multiple correlations that may be reported for this type of longitudinal data. I give three for each pair (X, Y) of variables: (a) correlations between dyads, testing whether the dyad-mean value of X is correlated with the dyad-mean value of Y , (b) correlations between sessions (within dyads), controlling for the overall correlation at the dyad level, and (c) total correlation of X and Y measured at the same session, including both between-dyad and within-dyad correlation. The first measure of correlation is similar to a naïve Pearson correlation on dyad mean values, while the last measure is similar to a naïve Pearson correlation on all observations, ignoring grouping.

All three correlations are estimated simultaneously by fitting a bivariate mixed-effect model to the observations [73]. This model assumes that dyad

means for X and Y are drawn from a bivariate normal distribution, and that each per-session observation is drawn from a second bivariate normal distribution, centered on the dyad mean. For observations x_{ij} and y_{ij} of dyad i and session j :

$$\begin{pmatrix} x_{ij} \\ y_{ij} \end{pmatrix} \stackrel{iid}{\sim} \text{Normal} \left(\begin{pmatrix} u_{xi} \\ u_{yi} \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} \right) \\ \begin{pmatrix} u_{xi} \\ u_{yi} \end{pmatrix} \stackrel{iid}{\sim} \text{Normal} \left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \varsigma_x^2 & \varsigma_{xy} \\ \varsigma_{xy} & \varsigma_y^2 \end{pmatrix} \right) \quad (3.2)$$

where μ_x, μ_y are the population means of x and y , μ_{xi}, μ_{yi} are the per-dyad means for dyad i , $\sigma_x^2, \sigma_y^2, \sigma_{xy}$ and the within-dyad variances and covariance, and $\varsigma_x^2, \varsigma_y^2, \varsigma_{xy}$ are the between-dyad variances and covariance. The three correlations of interest can now be estimated as:

$$\begin{aligned} r_{\text{dyad}} &= \frac{\varsigma_{xy}}{\varsigma_x \varsigma_y} \\ r_{\text{session}} &= \frac{\sigma_{xy}}{\sigma_x \sigma_y} \\ r_{\text{total}} &= \frac{\varsigma_{xy} + \sigma_{xy}}{\sqrt{\varsigma_x^2 + \sigma_x^2} + \sqrt{\varsigma_y^2 + \sigma_y^2}} \end{aligned} \quad (3.3)$$

For each pair of variables, the restricted maximum likelihood fit was computed using the nlme package [126] and R [136]. Confidence intervals and p -values were estimated with a residuals-resampling bootstrap, which is robust against violations of normality [36].

3.5.1 Therapeutic Alliance

Table 3.5 gives the correlations between counselor and client assessments of therapeutic alliance, reported separately for the three components of therapeutic alliance (goal, task, and bond), and for the overall alliance. In all cases, there is a positive estimated correlation at the dyad level. However, this correlation is not near significance in any case; this is expected, as the corpus has a small sample size at the dyad level and is therefore underpowered to detect between-dyad correlation.

		r	95% CI	p
Overall	Dyad	0.59	[-0.85, 1.00]	0.216
	Session	0.36	[0.03, 0.65]	0.036*
	Total	0.47	[0.03, 0.76]	0.036*
Goal	Dyad	0.68	[-0.97, 1.00]	0.204
	Session	-0.01	[-0.33, 0.30]	0.940
	Total	0.25	[-0.21, 0.57]	0.300
Task	Dyad	0.54	[-0.94, 1.00]	0.336
	Session	0.09	[-0.16, 0.37]	0.528
	Total	0.31	[-0.27, 0.72]	0.252
Bond	Dyad	0.61	[-0.97, 1.00]	0.256
	Session	0.53	[0.28, 0.76]	0.004**
	Total	0.55	[0.24, 0.80]	0.002**

Table 3.5: Correlations between counselor and client assessments of therapeutic alliance. Confidence intervals and p -values are estimated from a residuals-resampling bootstrap (500 replicates).

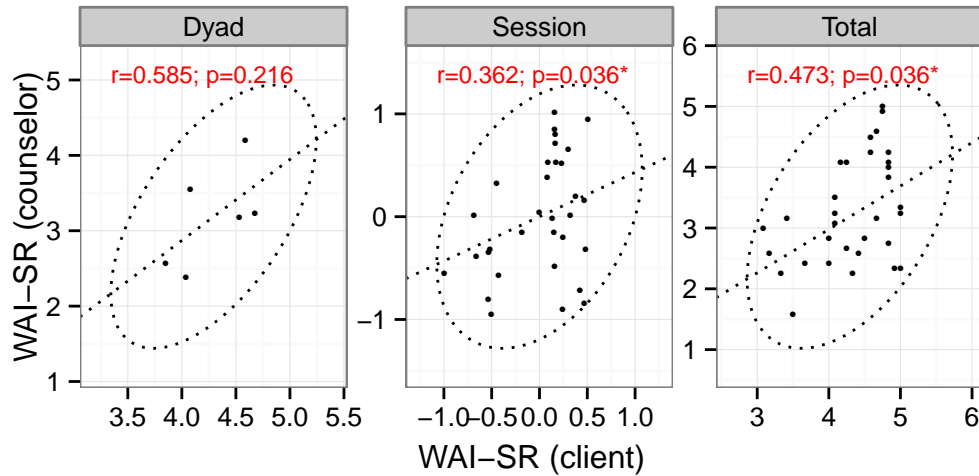


Figure 3.9: Correlations between counselor and client overall therapeutic alliance assessments. The plots show dyad mean values (Dyad), dyad-mean-adjusted values (Session), and raw values (Total), respectively.

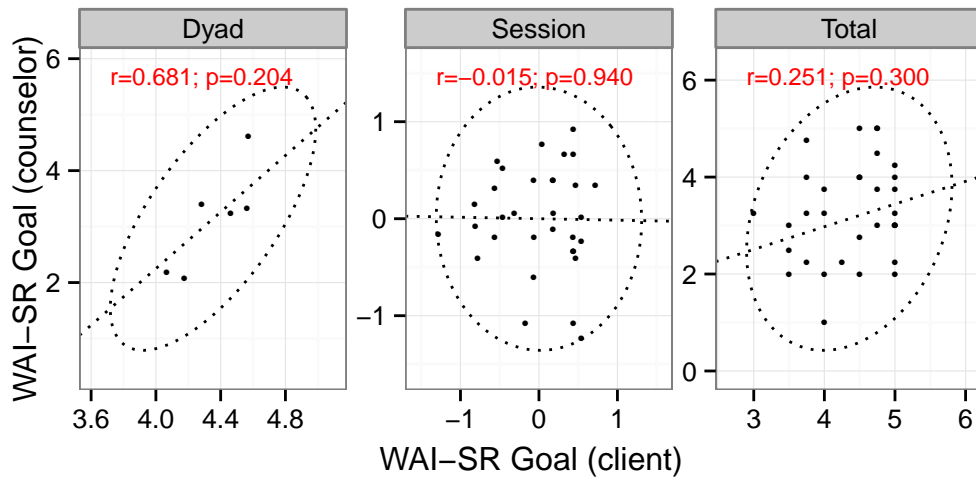


Figure 3.10: Correlations between counselor and client therapeutic alliance goal assessments. The plots show dyad mean values (Dyad), dyad-mean-adjusted values (Session), and raw values (Total), respectively.

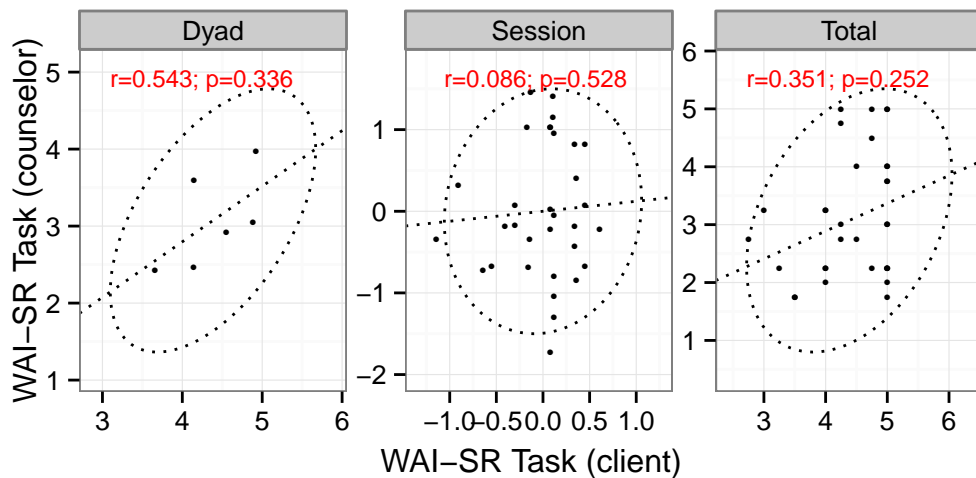


Figure 3.11: Correlations between counselor and client therapeutic alliance task assessments. The plots show dyad mean values (Dyad), dyad-mean-adjusted values (Session), and raw values (Total), respectively.

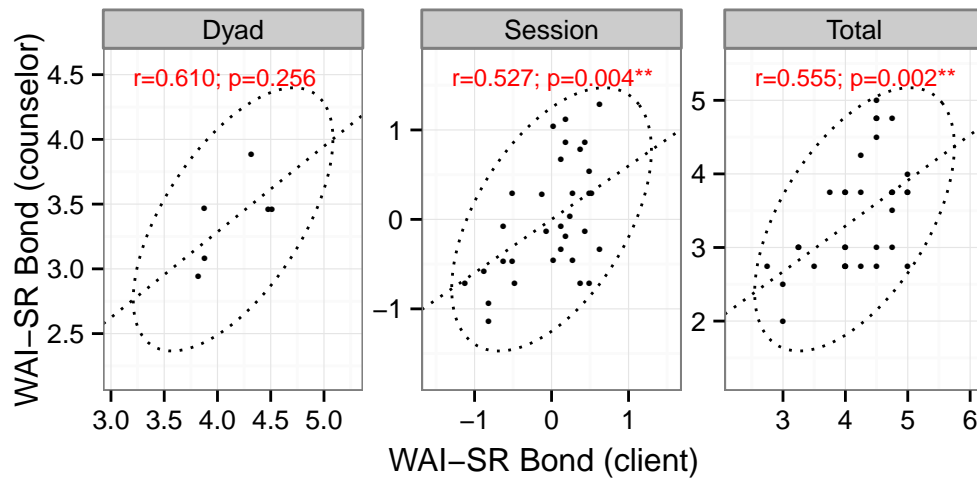


Figure 3.12: Correlations between counselor and client therapeutic alliance bond assessments. The plots show dyad mean values (Dyad), dyad-mean-adjusted values (Session), and raw values (Total), respectively.

On the overall assessment of alliance, there is a significant within-dyad correlation ($r = 0.36, p = 0.018$) as well as a significant total correlation ($r = 0.47, p = 0.024$). Counselor and client assessments of alliance are significantly positively associated at the session level: the counselor tended to give high (and low) assessments in the same session as the client, even when accounting for between-dyad differences. Figure 3.9 illustrates this result. However, inspection of the individual components of therapeutic alliance shows that there is no evidence of significant correlation at either the dyad or session level for the goal (Figure 3.10) or task (Figure 3.11) components. The reported overall association is due only to the bond component (Figure 3.12), where counselor and client assessments have significant within-dyad correlation ($r = 0.53, p = 0.002$) and total correlation ($r = 0.55, p = 0.002$).

3.5.2 Therapeutic Alliance and Conversation Duration

Table 3.6 gives the correlations between conversation durations and assessments of therapeutic alliance, shown separately for counselor and client assessments, and for overall alliance and the three components (goal, task, and

			r	95% CI	p
Client	Overall	Dyad	0.58	[-0.98, 1.00]	0.284
		Session	-0.42	[-0.70, -0.11]	0.004**
		Total	0.20	[-0.45, 0.67]	0.520
	Goal	Dyad	0.78	[-0.99, 1.00]	0.240
		Session	-0.28	[-0.60, 0.10]	0.160
		Total	0.16	[-0.31, 0.60]	0.416
	Task	Dyad	0.65	[-0.59, 1.00]	0.108
		Session	-0.41	[-0.68, -0.09]	0.024*
		Total	0.33	[-0.37, 0.78]	0.212
	Bond	Dyad	0.35	[-1.00, 1.00]	0.629
		Session	-0.43	[-0.68, -0.10]	0.012*
		Total	0.01	[-0.55, -.60]	0.938
Counselor	Overall	Dyad	0.54	[-0.80, 1.00]	0.124
		Session	-0.30	[-0.64, 0.04]	0.076
		Total	0.25	[-0.35, 0.60]	0.225
	Goal	Dyad	0.46	[-0.42, 1.00]	0.140
		Session	-0.17	[-0.49, 0.13]	0.212
		Total	0.30	[-0.21, 0.73]	0.172
	Task	Dyad	0.60	[-0.66, 1.00]	0.132
		Session	-0.22	[-0.56, 0.16]	0.228
		Total	0.20	[-0.30, 0.58]	0.260
	Bond	Dyad	0.67	[-0.87, 1.00]	0.160
		Session	-0.40	[-0.78, -0.04]	0.028*
		Total	0.11	[-0.38, 0.48]	0.480

Table 3.6: Correlations between conversation durations and assessments of therapeutic alliance. Confidence intervals and p -values are estimated from a residuals-resampling bootstrap (500 replicates).

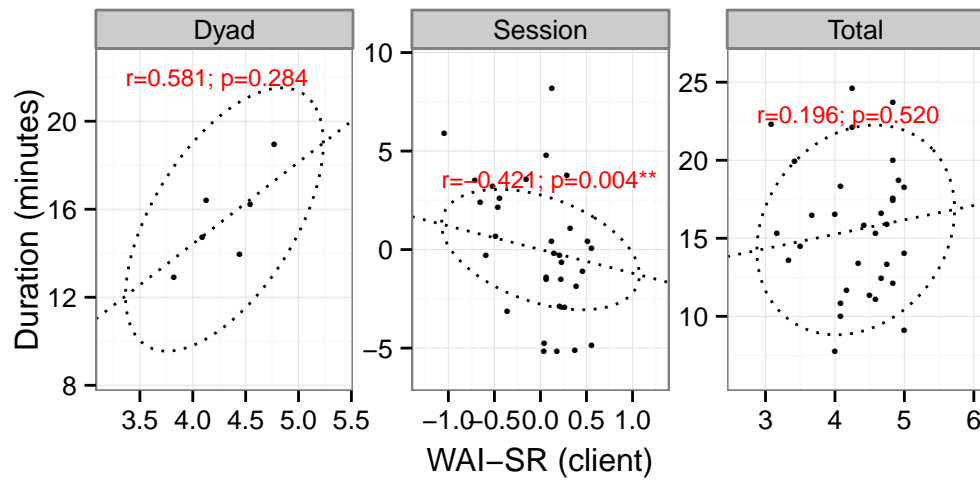


Figure 3.13: Correlations between conversation durations and client overall therapeutic alliance assessments. The plots show dyad mean values (Dyad), dyad-mean-adjusted values (Session), and raw values (Total), respectively.

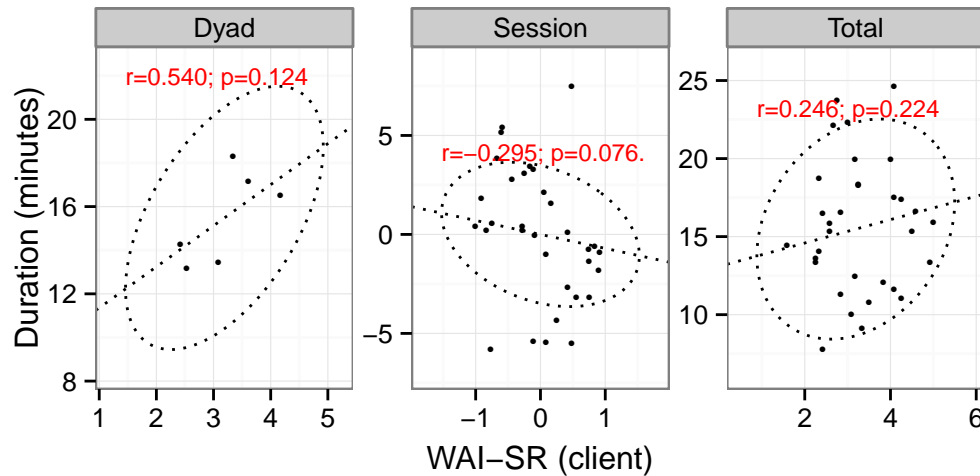


Figure 3.14: Correlations between conversation durations and counselor overall therapeutic alliance assessments. The plots show dyad mean values (Dyad), dyad-mean-adjusted values (Session), and raw values (Total), respectively.

	r	95% CI	p
Dyad	0.36	[-0.94, 1.00]	0.388
Session	-0.40	[-0.70, 0.08]	0.100
Total	0.07	[-0.49, 0.48]	0.976

Table 3.7: Correlations between assessments of decisional balance pros and cons. Confidence intervals and p -values are estimated from a residuals-resampling bootstrap (500 replicates).

bond). In all cases, there is a positive estimated correlation at the dyad level; i.e., dyads with longer conversations on average tended to report strong alliance on average. However, the correlation is not statistically significant; as above, the sample size of the corpus is small and it is underpowered.

At the session level, however, there is a significant negative correlation between conversation duration and overall therapeutic alliance reported by the client ($r = -0.42, p = 0.004$). Figure 3.13 illustrates this relationship: conversations that were longer than average for a particular dyad tended to have lower assessments of overall therapeutic alliance by the client. There are similar negative correlations at the session level for that task ($r = -0.41, p = 0.024$) and bond ($r = -0.43, p = 0.012$) components, and a similar but weaker negative correlation (which is not statistically significant) for the goal component ($r = -0.28, p = 0.160$).

There were similar but weaker negative correlations between conversation durations and the therapeutic alliance reported by the counselor (Figure 3.14). However, the correlation is significant only for the bond component ($r = -0.40, p = 0.028$), and near-significant for the overall alliance assessment ($r = -0.30, p = 0.076$).

3.5.3 Decisional Balance

I tested for correlations between client assessments of decisional balances pros and cons toward regular exercise (assessed at intake, and following the second, fourth, and sixth sessions). Results are shown in Table 3.7 and Fig-

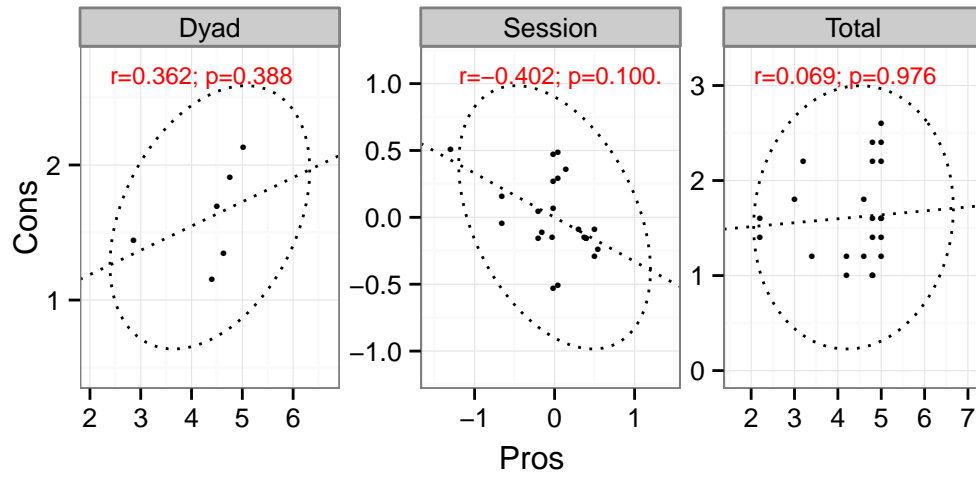


Figure 3.15: Correlations between assessments of decisional balance pros and cons. The plots show dyad mean values (Dyad), dyad-mean-adjusted values (Session), and raw values (Total), respectively.

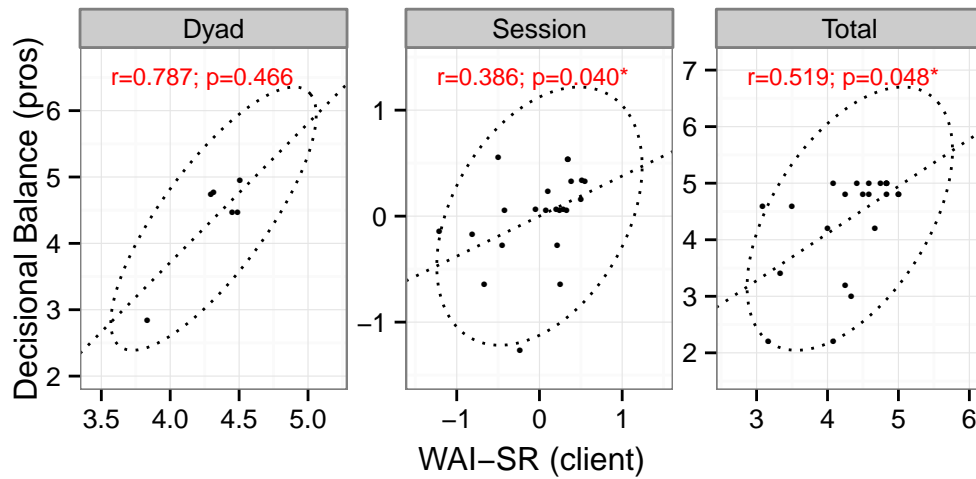


Figure 3.16: Correlations between client assessments of therapeutic alliance and decisional balance pros. The plots show dyad mean values (Dyad), dyad-mean-adjusted values (Session), and raw values (Total), respectively.

ure 3.15. There is a near-significant negative correlation at the session level ($r = -0.40, p = 0.100$); clients tended to report higher importance on the pros when they imported lower importance on cons, relative to their average assessments.

To explore possible associations between therapeutic alliance and attitudes

			r	95% CI	p
Client	Overall	Dyad	0.79	[-0.99, 1.00]	0.47
		Session	0.39	[0.01, 0.71]	0.040*
		Total	0.52	[0.00, 0.81]	0.050*
	Goal	Dyad	0.99	[-0.97, 1.00]	0.204
		Session	0.46	[0.00, 0.77]	0.050*
		Total	0.53	[0.13, 0.80]	0.008**
	Task	Dyad	0.75	[-0.99, 1.00]	0.356
		Session	0.32	[-0.13, 0.78]	0.160
		Total	0.57	[-0.10, 0.86]	0.104
	Bond	Dyad	0.42	[-0.99, 1.00]	0.732
		Session	0.29	[-0.14, 0.65]	0.164
		Total	0.31	[-0.18, 0.64]	0.244
Counselor	Overall	Dyad	0.66	[-0.98, 1.00]	0.320
		Session	0.20	[-0.25, 0.55]	0.336
		Total	0.44	[-0.19, 0.77]	0.140
	Goal	Dyad	0.58	[-0.96, 1.00]	0.176
		Session	0.27	[-0.14, 0.59]	0.188
		Total	0.50	[-0.05, 0.87]	0.076
	Task	Dyad	0.80	[-0.98, 1.00]	0.389
		Session	0.13	[-0.40, 0.55]	0.605
		Total	0.35	[-0.25, 0.67]	0.281
	Bond	Dyad	0.98	[-0.98, 1.00]	0.425
		Session	0.19	[-0.23, 0.54]	0.357
		Total	0.30	[-0.21, 0.65]	0.200

Table 3.8: Correlations between decisional balance pros and assessments of therapeutic alliance. Confidence intervals and p -values are estimated from a residuals-resampling bootstrap (500 replicates).

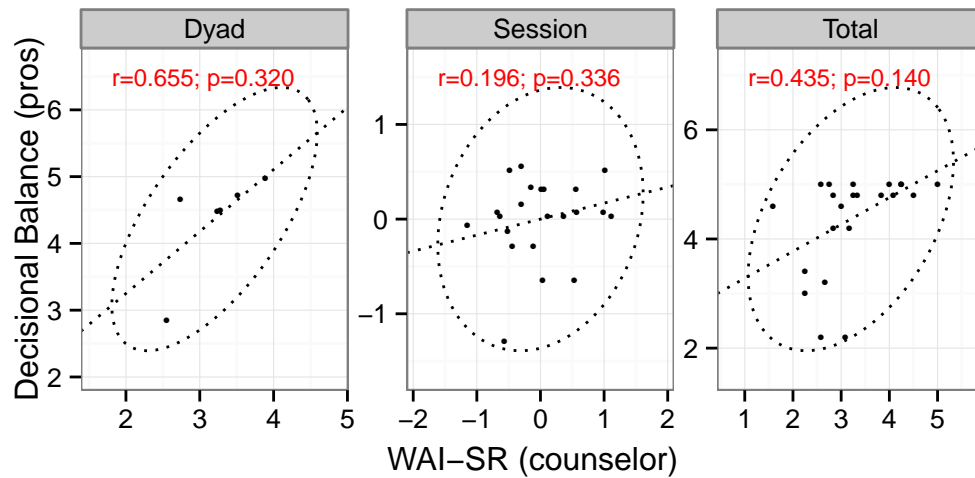


Figure 3.17: Correlations between counselor assessments of therapeutic alliance and client assessments of decisional balance pros. The plots show dyad mean values (Dyad), dyad-mean-adjusted values (Session), and raw values (Total), respectively.

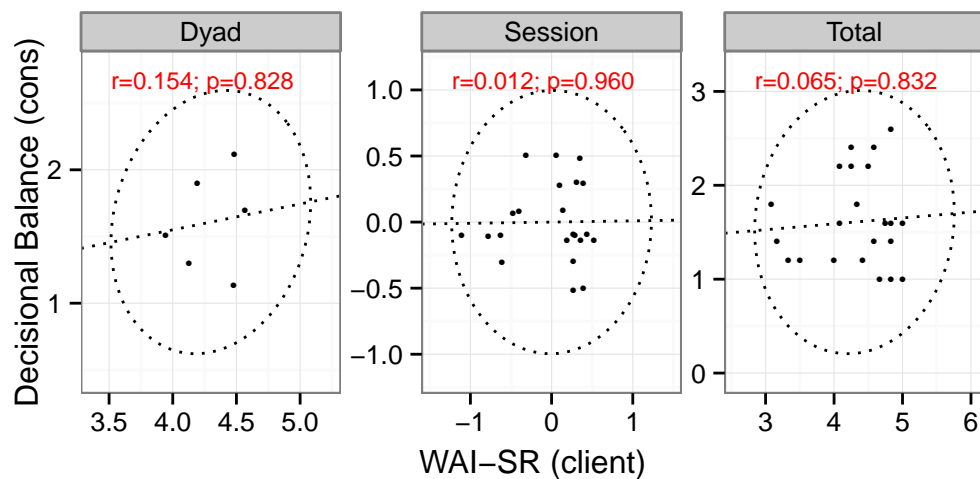


Figure 3.18: Correlations between client assessments of therapeutic alliance and decisional balance cons. The plots show dyad mean values (Dyad), dyad-mean-adjusted values (Session), and raw values (Total), respectively.

			r	95% CI	p
Client	Overall	Dyad	0.15,	[-0.99, 1.00]	0.828
		Session	0.01	[-0.41, 0.43]	0.960
		Total	0.07	[-0.39, 0.48]	0.832
	Goal	Dyad	0.39	[-0.99, 0.99]	0.888
		Session	-0.12	[-0.52, 0.34]	0.624
		Total	-0.05	[-0.50, 0.34]	0.752
	Task	Dyad	0.31	[-0.99, 1.00]	0.652
		Session	0.16	[-0.16, 0.46]	0.304
		Total	0.24	[-0.25, 0.61]	0.368
	Bond	Dyad	-0.19	[-0.99, 0.99]	0.828
		Session	0.01	[-0.40, 0.45]	0.956
		Total	-0.05	[-0.45, 0.45]	0.864
Counselor	Overall	Dyad	0.95	[-0.37, 1.00]	0.076
		Session	-0.35	[-0.71, 0.18]	0.164
		Total	0.25	[-0.38, 0.56]	0.548
	Goal	Dyad	0.77	[-0.40, 1.00]	0.120
		Session	-0.15	[-0.49, 0.26]	0.472
		Total	0.44	[-0.22, 0.74]	0.200
	Task	Dyad	0.99	[-0.74, 1.00]	0.124
		Session	-0.28	[-0.72, 0.29]	0.336
		Total	0.16	[-0.47, 0.50]	0.856
	Bond	Dyad	0.99	[-0.99, 1.00]	0.293
		Session	-0.40	[-0.72, 0.13]	0.104
		Total	-0.04	[-0.56, 0.39]	0.665

Table 3.9: Correlations between decisional balance cons and assessments of therapeutic alliance. Confidence intervals and p -values are estimated from a residuals-resampling bootstrap (500 replicates).

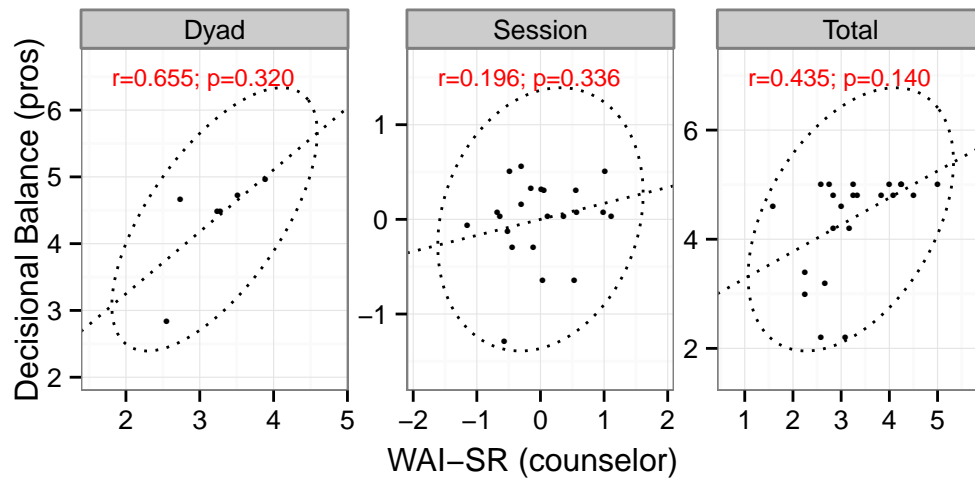


Figure 3.19: Correlations between counselor assessments of therapeutic alliance and client assessments of decisional balance cons. The plots show dyad mean values (Dyad), dyad-mean-adjusted values (Session), and raw values (Total), respectively.

toward regular exercise in the corpus, I tested for correlations between counselor and client assessments of therapeutic alliance and client assessment of decisional balance pros and cons. These assessments were not performed at the same time points: therapeutic alliance was assessed following each weekly conversation, while decisional balance was assessed at intake (before the first conversation), and following the second, fourth, and sixth conversations. For simplicity, the reported correlations ignore therapeutic alliance assessments from the third and fifth conversation, and treat the intake assessment of decisional balance as paired with the first conversation assessment of therapeutic alliance.

Table 3.8 shows correlations with client assessments of decisional balance pros. Stronger reported overall therapeutic by the client is significantly correlated with higher importance placed on the pros of regular exercise (Figure 3.16) at both the session level ($r = 0.39$, $p = 0.04$) and in total ($r = 0.52$, $p = 0.05$). Of the components of therapeutic alliance, only the goal component is significantly correlated, although both task and bond have trends

		r	95% CI	p
Pros	Dyad	0.44	[-0.96, 1.00]	0.316
	Session	0.32	[-0.21, 0.67]	0.220
	Total	0.38	[-0.07, 0.75]	0.108
Cons	Dyad	0.41	[-0.99, 1.00]	0.564
	Session	-0.12	[-0.52, 0.34]	0.616
	Total	0.11	[-0.48, 0.50]	0.760

Table 3.10: Correlations between assessments of decisional balance and stage of change. Confidence intervals and p -values are estimated from a residuals-resampling bootstrap (500 replicates).

in the same direction. The counselor's assessment of therapeutic alliance is not significantly correlated with assessments of decisional balance pros (Figure 3.17).

Table 3.9 show correlations with client assessments of decisional balance cons. There is no significant correlation, either for the client's assessment of therapeutic alliance (Figure 3.18) or for the counselor's assessment (Figure 3.19). Note that there is a very high (and near-significant) estimated positive correlation between counselor assessments of therapeutic alliance and decisional balance cons at the dyad level ($r = 0.95$, $p = 0.076$).

3.5.4 Stage of Change

For simplicity, the same statistical method as in previous sections is used to examine associations between stage of change assessments and other variables. However, I note that since stage of change is an ordinal variable (assessed as one of the five stages, ordered from precontemplation to maintenance) rather than continuous, estimates of correlation may be attenuated.

Table 3.10 shows correlations between assessments of stage of change and assessment of decisional balance (both pros and cons). There are no significant correlations at either the dyad or session level, for either pros (Figure 3.20) or for cons (Figure 3.21).

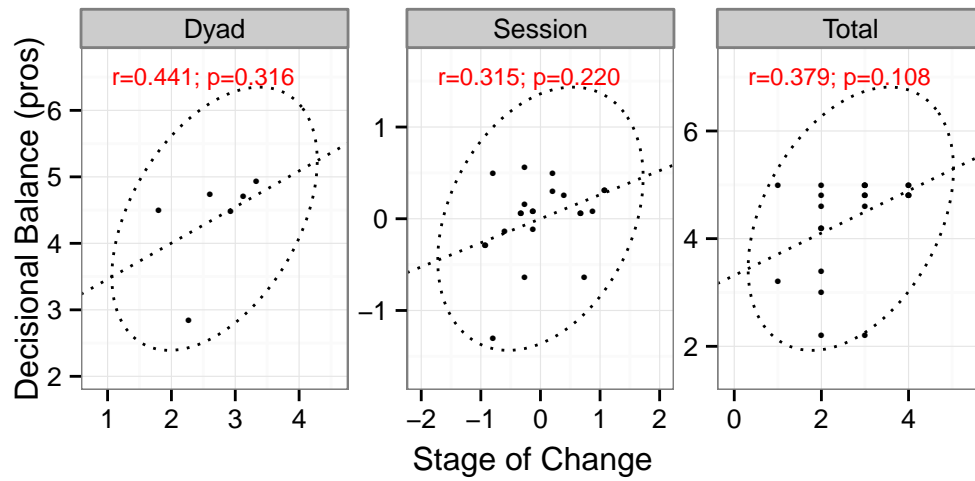


Figure 3.20: Correlations between stage of change assessments and decisional balance pros. The plots show dyad mean values (Dyad), dyad-mean-adjusted values (Session), and raw values (Total), respectively.

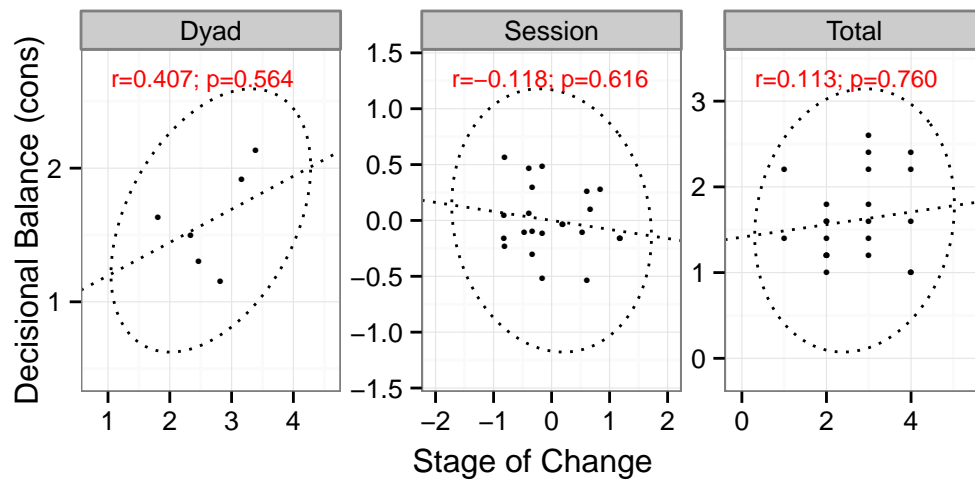


Figure 3.21: Correlations between stage of change assessments and decisional balance cons. The plots show dyad mean values (Dyad), dyad-mean-adjusted values (Session), and raw values (Total), respectively.

			r	95% CI	p
Client	Overall	Dyad	-0.23	[-0.99, 1.00]	0.990
		Session	0.52	[0.23, 0.78]	0.004**
		Total	0.26	[-0.14, 0.75]	0.200
	Goal	Dyad	0.02	[-0.99, 0.99]	0.688
		Session	0.36	[-0.11, 0.67]	0.092
		Total	0.26	[-0.13, 0.65]	0.136
	Task	Dyad	-0.13	[-0.99, 1.00]	0.928
		Session	0.58	[0.22, 0.78]	0.002**
		Total	0.26	[-0.24, 0.75]	0.260
	Bond	Dyad	-0.33	[-0.99, 0.99]	0.838
		Session	0.48	[0.15, 0.74]	0.016*
		Total	-0.21	[-0.21, 0.68]	0.240
Counselor	Overall	Dyad	0.60	[-0.97, 1.00]	0.252
		Session	0.47	[0.04, 0.77]	0.028*
		Total	0.53	[0.05, 0.82]	0.032*
	Goal	Dyad	0.56	[-0.98, 1.00]	0.280
		Session	0.34	[-0.08, 0.73]	0.088
		Total	0.45	[-0.14, 0.79]	0.100
	Task	Dyad	0.85	[-0.31, 1.00]	0.096
		Session	0.44	[0.12, 0.77]	0.004**
		Total	0.58	[0.24, 0.81]	0.002**
	Bond	Dyad	0.70	[-0.98, 1.00]	0.464
		Session	0.48	[0.12, 0.76]	0.012*
		Total	0.41	[0.00, 0.73]	0.050*

Table 3.11: Correlations between assessments of stage of change and assessments of therapeutic alliance. Confidence intervals and p -values are estimated from a residuals-resampling bootstrap (500 replicates).

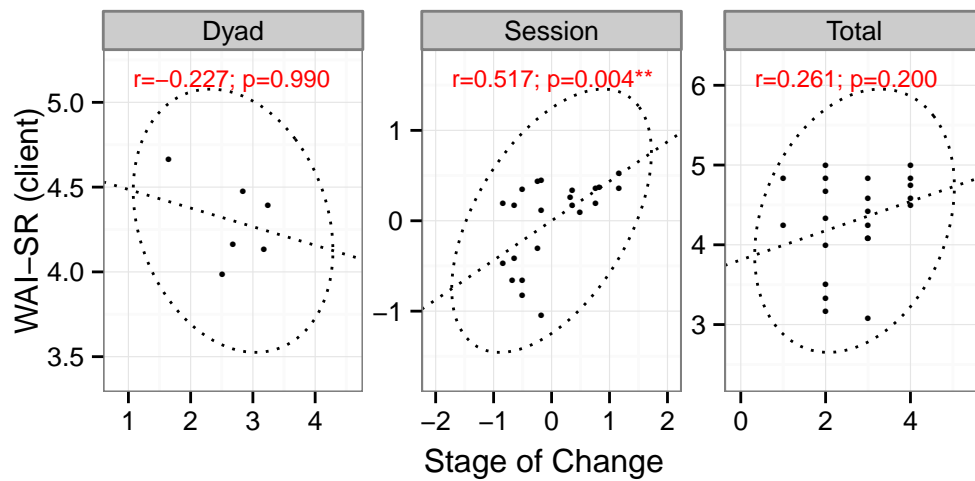


Figure 3.22: Correlations between stage of change assessments and client assessments of therapeutic alliance. The plots show dyad mean values (Dyad), dyad-mean-adjusted values (Session), and raw values (Total), respectively.

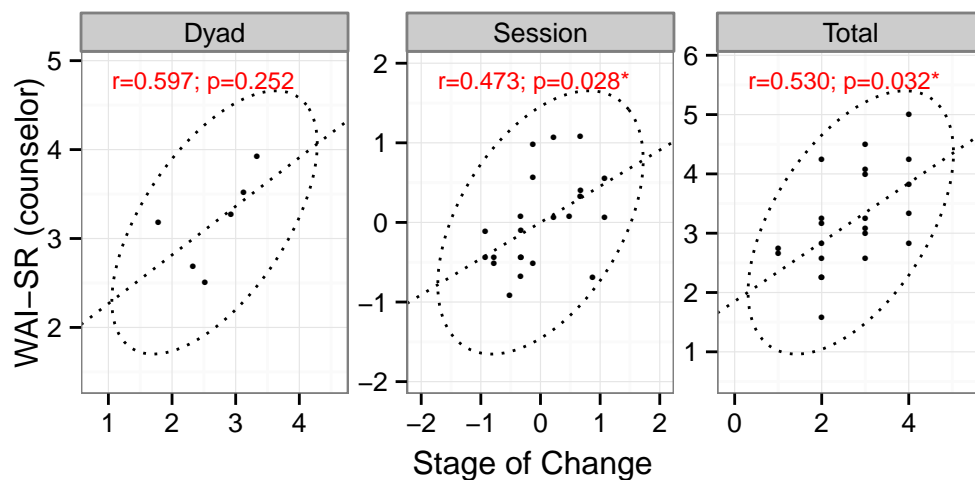


Figure 3.23: Correlations between stage of change assessments and counselor assessments of therapeutic alliance. The plots show dyad mean values (Dyad), dyad-mean-adjusted values (Session), and raw values (Total), respectively.

Correlations between stage of change assessments and therapeutic alliance assessments were estimated, as for decisional balance and alliance above, by examining therapeutic alliance assessments only for sessions in which there was also a stage of change assessments (i.e., ignoring the third and fifth sessions), and taking the intake assessment of stage of change to be the same point of observation as the therapeutic alliance assessment at the end of the first session. Table 3.11 shows the resulting estimates. There is a significant correlation between client assessments of therapeutic alliance and assessments of stage of change at the session level: clients reported later stages of change (relative to their mean assessment) in sessions when they reported stronger alliance (Figure 3.22). This association was found both for overall alliance ($r = 0.052$, $p = 0.004$) and for the task ($r = 0.58$, $p = 0.002$) and bond ($r = 0.48$, $p = 0.016$) components; for the goal component, the estimated correlation was also positive and near-significant ($p = 0.36$, $p = 0.092$). Similarly, there is a significant positive correlation at the session level between counselor assessments of therapeutic alliance and assessments of stage of change (Figure 3.23); this association was found, as with client assessments, for the counselor's overall assessment of alliance ($r = 0.47$, $p = 0.028$) and for the task ($r = 0.44$, $p = 0.004$) and bond ($r = 0.48$, $p = 0.012$) components, and was near-significant for the goal component ($r = 0.34$, $p = 0.088$).

3.6 Discussion

In summary, the Exercise Counseling Corpus is a longitudinal corpus, containing multiple conversations between each of several dyads, with sufficient length of interaction to observe changes in interpersonal relationship, task familiarity, task outcomes, and other contextual features of interaction. In the rest of this section, I give a more detailed discussion of the suitability and limitations of the corpus for research tasks related to the overall research programme of this dissertation.

3.6.1 The Size and Scope of the Corpus

In terms of the total amount of recorded conversation (about 8.3 hours), the Exercise Counseling Corpus is a large corpus relative to other multimodal corpora collected for the modeling of broadly-applicable patterns of specific verbal and nonverbal behaviors. For example, Cassell et al. [41] used a corpus of 70.5 minutes of video to create a (non-longitudinal) model of posture shifts as predicted by discourse structure; this is comparable to the amount of video of initial sessions alone in the Exercise Counseling Corpus. However, the corpus is small relative to corpora intended for general-purpose reuse across multiple research projects: the AMI Meeting Corpus [35] is an order of magnitude larger, containing over 100 hours of video. Despite this, the corpus should contain sufficient examples of patterns of behavior which occur broadly in most conversation.

The number of participants is a limitation of the corpus, as there are conversations with only 6 different clients, and a single counselor. The corpus is consequently a weaker tool for studying differences between participants. In particular, any systematic patterns of behavior which are observed in the counselor and not the clients may be idiosyncrasies of one individual which would not be produced by others. Such observations may be useful for developing models of behavior of a single individual (e.g., to reproduce an individual's conversational "style," as done by Neff et al. [120]) rather than more general models of human behavior.

For each dyad, the corpus contains a maximum of six conversations, spanning a total time period of 6 weeks. Some potential long-term applications of conversational agents may span much longer periods: months or years, with hundreds of conversations. However, six conversations is a long enough time period to see qualitative changes in the participants' interpersonal relationship. For example, Tickle-Degnen and Gavett [160] describe a 3-stage model of the development of a counselor-client working relationship (development of

rapport, development of working alliance, and ongoing working relationship). In an observational study of speech and language therapy sessions, they identify these three stages within 8 weekly sessions (for each of 6 dyads); the first stage occurs entirely within initial sessions.

Within the corpus, we can identify conversations representing, at minimum, three different points in the development of the counselor-client interpersonal relationship: initial conversations, where the previously unacquainted dyad begins a long-term interaction and the associated interpersonal relationship, middle conversations where the long-term interaction is ongoing and has continuity to both past and future, and the final session, in which the dyad terminates their interaction.

A dyad's participation in the corpus has a maximum number of sessions and a defined ending point. At the last conversation, both participants are aware it is their last conversation, and this may influence their behavior. A pre-defined ending point and a fixed number of sessions is common (although certainly not universal) in many counseling relationships, motivated either by theoretical considerations or by financial or other constraints. Many other types of long-term interaction, including other types of counseling, and friendship and companionship relationships, lack a defined end point, and consequently persons in such interaction may show different behavior than can be observed in the corpus.

3.6.2 Characterizing the Corpus by Self-Report

Assessments

Given the small number of participants, it is not feasible to use the trait measures (personality and attachment style) as predictors of conversational behavior in the corpus. However, I observe that on all five personality dimensions (Figure 3.4), participants do not vary far from the population mean, with no indication that any are outliers relative to the larger population. On

attachment style (Figure 3.5), participants are more dispersed from the population mean, although in different directions; the group mean is still near the population mean. These observations somewhat increase confidence that the participants as a group are fairly representative of a larger population, and increase confidence that behaviors observed in the corpus may be generalizable to a larger population.

There is a strong trend, consistent across dyads, for increasing therapeutic alliance with an increasing number of prior conversations. This trend accounts for some, but not all variability in therapeutic alliance within each participant's reports. Therefore, the corpus may be suitable for considering therapeutic alliance as a predictor of behavior separate from the number of prior conversations, although with reduced power. However, the corpus contains no examples of dyads which do *not* increase therapeutic alliance over time. It may be suitable for developing models of behavior in a successful counseling relationship, but less suitable for developing models that can predict behavior in a failed relationship.

The Stage of Change and decisional balance assessments indicate a moderate change in clients' attitudes toward exercise, although this change was fairly consistent across dyads. As with changes in therapeutic alliance, these results indicate that the corpus contains examples of moderately successful counseling and behavior change, and should be suitable for developing models of behavior in similar scenarios.

CHAPTER 4

Speaking Rates and Coordination

My initial investigation of conversational behavior in long-term interaction focused on one aspect of verbal behavior in the exercise counseling corpus: changes in the articulation rate of speech, defined here in terms of the duration of words spoken, excluding any silence or pauses. In this chapter, I construct a model of changes in articulation rates, both within and across conversations, and present a preliminary evaluation in which some of the changes predicted by this model are incorporated in an ECA.

4.1 Related Work

A number of studies have examined the effects of speaking rates on listeners, focusing particularly on how changes in speaking rate affect a listener's perceptions of the speaker's personality. Smith et. al. found that increased speaking rate was perceived as increased competence, while perceived benevolence was highest at normal speech rates, and lower otherwise [153]. Nass and Lee showed that users perceived synthesized computer speech as more extroverted when it was generated with greater speech rate, volume, pitch, and pitch variation. In the same study, users tended to prefer speech perceived as matching their own introversion/extroversion [116].

Several researchers have examined differences in speaking rates, and other features of verbal behavior, usually with a cross-sectional design (comparing

friends to strangers or acquaintances). Planalp and Benson, in their comparison of audiotaped conversations between friends and strangers (discussed in Section 2.2.1), found that observers trying to discriminate between friends and strangers commonly used cues related to articulation rate (although a small percentage of all cited), such as “pace”, “tone of voice,” and “smoothness” [129]. Yuan et. al. compared several large corpora of conversational telephone speech in English and Chinese. Corpora consisting primarily of conversations between friends or family members had a higher average speaking rate than those consisting primarily of conversations between strangers [174].

Quené, in a study of a large corpus of spoken Dutch speech, reported that articulation rates tended to increase during conversations [135]. His corpus did not include any examples of multiple conversations between the same speakers, so did not analyze changes in articulation rates across conversations.

4.2 A Model of Articulation Rates

We¹ performed a full word-aligned orthographic transcription of the corpus, producing an estimate of the duration of every spoken word. The transcript was divided into pause groups (a sequence of words by a speaker uninterrupted by silence). Features of the pause groups a word belonged to were important predictors in the model, as discussed below. Note that a single turn or utterance by a speaker may contain multiple pause groups, if it included any intra-turn pauses.

In order to account for differences in word lengths, the duration of each word was normalized by the number of phonemes. The phonemes per word were determined using the CMU pronouncing dictionary (version 0.7a²), with manual correction for words that did not appear in the dictionary.

I used a linear mixed-effect model to account for the longitudinal nature of the data [166], analyzed with Bayesian methods using R [136] 2.10 and the

¹The transcription was performed by the myself and one additional naïve transcriber.

²<http://www.speech.cs.cmu.edu/cgi-bin/cmudict>

MCMCglmm package [70]. Uninformative or very weakly-informative prior distributions were used for all effects. For fixed effects, as an uninformative prior distribution, I use normal distributions with high variance (10^8). For random effects, since the number of dyads is small, and therefore sensitive to the choice of prior distribution, I use weakly informative half-Cauchy prior distributions [62].

To model change across conversations, the model includes a fixed effect of the number of previous sessions, while random effects allow for variability across subjects. Two covariates were motivated by prior work: (a) the position of a pause group within a conversation [135], and (b) the length of a pause group [115]. Inspection of preliminary models showed that predictions were poor for single-word pause groups (individual words bounded by silence); these had longer duration than predicted, even including group length as a covariate. Therefore, I included, as an additional predictor (“Multiword”), whether a word was in a multi-word pause group.

4.2.1 Results

Table 4.1 shows the full regression model. Word durations in later conversations are shorter than word durations in earlier conversations. However, this change was observed *only* for single-word pause groups (shown by the fixed effects “Session” and “Session \times Multiword”). Similarly, within conversations, words near the end of a conversation tended to be shorter, again largely for single word pause groups (shown by “Pos” and “Multiword \times Pos”). Figure 4.2 illustrates these changes; the solid lines indicate predicted articulation rates for a population-average speaker. The client spoke more slowly than the counselor (shown by the fixed effect “Who”), again largely for single word pause groups (shown by “Who \times Multiword”).

Given these results, I next examined the occurrences of single-word pause groups within the corpus (Table 4.2). The most common such words (“okay,”

Model Specification

$$\begin{aligned}
\log(y_{ij}) = & \beta_0 + u_{0j} + (\beta_1 + u_{1j}) \cdot \text{Session}_{ij} + (\beta_2 + u_{2j}) \cdot \text{Who}_{ij} \\
& + \beta_3 \cdot \text{Multiword}_{ij} + \beta_4 \cdot \text{Pos}_{ij} \\
& + \beta_5 \cdot \text{Session}_{ij} \cdot \text{Who}_{ij} + \beta_6 \cdot \text{Who}_{ij} \cdot \text{Multiword}_{ij} \\
& + \beta_7 \cdot \text{Session}_{ij} \cdot \text{Multiword}_{ij} \\
& + \beta_8 \cdot \text{Multiword}_{ij} \cdot \text{Len}_{ij} + \beta_9 \cdot \text{Multiword}_{ij} \cdot \text{Pos}_{ij} \\
& + \beta_{10} \cdot \text{Multiword}_{ij} \cdot \text{Len}_{ij} \cdot \text{Pos}_{ij} \\
& + \epsilon_{ij} \\
\begin{pmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{pmatrix} & \stackrel{iid}{\sim} N(0, \Sigma) \\
\epsilon_{ij} & \stackrel{iid}{\sim} N(0, \sigma_\epsilon^2)
\end{aligned}$$

Prior Distribution

$$\begin{aligned}
\beta_1, \dots, \beta_{10} & \sim N(0, 10^8) \\
\sigma_\epsilon^2 & \sim \text{Gamma}^{-1}(0.05, 0.05) \\
\Sigma & = \text{diag}(\varsigma_1, \varsigma_2, \varsigma_3) \cdot Q \cdot \text{diag}(\varsigma_1, \varsigma_2, \varsigma_3) \\
Q & \sim \text{Wishart}^{-1}(I_3, 3) \\
\varsigma_1, \varsigma_2, \varsigma_3 & \sim N(0, 100)
\end{aligned}$$

Predictors

Session_{ij}	# of previous sessions (starts at 0)
Who_{ij}	Interaction role: 0=counselor, 1=client
Pos_{ij}	$\frac{\text{pos}_{ij}-152.9}{109.4}$, where pos_{ij} is the sequential number of the enclosing pause group in the conversation
Len_{ij}	$\frac{\log(\text{len}_{ij})-2.23}{0.87}$, where len_{ij} is the length of the enclosing pause group in words
Multiword_{ij}	1 if in a pause group of more than 1 word, and 0 otherwise

Figure 4.1: A model of y_{ij} , the normalized (by # of phonemes) articulation rate of word i in dyad j .

Parameter		Est. ²	95% CI ^{1,2}
Fixed Effects	Intercept	-2.070	[-2.120,-2.021]***
	Session ³	-0.015	[-0.024,-0.005]**
	Who ⁴	0.100	[-0.002,0.176]*
	Multiword ⁵	-0.592	[-0.616,-0.564]***
	Pos ⁶	-0.045	[-0.057,-0.035]***
	Session \times Who	0.002	[-0.001,0.005]
	Who \times Multiword	-0.065	[-0.091,-0.037]***
	Session \times Multiword	0.012	[0.005,0.019]**
	Multiword \times Len ⁷	-0.105	[-0.108,-0.100]***
	Multiword \times Pos	0.038	[0.025,0.049]***
	Multiword \times Len \times Pos	-0.004	[-0.008,-0.001]*
Random Effects (SD)	Intercept	0.031	[0.017,0.087]
	Session	0.003	[0.001,0.011]
	Who	0.066	[0.034,0.208]
Random Effects (correlation)	Intercept,Session	-0.797	[-0.955,0.256]
	Intercept,Who	-0.073	[-0.693,0.607]
	Session,Who	0.194	[-0.537,0.830]
Residual		0.446	[0.444,0.448]

¹ *p<.05, **p<.01, ***p<.001

² Posterior mode and 95% credible interval.

³ Previously completed sessions (starts at zero).

⁴ 0=counselor, 1=client.

⁵ 1 if the word is part of a longer pause group, 0 otherwise.

⁶ Number of pause group within a conversation, centered and standardized (mean=152.9, SD=109.4).

⁷ Length of pause group in words, log-transformed, centered, and standardized (mean=2.23, SD=0.87).

Table 4.1: Mixed-effect regression model predicting articulation rates of words (average seconds per phoneme, log-transformed).

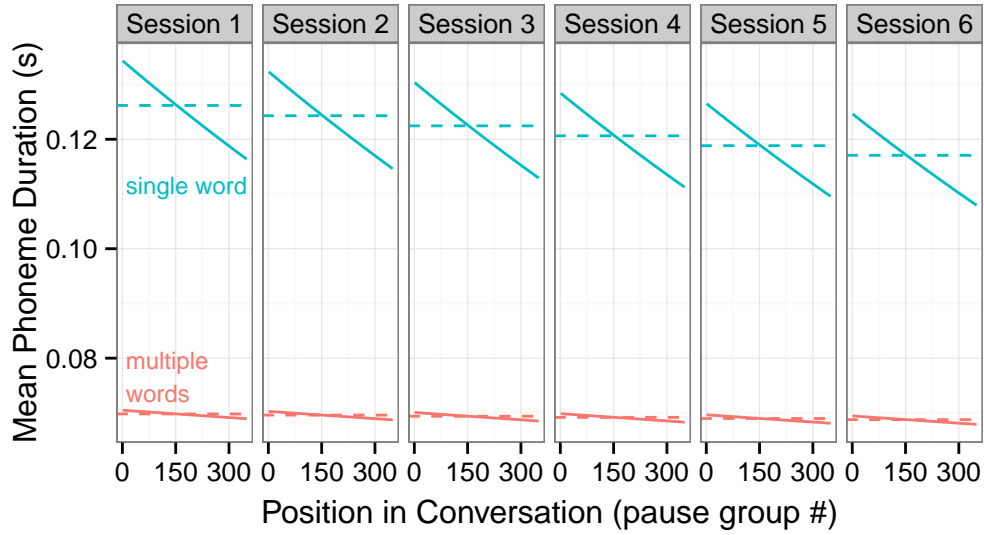


Figure 4.2: Predicted changes in articulation rate, for a (estimated) population-average speaker. The dotted lines indicate mean word duration within a conversation.

Word	Count	% of Total
Okay	857	15.9%
Yeah	817	15.1%
Mm-Hmm	640	11.9%
Um	612	11.3%
So	209	3.9%
Yknow	173	3.2%
And	131	2.4%
Right	112	2.1%
But	107	2.0%
Great	101	1.9%

Table 4.2: Common words appearing as single-word pause groups.

“yeah,” “mm-hmm,” “um,” “so,” “yknow,” “and,” “right,” “but,” and “great”, accounting for approximately 70% of all instances), appear to consist mainly of backchannels and acknowledgements (e.g., “okay,” “yeah”), and discourse markers (e.g., “so,” “and”) [146].

In sum, I observe that the durations of single-word acknowledgements and discourse markers decreased over time, both within a single conversation and across multiple conversations.

4.3 Discussion

The corpus shows evidence of changes in articulation rates over time, both within conversations and across multiple conversations. This complements results from earlier, cross-sectional studies, providing evidence that these changes in verbal behavior are in fact changes over time rather than pre-existing differences.

I also show a previously unreported nuance: increases in articulation rates were observed mainly in words bordered by silence, and these words were often acknowledgments or discourse markers. One possible explanation is that these words are the ones most easily spoken faster; longer pause groups already tend to have faster articulation rates [115]. Alternatively, markers such as “so” may be used by speakers to coordinate their interaction as a collaborative task [68]. Faster articulation rates may indicate a decrease in explicit coordination as speakers increase in familiarity.

CHAPTER 5

Posture and Discourse Structure

In this chapter, I focus on a single nonverbal behavior — posture shifts, defined here as gross movements of the body, including the trunk, the legs and lower body, or both – which is both a part of the standard repertoire of many ECAs, and may also be an indicator of the interpersonal relationship of a dyad. Following the approach developed while examining articulation rates (Chapter 4), I construct a statistical model of changes in the rate of posture shifts, including change both within and across conversations.

There is a well-established association between the occurrence of posture shifts in a conversation and the discourse structure (e.g., topic shifts) of that conversation; this association is discussed in more detail in the following section. While verbal and nonverbal behavior in conversation may vary in long-term interaction, discourse structure may also vary: for example, later conversations may tend to have more or fewer topic shifts than earlier conversations.

In order to examine changes in nonverbal behavior over time in the presence of possible changes in discourse structure, I construct a model that predicts the rate of posture shifts in conversation as a function of discourse structure, position within a conversation, and the number of previous conversations a dyad has had.

5.1 Background and Related Work

Several decades of research, dating at least to work by Schefflen [144], has focused on examining postural alignment or mirroring as an indicator of rapport. To the extent that a dyad is likely to build stronger rapport over multiple conversations, this predicts increasing postural alignment over time. However, empirical tests have been mixed [93, 94]. Bernieri reports that movement synchrony (i.e., similarity in timing) may be an indicator of rapport while behavior matching (i.e., taking the same position at the same time) is not [12]. Tickle-Degnen and Gavett suggest that postural alignment may have a positive association with rapport only in later interactions, rather than initial interactions [160].

In this study, I do not focus on alignment or mirroring. I focus instead on extending prior work which has established an association between posture shifts and discourse structure. Many authors have noted that posture shifts tend to occur at topic boundaries (e.g. [86]). Cassell and Nakano et al. give empirical evidence of this phenomenon, based on an examination of direction-giving dialogues [41].

Little prior work examines any possible association between posture shifts and aspects of interpersonal relationship other than rapport. To my knowledge, no prior empirical work examines simultaneously an association between posture shifts and discourse structure and an association between posture and interpersonal relationship.

5.2 Coding

The exercise counseling corpus was annotated to identify occurrences of posture shifts and of topic shifts independently, as two separate coding tasks. Both posture shifts and topic shifts were coded for both the counselor and the clients. Coding was performed by the author. To check reliability, three conversations were randomly selected for each coding task and analyzed by a

second (naïve) coder.

5.2.1 Coding of Topic Shifts

Topic shifts were coded using transcripts of the corpus produced when modeling changes in articulation rates (Chapter 4). Video was not viewed in order to avoid confounding topic shifts with visible posture shifts or other nonverbal indicators of discourse structure. The transcripts were segmented based on the occurrence of silence, and topic shifts were coded as occurring at the beginning of the segment that introduced a new topic.

Following Grosz and Sidner [68], we defined a topic as a shared conversational goal to which the participants were mutually committed. A topic shift was marked whenever the coder believed a participant was attempting to introduce such a shared goal (whether or not the attempt was successful). The agreement rate between coders was 96.4% (Cohen’s $\kappa = 0.68$).¹

The complete coding manual used to annotate topic shifts is included as Appendix B.

5.2.2 Coding of Posture Shifts

Posture shifts were coded using muted video, in order to avoid confounding posture shifts with audible topic shifts. Movements that appeared to be caused by the performance of a communicative gesture (e.g., a large hand gesture) were excluded, as were repetitive motions lasting more than a couple seconds (e.g., repeatedly rocking back and forth in the chair). Initially, chair rotation was coded as a posture shift, but preliminary examination revealed that these movements were very difficult to code reliably, thus they were excluded.

Coders were asked to judge the start and end times of each posture shift, and several additional features, including movement direction, co-occurrence

¹There is not a universally accepted convention for an acceptable level of Cohen’s κ [9]. For this dissertation, I have generally chosen to take $\kappa \geq 0.65$ as acceptable, and $\kappa \geq 0.55$ as minimally acceptable with reservations. For comparison, a widely-cited set of guidelines [96] specifies $\kappa > 0.6$ as “substantial”, and $\kappa > 0.4$ as “moderate”.

of grooming behavior (e.g., brushing hair or adjusting clothes), and an estimated energy level. Shifts were coded in continuous time (to the nearest video frame). Energy was judged on a linear scale ranging from 1 (the smallest perceptible shift) to 10 (the most energetic possible shift without leaving the chair).

An initial examination showed that reliability was very poor on low-energy shifts, and consequently all shifts with an energy of less than 4 were discarded. Aside from this, features of posture shifts besides the time of occurrence were not used in the present study. To compute inter-rater reliability, the corpus was divided into 1-second intervals, and each interval was considered to have been marked as a shift if the majority of it was covered by any coded posture shift. Cohen's κ was 0.58.²

The complete coding manual used to annotate posture shifts is included as Appendix C.

5.2.3 Results

Thirty-one conversations were coded for both posture and topic shifts; one conversation had large portions of unintelligible speech, and could not be coded for topic shifts. A total of 803 posture shifts were identified in the remaining conversations. The rate of posture shifts varied widely across conversations, ranging from 0.035 to 4.92 per minute (median 0.71).

5.3 A Model of Posture Shifts

The start time of each posture shift was aligned to the nearest second. I then modeled the occurrence of a posture shift as a binary outcome, with one observation per second. A logistic mixed-effect regression model was used. This

²This value of κ is relatively low, and may indicate noisy data that can cause attenuation or bias of results. Informal inspection of the mutually-coded videos indicated that many disagreements between coders concerned the start time of a shift rather than the existence of a shift.

model extends logistic regression to account for observations that are non-independent due to being grouped or nested — in this case, within conversations and dyads — by adding “random effects” which model the group-level variance.

The model was estimated with Bayesian methods, using R 2.15 and the MCMCglmm package [70]. Weakly-informative prior distributions were used: For fixed effects, I use normal distributions with high variance (10^8). For random effects, since the number of dyads is small, and therefore sensitive to the choice of prior distribution, I use weakly informative half-Cauchy prior distributions [62].

In order to model change over time, I include the number of previous sessions, and the time since the start of the conversation as predictors. To control for varying discourse structure, I include the co-occurrence of topic shifts, by both a speaker and their conversation partner, as predictors. To allow for variability among different speakers, dyads, and conversations, I include the speaker (counselor or client) as a predictor, along with random effects on dyads and individual conversations. Finally, I include two-way interactions among all predictors; a model comparison by the Deviance Information Criterion [156] preferred this more complex model ($\Delta\text{DIC} = 7.03$)

Table 5.1 shows the full regression model. Posture shifts were significantly more likely to occur at topic shifts (the coefficient “Tshift_{self}”); this replicates results by Cassell et al. [41]. There was no significant effect for topic shifts (“Tshift_{other}”) introduced by the conversation partner rather than a participant, although there was a trend in the same direction.

There were significant changes over time, both within and across conversations. Figure 5.2 and Figure 5.3 show the predicted probability of a posture shift (for a client and for the counselor, respectively) at different points over six weekly sessions, averaged over dyads. Each column shows a single session. Posture shifts occurred much more frequently in the beginning of a conversation, as indicated by the steeply downward sloping lines in each column

Model Specification

$$\begin{aligned}
y_{ijk} &\overset{iid}{\sim} \text{Bernoulli}(p_{ijk}) \\
\text{logit}(p_{ijk}) &= \beta_0 + u_j + v_k \\
&+ \beta_1 \cdot \text{Sessions}_{ijk} + \beta_2 \cdot \text{Speaker}_{ijk} + \beta_3 \cdot \text{Minutes}_{ijk} \\
&+ \beta_4 \cdot \text{Tshift}_{(\text{self})_{ijk}} + \beta_5 \cdot \text{Tshift}_{(\text{other})_{ijk}} \\
&+ \beta_6 \cdot \text{Sessions}_{ijk} \cdot \text{Speaker}_{ijk} + \beta_7 \cdot \text{Sessions}_{ijk} \cdot \text{Minutes}_{ijk} \\
&+ \beta_8 \cdot \text{Sessions}_{ijk} \cdot \text{Tshift}_{(\text{self})_{ijk}} + \beta_9 \cdot \text{Sessions}_{ijk} \cdot \text{Tshift}_{(\text{other})_{ijk}} \\
&+ \beta_{10} \cdot \text{Speaker}_{ijk} \cdot \text{Minutes}_{ijk} + \beta_{11} \cdot \text{Speaker}_{ijk} \cdot \text{Tshift}_{(\text{self})_{ijk}} \\
&+ \beta_{12} \cdot \text{Speaker}_{ijk} \cdot \text{Tshift}_{(\text{other})_{ijk}} + \beta_{13} \cdot \text{Minutes}_{ijk} \cdot \text{Tshift}_{(\text{self})_{ijk}} \\
&+ \beta_{14} \cdot \text{Minutes}_{ijk} \cdot \text{Tshift}_{(\text{other})_{ijk}} \\
&+ \beta_{15} \cdot \text{Tshift}_{(\text{self})_{ijk}} \cdot \text{Tshift}_{(\text{other})_{ijk}} \\
u_j &\overset{iid}{\sim} N(0, \sigma_{\text{dyad}}^2) \\
v_k &\overset{iid}{\sim} N(0, \sigma_{\text{conv}}^2)
\end{aligned}$$

Prior Distribution

$$\begin{aligned}
\beta_0, \dots, \beta_{15} &\sim N(0, 10^8) \\
\sigma_{(\text{dyad})}^2, \sigma_{(\text{conv})}^2 &\sim 1000 \cdot F(1, 1)
\end{aligned}$$

Predictors

Sessions_{ijk}	# of previous sessions (starts at 0)
Speaker_{ijk}	Interaction role: 0=counselor, 1=client
Minutes_{ijk}	Minutes from start of conversation
$\text{Tshift}_{(\text{self})_{ijk}}$	The participant introduced a topic shift within 2 seconds
$\text{Tshift}_{(\text{other})_{ijk}}$	The conversation partner introduced a topic shift within 2 seconds

Figure 5.1: A model of y_{ijk} , the occurrence of a posture shift by a participant in a 1-second window (i) in dyad j and conversation k .

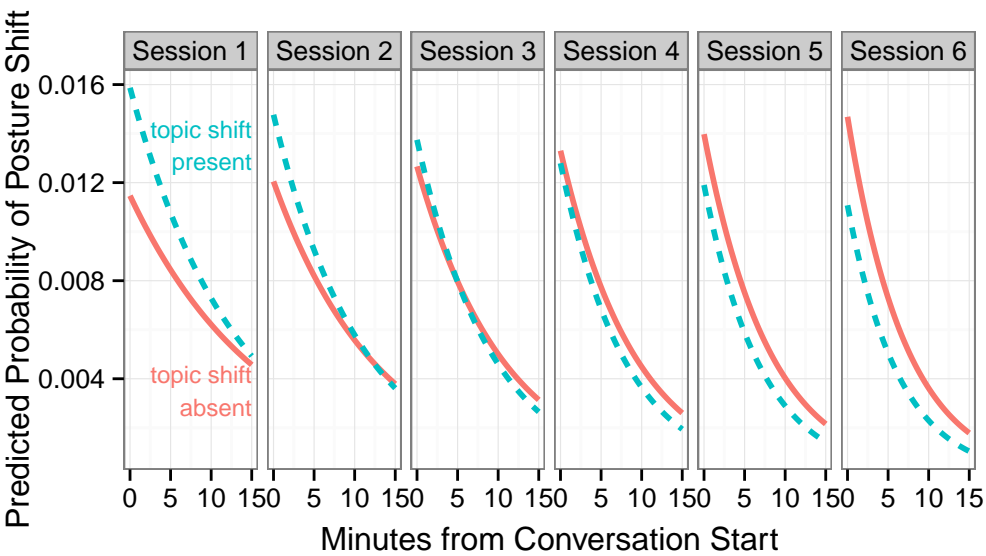


Figure 5.2: Predicted changes in rate of posture shifts for a client, in a (estimated) population-average speaker.

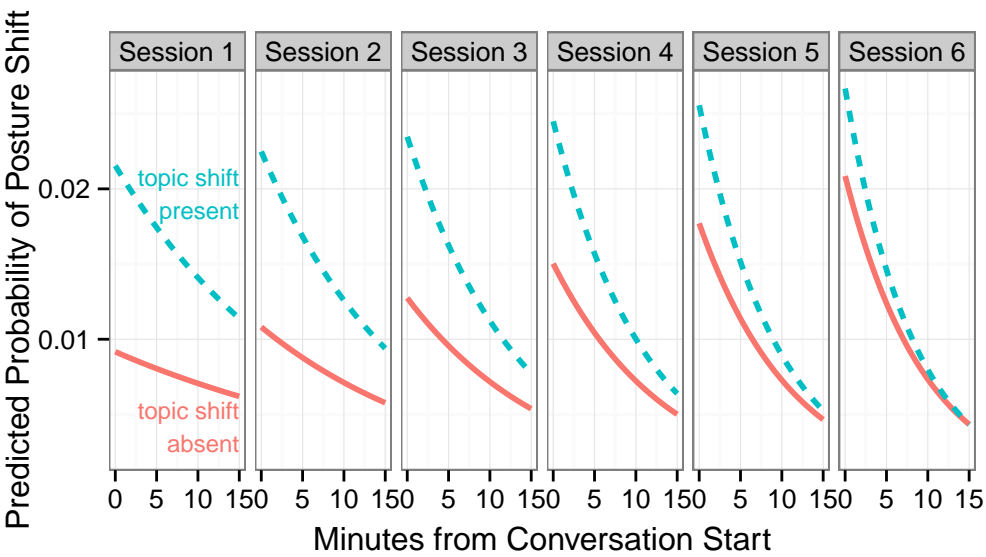


Figure 5.3: Predicted changes in rate of posture shifts for the counselor, in a (estimated) population-average speaker.

Parameter		Est. ²	95% CI ^{1,2}
Fixed Effects	Intercept	-4.683	[-4.101,-5.227]***
	Sessions ³	0.167	[0.053,0.271]**
	Speaker ⁴	0.226	[-0.205,0.554]
	Minutes ⁵	-0.026	[-0.051,-0.002]*
	Tshift _{self} ⁶	0.867	[0.338,1.648]**
	Tshift _{other} ⁶	0.368	[-0.148,1.053]
	Sessions \times Speaker	-0.117	[-0.199,-0.025]**
	Sessions \times Minutes	-0.016	[-0.026,-0.008]***
	Sessions \times Tshift _{self}	-0.123	[-0.256,0.065]
	Sessions \times Tshift _{other}	-0.125	[-0.289,0.037]
	Speaker \times Minutes	-0.036	[-0.073,-0.010]*
	Speaker \times Tshift _{self}	-0.538	[-1.103,-0.110]*
	Speaker \times Tshift _{other}	0.299	[-0.365,0.712]
	Minutes \times Tshift _{self}	-0.017	[-0.075,0.019]
	Minutes \times Tshift _{other}	-0.036	[-0.079,0.019]
	Tshift _{self} \times Tshift _{other}	0.272	[-1.404,1.393]
Random Intercept			
(SD)			
	Dyad	0.331	[0.155,0.869]
	Conversation	0.261	[0.158,0.412]

¹ *p<.05, **p<.01, ***p<.001

² Posterior mode and 95% credible interval.

³ # of previous conversations (starts at 0).

⁴ 0=client, 1=counselor.

⁵ Minutes from start of conversation.

⁶ A topic shift occurs within 2 seconds.

Table 5.1: Mixed-effect logistic regression model predicting the onset of a posture shift within a 1-second window.

(and the coefficient “Minutes” in Table 5.1). There was an interaction with the number of previous sessions: the rate of decrease was greater in later conversations (“Sessions \times Minutes”).

5.4 Discussion

I show evidence of changes in the occurrence of posture shifts over time, both within conversations and across multiple conversations. These changes appear to occur independently from previously-observed associations between posture shifts and discourse structure. While I observed that posture shifts were more

frequent in the presence of topic shifts (as previously reported), the rate of posture shifts decreased over time both in the presence and absence of topic shifts.

I do not yet have clear evidence for a mechanism explaining these effects, but instead offer some conjectures, based partially on subjective examination of the corpus. The early portion of many conversations included posture shifts that appeared to be part of a process of “settling in,” with most shifts leaving the participant in a more relaxed body posture. A relaxed body posture is an indicator of nonverbal immediacy [5] (i.e., intimacy, warmth, or closeness). Increasingly rapid decreases in the rate of posture shifts in later conversations may indicate that, as a stronger interpersonal relationship develops over time, participants will more quickly and easily adopt a body posture that indicates high immediacy. Future work may investigate these conjectures through more detailed coding of posture shifts and by examining other indicators of immediacy for similar patterns of change.

CHAPTER 6

Openings and Reopenings

In previous chapters I examined changes in conversational behavior in long-term, multi-conversation interaction, but restricted my focus to examining single aspects of conversational behavior: articulation rate (Chapter 4) and posture shifts (Chapter 5). I also examined only a single aspect of interpersonal relationship as a predictor of these behaviors: interaction history, operationalized as the number of prior conversations a dyad has had.

In the current chapter, I broaden my focus to examine changes in multiple types of verbal and nonverbal behaviors and to model such changes as a function of multiple aspects of interpersonal relationship: interaction history and the strength or quality of the interpersonal relationship. Interaction history includes the number, pattern, and purpose of the series of conversations two interactants have had. The strength or quality of an interpersonal relationship has been conceptualized in many ways, including such longitudinal constructs as trust, intimacy, and working relationship (e.g., therapeutic alliance in healthcare). These two variables — history and relationship — are related, but often separate factors in influencing the behavior of dyads over time [162].

I focus on behavior occurring within conversation openings, defined here as the first minute of conversations. Openings are a particularly important segment of conversation, in which effects of relationship status may be most pronounced. At the beginning of a conversation, participants' beliefs about

their interpersonal relationship may be communicated and/or negotiated [65]. The earlier modeling efforts (Chapter 4 and Chapter 5), when predicting differences in behavior across multiple conversations, showed that such differences were largest at the beginning of conversations.

6.1 Coding

The goal of the analysis was to identify whether there were systematic changes in counselor and client nonverbal behavior across conversations, as a function of interaction history (the number of conversations, and whether the current conversation is the last), relationship strength (measured by therapeutic alliance), or both. The analysis did not attempt to account for all variability in behavior, and substantial unexplained variability remains.

A one-minute segment of each conversation was selected, beginning from the first point at which both participants were judged to be fully seated; participants sat facing each other immediately after entering the room in all conversations. The resulting 32 minutes of video were manually annotated for various nonverbal behaviors (detailed below) using ANVIL [88]. The word-aligned orthographic transcription of the corpus, performed when examining articulation rates (Chapter 4), was used to identify segments where each participant was speaking.

6.1.1 Outcome Variables

I chose the following set of outcome variables for analysis based on those behaviors which prior work suggested might show changes associated with varying interpersonal relationship:

- *The proportion of time spent speaking:* friends are reported to share speaking time more equally than strangers [129].

- *The number of gaze-aways during speech*: the amount of gaze-away during speech is reported to be associated with topic intimacy [1].
- *The proportion of time, when not speaking, spent nodding*: friends are reported to use less nodding for acknowledgement than strangers [40]. Restricting to time when not speaking controls for varying opportunity to show acknowledgment in different videos.
- *The proportion of time spent smiling or frowning*, or more generally with the mouth in a non-neutral position: increased facial expressivity is associated with higher immediacy [5].
- *The proportion of time spent performing self-adaptors, when not speaking*: the use of self-adaptors — self-touching gestures that do not signal meaning in conversation, and often serve to release bodily tension — is associated with perceptions of anxiety [170]. A qualitative inspection of the corpus indicated that most self-adaptors occurred when not speaking.
- *The proportion of time spent performing gestures other than self-adaptors, during speech*: frequent and expressive gestures are associated with immediacy [5], and most hand gestures co-occur with speech.
- *The proportion of time spent with eyebrows raised or lowered, during speech*: eyebrows raises and frowns are a component of displays of affect and other facial expressivity, associated with immediacy.

Two additional aspects of behavior (shoulder movement and arm position) were annotated but were not used in subsequent analysis due to low interrater reliability.

A preliminary analysis indicated no significant changes in behavior within a single one-minute video. Therefore, all outcome variables are aggregates of behaviors over a video clip.

6.1.2 Development of a Coding Manual

A coding manual was developed iteratively, beginning with an initial draft intended to capture the outcome variables detailed above. The initial draft was heavily based on the MUMIN coding scheme [2], which has been previously used for annotation of nonverbal behavior with good reliability on most behaviors. Relative to MUMIN, the manual omits all coding related to communicative function and simplified the coding of facial expressions and hand gestures. It adds annotation of hand gestures used as self-adaptors, and modifies the annotation of posture.

The initial draft was refined by choosing at random 3 video clips, from 3 different dyads, which were coded by the author using the draft coding manual, and noting any behaviors that were ambiguous or difficult to code. After each clip was coded, the manual was revised to attempt to eliminate or mitigate any such problems. After the third test clip was coded, only minor revisions were made to the draft manual, and it was declared to be the final coding manual.

All 3 test clips were later recoded using the final coding manual, and only these final annotations are used in subsequent analysis.

6.1.3 Description

A brief summary of the annotation scheme follows; for full details, see the coding manual included as Appendix D. All event start and end times were coded in continuous time (to the nearest video frame).

Gaze An event was coded whenever a participant looked away from the partner's eyes, in any direction. The approximate direction of gaze was recorded as: up, down, or sideways (either left or right).

Eyebrows An event was coded whenever a participant raised or lowered his or her eyebrows away from a neutral facial expression. The position of the

eyebrows (raised, frown, or “other”) was recorded.

Head Movement An event was coded for any head movement which caused any part of the head to move at least two inches in any direction. However, nodding, shaking, and other rhythmic and repetitive movements were always coded, even when resulting in less than two inches of movement. Each event was categorized as one of: nod (up-and-down movement), jerk (single quick upward movement), back (movement away from the partner), forward (toward the partner), turn (rotation either left or right), or tilt (leaning to either side).

Mouth Shape An event was coded whenever a participant’s mouth took a shape that differed from a neutral facial expression (e.g., corners up or down, lips protruded or retracted), other than to open during speech.

Hand Gesture Based on semiotic categories as described by McNeill [108], gestures were coded as deictic (pointing or other indications of physical location), iconic (a representational gesture, indicating meaning by similarity in shape, trajectory, or speed), emblematic (a gesture with meaning based on social convention), beat (a gesture which marks emphasis or timing in speech, and appears to have no additional communicative function), self-adaptor (idiosyncratic gestures which appear to relieve bodily tension), or “other.”

Shoulders An event was coded whenever a participant raised his or her shoulders, as in a shrug. We noted whether a single shoulder or both were raised.

Arm Position The open-closed position of the arms was coded continuously, in segments that covered an entire video clip. Openness was defined in terms of the lateral position of the upper arms and elbows relative to the arms of the chair: an arm was “closed” if inside the chair arm, “open” if at the chair arm, and “back” if outside the chair arm.

Table 6.1: Interrater reliability for coding of nonverbal behavior

Behavior	Cohen's κ
Gaze-away	0.71
Eyebrows	0.65
<i>Head movement (occurrence)^a</i>	<i>0.68</i>
Head movement (categorized)	0.67
Mouth	0.81
<i>Gesture (occurrence)^a</i>	<i>0.71</i>
<i>Gesture (categorized)^a</i>	<i>0.57</i>
Gesture (self-adaptor) ^b	0.91
<i>Shoulder movement^a</i>	0.44
<i>Arm position^a</i>	0.85

^a not used in subsequent analysis^b a composite of several categories

6.1.4 Interrater Reliability

Three randomly-selected videos were chosen, containing sessions with three different clients and not previously used during the development of the coding manual. These were coded separately by the author and by a second coder who was not involved in the development of the coding manual. The start and end times of all coded events were aligned to the nearest quarter second, and Cohen's κ was computed, treating a quarter second segment as one observation. Reliability was considered acceptable when $\kappa \geq 0.65$.

Table 6.1 summarizes the results. For head movement and hand gesture, Cohen's κ is reported separately for coding the occurrence of an event at the same time, and for coding the same event category at the same time. Reliability was low for categorized hand gestures: beat, iconic, and deictic gestures were all frequently confused, and emblematic gestures were rare. Combining all categories other than "adaptor" yielded good reliability, and all subsequent analysis uses only the categories of self-adaptor and non-adaptor gestures.

Shoulder movement was not coded reliably ($\kappa = 0.44$); inspection of the videos indicated that this was due to difficulty in identifying shoulder movement in the presence of other bodily movement, such as large gestures or

	Counselor		Client		Overall	
	Mean	SD	Mean	SD	Mean	SD
Mouth ^a	0.35	0.18	0.37	0.24	0.36	0.21
Gaze-Away ^b	0.04	0.06	0.11	0.08	0.08	0.08
Nod ^c	0.24	0.08	0.14	0.10	0.19	0.10
Speech ^a	0.20	0.10	0.57	0.12	0.39	0.21
Adaptor ^c	0.09	0.07	0.11	0.18	0.10	0.14
Gesture ^d	0.04	0.04	0.12	0.11	0.08	0.09
Brows ^d	0.02	0.05	0.10	0.09	0.06	0.08

^a Proportion of time

^b Count of events during speech

^c Proportion of time not speaking

^d Proportion of time speaking

Table 6.2: Descriptive statistics for all behaviors

postural shifts.

6.1.5 Descriptive Results

Table 6.2 gives descriptive statistic for each behavior of interest, aggregated over all conversations. Clients show greater variability than the counselor on all behavioral variables, and there are large differences between the counselor and client means on several behaviors.

6.2 Models of Behavior in Conversation Openings

For each outcome variable listed in 6.1.1, I fit a series of regression models. Ordinary linear regression is inappropriate here, as it assumes that observations are independent, whereas here conversations are grouped within dyads. I use generalized linear mixed-effect regression [107], an extension of ordinary linear regression that accounts for grouped data by adding “random effects,” or per-group means that are assumed to be normally distributed around the population mean. The models treat counselor and client behavior as sepa-

rate but correlated per-conversation outcome variables, and include separate per-dyad means for the counselor and client, which may also be correlated.

I considered four variants of this model for each behavior, differing in the set of predictors included. All models include two predictors, modeling change over time: the number of prior conversations, and whether the current conversation is the last conversation for that dyad.¹ From this basic model, I consider:

- A. The predictors above, with the added assumption that the effect of these predictors is the same on the counselor and the client.
- B. As in A, and including self-reported therapeutic alliance from the previous conversation.
- C. As in A, but with no assumption that effects on the client and counselor are the same.
- D. As in B, but with no assumption that effects on the client and counselor are the same.

The number of gaze-aways during speech was modeled as a Poisson-distributed count outcome with an added Gaussian random effect to allow for overdispersion [22]. For all other behaviors, the proportion of time during which the behavior was observed was modeled as a Gaussian-distributed outcome, following two transformations: first, by “squeezing” all values toward 0.5 slightly to avoid proportions exactly equal to 0 or 1 [155], and then by applying the logit function:

$$y' = \text{logit} \left(\frac{y * (N - 1) + 0.5}{N} \right)$$

where $N = 64$ is the total number of observations.

¹This predictor was added after observing that the final sessions appeared qualitatively different from others in the corpus. Omitting these sessions gives estimates similar to those reported here, although the interaction of therapeutic alliance and number of conversations on nodding (Figure 6.6) is only near-significant.

Model Specification

$$\begin{aligned}
\text{logit} \left(\frac{63 \cdot a_{ij} + 0.5}{64} \right) &= \beta_{a0} + u_{aj} + \beta_1 \cdot \text{Session}_{ij} + \beta_2 \cdot \text{Last}_{ij} + \epsilon_{aij} \\
\text{logit} \left(\frac{63 \cdot b_{ij} + 0.5}{64} \right) &= \beta_{b0} + u_{bj} + \beta_1 \cdot \text{Session}_{ij} + \beta_2 \cdot \text{Last}_{ij} + \epsilon_{bij} \\
\begin{pmatrix} u_{aj} \\ u_{bj} \end{pmatrix} &\stackrel{iid}{\sim} N(0, \Sigma_{\text{dyad}}) \\
\begin{pmatrix} \epsilon_{aij} \\ \epsilon_{bij} \end{pmatrix} &\stackrel{iid}{\sim} N(0, \Sigma_{\text{conv}})
\end{aligned}$$

Prior Distribution

$$\begin{aligned}
\beta_{a0}, \beta_{b0}, \beta_1, \beta_2 &\sim N(0, 10^{10}) \\
\Sigma_{\text{dyad}}, \Sigma_{\text{conv}} &\sim \text{Wishart}^{-1}(I_2, 3)
\end{aligned}$$

Predictors

Sessions_{ij} # of previous sessions (starts at 0)
 Last_{ij} 1 if this is a final (6th) session, 0 otherwise

Figure 6.1: Model A of a_{ij} and b_{ij} , the proportion of time the counselor or client, respectively, perform a behavior in conversation i of dyad j .

The models were fit to the data using Bayesian estimation with weak prior distributions: normal distributions with high variance (10^{10}) for fixed effects of parameters and inverse Wishart distributions (3 d.f.) for the dyad-level and conversation-level covariance matrices. Models were compared using the Deviance Information Criterion (DIC) [156].

Figure 6.1 and Figure 6.2 summarize the models for all outcome variables other than the number of gaze-aways during speech. Figure 6.3 summarizes model A applied to the number of gaze aways; Models B,C, and D are constructed analogously.

Gaze-aways, nodding, and smiling and frowning were best predicted by models in which interaction history and relationship strength have the same

Model B

$$\begin{aligned}
\text{logit} \left(\frac{63 \cdot a_{ij} + 0.5}{64} \right) &= \beta_{a0} + u_{aj} + \beta_1 \cdot \text{Session}_{ij} + \beta_2 \cdot \text{Last}_{ij} \\
&\quad + \beta_3 \cdot \text{Alliance}_{ij} + \beta_4 \cdot \text{Session}_{ij} \cdot \text{Alliance}_{ij} + \epsilon_{aij} \\
\text{logit} \left(\frac{63 \cdot b_{ij} + 0.5}{64} \right) &= \beta_{b0} + u_{bj} + \beta_1 \cdot \text{Session}_{ij} + \beta_2 \cdot \text{Last}_{ij} \\
&\quad + \beta_3 \cdot \text{Alliance}_{ij} + \beta_4 \cdot \text{Session}_{ij} \cdot \text{Alliance}_{ij} + \epsilon_{aij}
\end{aligned}$$

Model C

$$\begin{aligned}
\text{logit} \left(\frac{63 \cdot a_{ij} + 0.5}{64} \right) &= \beta_{a0} + u_{aj} + \beta_{a1} \cdot \text{Session}_{ij} + \beta_{a2} \cdot \text{Last}_{ij} + \epsilon_{aij} \\
\text{logit} \left(\frac{63 \cdot b_{ij} + 0.5}{64} \right) &= \beta_{b0} + u_{bj} + \beta_{b1} \cdot \text{Session}_{ij} + \beta_{b2} \cdot \text{Last}_{ij} + \epsilon_{aij}
\end{aligned}$$

Model D

$$\begin{aligned}
\text{logit} \left(\frac{63 \cdot a_{ij} + 0.5}{64} \right) &= \beta_{a0} + u_{aj} + \beta_{a1} \cdot \text{Session}_{ij} + \beta_{a2} \cdot \text{Last}_{ij} \\
&\quad + \beta_{a3} \cdot \text{Alliance}_{ij} + \beta_{a4} \cdot \text{Session}_{ij} \cdot \text{Alliance}_{ij} + \epsilon_{aij} \\
\text{logit} \left(\frac{63 \cdot b_{ij} + 0.5}{64} \right) &= \beta_{b0} + u_{bj} + \beta_{b1} \cdot \text{Session}_{ij} + \beta_{b2} \cdot \text{Last}_{ij} \\
&\quad + \beta_{b3} \cdot \text{Alliance}_{ij} + \beta_{b4} \cdot \text{Session}_{ij} \cdot \text{Alliance}_{ij} + \epsilon_{aij}
\end{aligned}$$

Predictors

Sessions_{ij}	# of previous sessions (starts at 0)
Last_{ij}	1 if this is a final (6 th) session, 0 otherwise
Alliance_{ij}	WAI-SR score at previous assessment, centered and standardized

Figure 6.2: Models B, C, and D of a_{ij} and b_{ij} , the proportion of time the counselor or client, respectively, perform a behavior. Variance components (σ , u) and prior distributions are unchanged from Model A, and are not shown.

Model Specification

$$\begin{aligned}
a_{ij} &\stackrel{iid}{\sim} \text{Pois}(\lambda_{a_{ij}}) \\
\log(\lambda_{a_{ij}}) &= \beta_{a0} + u_{aj} + \beta_1 \cdot \text{Session}_{ij} + \beta_2 \cdot \text{Last}_{ij} \\
&\quad + \log(\text{SpeechDur}_{a_{ij}}) + \epsilon_{a_{ij}} \\
b_{ij} &\stackrel{iid}{\sim} \text{Pois}(\lambda_{b_{ij}}) \\
\log(\lambda_{b_{ij}}) &= \beta_{b0} + u_{bj} + \beta_1 \cdot \text{Session}_{ij} + \beta_2 \cdot \text{Last}_{ij} \\
&\quad + \log(\text{SpeechDur}_{b_{ij}}) + \epsilon_{b_{ij}} \\
\begin{pmatrix} u_{aj} \\ u_{bj} \end{pmatrix} &\stackrel{iid}{\sim} N(0, \Sigma_{\text{dyad}}) \\
\begin{pmatrix} \epsilon_{a_{ij}} \\ \epsilon_{b_{ij}} \end{pmatrix} &\stackrel{iid}{\sim} N(0, \Sigma_{\text{conv}})
\end{aligned}$$

Prior Distribution

$$\begin{aligned}
\beta_{a0}, \beta_{b0}, \beta_1, \beta_2 &\sim N(0, 10^{10}) \\
\Sigma_{\text{dyad}}, \Sigma_{\text{conv}} &\sim \text{Wishart}^{-1}(I_2, 3)
\end{aligned}$$

Predictors

Sessions _{ij}	# of previous sessions (starts at 0)
Last _{ij}	1 if this is a final (6 th) session, 0 otherwise
SpeechDur _{a_{ij}}	Total duration of speech (counselor)
SpeechDur _{b_{ij}}	Total duration of speech (client)

Figure 6.3: Model A of a_{ij} and b_{ij} , the number of times the counselor or client, respectively, gazes away during speech in conversation i of dyad j .

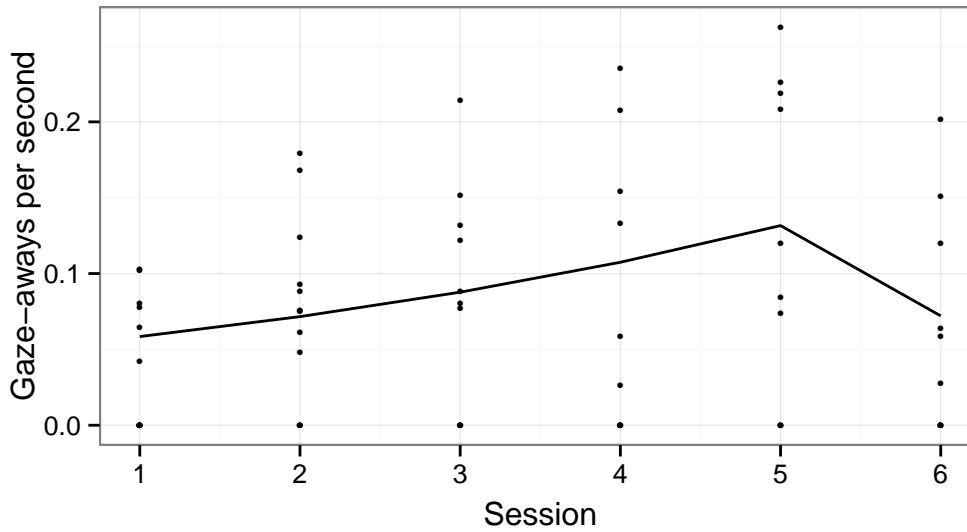


Figure 6.4: Gaze-aways during speech, by session. There were multiple observations of no gaze-aways by a participant: 6 in first sessions, and 3–4 in each of second through last sessions. The line is the model-based prediction for the average participant.

effect on counselor and client behavior. However, the proportion of time spent speaking, the use of both self-adaptor and non-adaptor gestures, and eyebrow raises and frowns were best predicted by models in which effects on the counselor and client differed.

6.2.1 Gaze

The number of gaze-aways during speech is predicted about equally well by models with and without therapeutic alliance (Table 6.3), although both give similar predictions: There is an increase in the rate of gaze-aways over time, which reverses in the last session (Figure 6.4).

6.2.2 Mouth

Participants used fewer non-neutral mouth positions in later sessions (Table 6.4), but reversed this trend in the last session (Figure 6.5).

Model A		Model B	Model C	Model D
DIC		195.70	195.84	198.15
Intercept	Counselor	-3.74 [-4.55,-2.79]	-4.17 [-5.18,-3.10]	-4.08 [-5.37,-3.02]
	Client	-2.84 [-3.71,-2.07]	-2.85 [-3.64,-1.96]	-2.76 [-3.63,-1.96]
Session	Counselor	0.20 [0.03,0.38]	0.28 [0.07,0.47]	0.37 [-0.01,0.79]
	Client			0.17 [-0.02,0.35]
Last	Counselor	-0.80 [-1.59,-0.01]	-0.91 [-1.71,-0.08]	-1.68 [-3.96,0.39]
	Client			-0.67 [-1.53,0.22]
Alliance	Counselor		-0.32 [-1.11,0.38]	-0.62 [-1.97,0.64]
	Client			-0.30 [-1.00,0.46]
All \times Sess	Counselor		-0.01 [-0.23,0.18]	0.03 [-0.43,0.45]
	Client			-0.02 [-0.13,0.08]
σ_{dyad}	Counselor	0.62 [0.25,1.06]	0.62 [0.27,1.10]	0.62 [0.29,1.08]
	Client	0.78 [0.32,1.36]	0.81 [0.36,1.40]	0.78 [0.35,1.34]
ρ_{dyad}	Counselor	-0.02 [-0.79,0.76]	-0.09 [-0.87,0.69]	-0.03 [-0.81,0.74]
	Client			-0.11 [-0.88,0.70]
σ_{conv}	Counselor	0.56 [0.28,0.92]	0.56 [0.27,0.93]	0.58 [0.28,0.95]
	Client	0.40 [0.24,0.59]	0.41 [0.24,0.59]	0.41 [0.25,0.60]
ρ_{conv}	Counselor	0.03 [-0.60,0.59]	0.01 [-0.59,0.59]	0.02 [-0.56,0.64]
	Client			0.00 [-0.58,0.64]

Table 6.3: Mixed-effect regression models predicting the rate of gaze-aways during speech. Posterior means and 95% credible intervals are shown. Bold values indicate a credible interval which excludes zero ($p \leq 0.05$)

	Model A	Model B	Model C	Model D
DIC	17.66	20.39	21.45	23.17
Intercept	Counselor Client	-0.43 [-1.20,0.31] -0.37 [-1.28,0.59]	-0.52 [-1.23,0.28] -0.43 [-1.37,0.50]	-0.42 [-1.15,0.34] -0.18 [-1.21,0.81]
Session	Counselor Client	-0.16 [-0.31,0.00] -0.14 [-0.30,0.01]	-0.16 [-0.33,-0.02] -0.27 [-0.54,-0.01]	-0.14 [-0.30,0.02] -0.19 [-0.52,0.12]
Last	Counselor Client	0.97 [0.25,1.61] 1.04 [0.29,1.71]	0.99 [0.30,1.76] 1.36 [0.15,2.54]	1.18 [0.43,1.91] 1.41 [0.10,2.74]
Alliance	Counselor Client	-0.19 [-0.52,0.16]		-0.27 [-0.62,0.08] -0.29 [-1.06,0.41]
All×Sess	Counselor Client	0.06 [-0.03,0.14]		0.12 [0.01,0.24] 0.05 [-0.08,0.20]
σ_{dyad}	Counselor Client	0.78 [0.37,1.30] 0.95 [0.40,1.62]	0.76 [0.38,1.27] 0.94 [0.43,1.62]	0.77 [0.37,1.31] 0.93 [0.38,1.60]
ρ_{dyad}		0.60 [0.04,0.97]	0.58 [0.02,0.97]	0.59 [0.04,0.97]
σ_{counv}	Counselor Client	0.56 [0.41,0.73] 0.96 [0.72,1.22]	0.56 [0.41,0.73] 0.96 [0.72,1.24]	0.57 [0.41,0.73] 0.97 [0.72,1.25]
ρ_{counv}		0.66 [0.44,0.85]	0.66 [0.44,0.85]	0.66 [0.43,0.86]
				0.67 [0.45,0.89]

Table 6.4: Mixed-effect regression models predicting the proportion of time (logit-transformed) displaying non-neutral mouth positions. Posterior means and 95% credible intervals are shown. Bold values indicate a credible interval which excludes zero ($p \leq 0.05$)

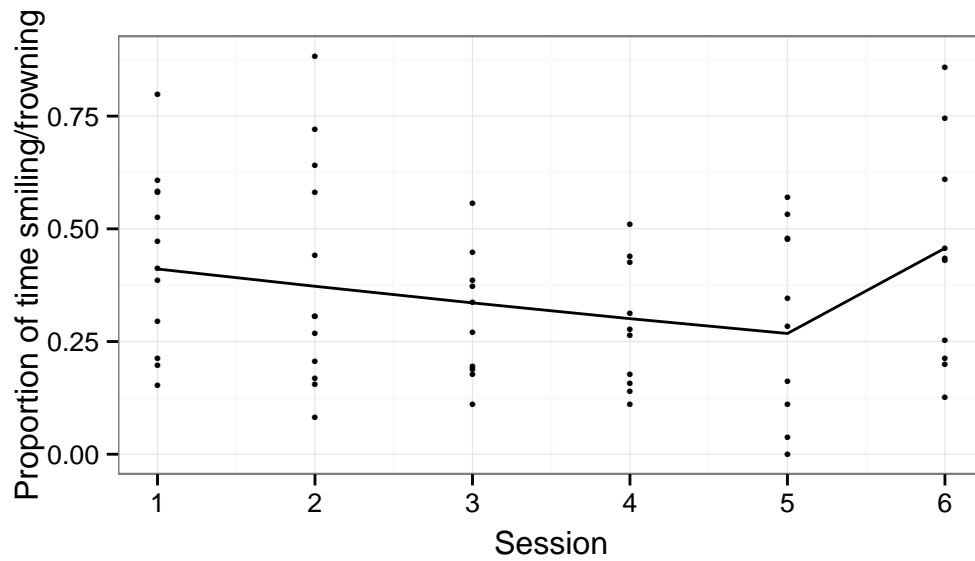


Figure 6.5: Occurrence of non-neutral mouth positions, by session. The line is the model-based prediction for the average participant.

6.2.3 Nodding

The proportion of time spent nodding was best predicted by a model which included therapeutic alliance (Table 6.5). Participants nodded less when they reported higher therapeutic alliance, but this effect was moderated by the number of sessions: in later sessions, all participants tended to nod more. There was a non-significant trend toward less nodding in the last session (Figure 6.6).

6.2.4 Speech

The proportion of time spent speaking was best predicted by a model which allowed different trends for the counselor and the clients (Table 6.6). The counselor spoke less in later sessions, and the clients spoke more, but this reversed in the last session for the counselor and there was a non-significant trend for it to reverse for the client (Figure 6.7).

	Model A	Model B	Model C	Model D
DIC	26.90	19.72	28.94	27.94
Intercept	Counselor Client	-1.24 [-1.76,-0.72] -2.26 [-3.14,-1.41]	-1.48 [-2.05,-0.97] -2.30 [-3.16,-1.48]	-1.21 [-1.79,-0.72] -2.35 [-3.27,-1.32]
Session	Counselor Client	0.03 [-0.09,0.14]	0.06 [-0.05,0.17]	0.03 [-0.11,0.16] 0.03 [-0.22,0.29]
Last	Counselor Client	-0.45 [-0.98,0.11]	-0.44 [-1.01,0.07]	-0.63 [-1.27,-0.03] 0.14 [-1.02,1.41]
Alliance	Counselor Client	-0.28 [-0.46,-0.10]		-0.21 [-0.40,-0.02] -0.33 [-0.78,0.13]
All×Sess	Counselor Client	0.06 [0.01,0.11]		0.04 [-0.02,0.10] 0.02 [-0.05,0.08]
σ_{dyad}	Counselor Client	0.52 [0.25,0.85] 0.91 [0.39,1.54]	0.53 [0.27,0.87] 0.91 [0.38,1.55]	0.51 [0.26,0.86] 0.92 [0.39,1.56]
ρ_{dyad}		0.23 [-0.49,0.84]	0.23 [-0.50,0.85]	0.22 [-0.49,0.83]
σ_{counv}	Counselor Client	0.48 [0.36,0.61] 0.93 [0.70,1.20]	0.45 [0.32,0.58] 0.87 [0.65,1.13]	0.48 [0.36,0.62] 0.93 [0.69,1.21]
ρ_{counv}		-0.02 [-0.42,0.32]	0.07 [-0.29,0.46]	-0.02 [-0.39,0.35]
				0.03 [-0.39,0.39]

Table 6.5: Mixed-effect regression models predicting the proportion of time (logit-transformed) spent nodding when not speaking. Posterior means and 95% credible intervals are shown. Bold values indicate a credible interval which excludes zero ($p \leq 0.05$)

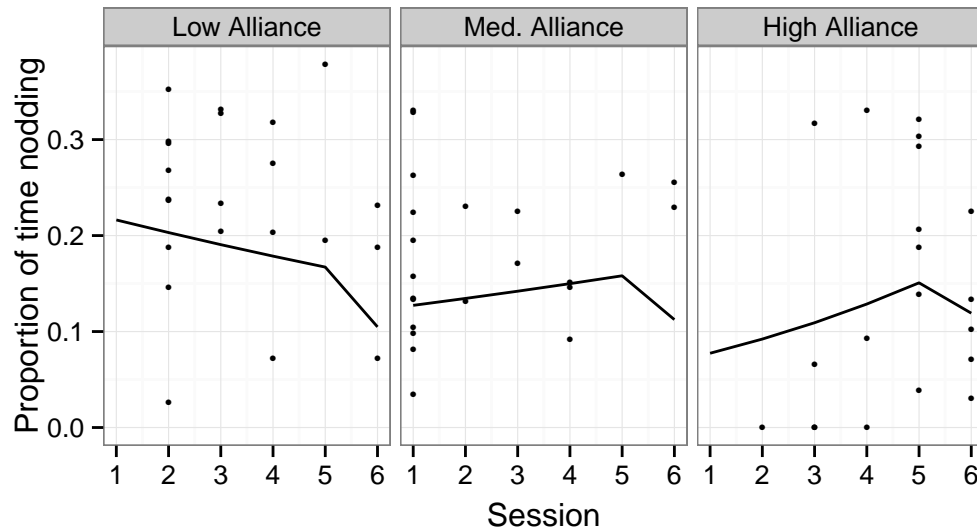


Figure 6.6: Nodding when not speaking, by session and alliance. The lines are model-based predictions for the average participant, at 25th, 50th, and 75th percentile therapeutic alliance.

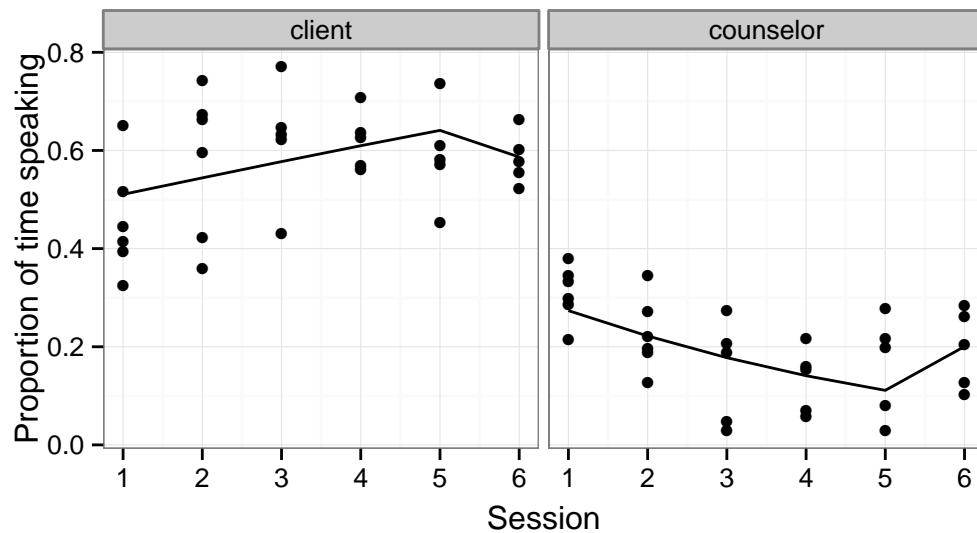


Figure 6.7: Proportion of time speaking, by session and role. The lines are the model-based predictions for the estimated average dyad.

	Model A	Model B	Model C	Model D	
DIC	-27.31	-24.63	-33.00	-27.69	
Intercept	Counselor	-1.46 [-2.04,-0.94]	-1.50 [-2.03,-0.90]	-1.00 [-1.61,-0.35]	-1.01 [-1.73,-0.34]
	Client	0.30 [-0.22,0.79]	0.28 [-0.25,0.74]	0.04 [-0.46,0.58]	0.01 [-0.53,0.55]
Session	Counselor	-0.01 [-0.07,0.06]	-0.00 [-0.08,0.07]	-0.27 [-0.46,-0.09]	-0.26 [-0.44,-0.07]
	Client			0.13 [0.03,0.24]	0.16 [0.04,0.29]
Last	Counselor	0.09 [-0.18,0.39]	0.07 [-0.23,0.35]	0.93 [0.11,1.80]	1.02 [0.15,1.90]
	Client			-0.36 [-0.88,0.12]	-0.44 [-0.97,0.06]
Alliance	Counselor		-0.17 [-0.48,0.13]		-0.08 [-0.55,0.39]
	Client				-0.13 [-0.46,0.17]
All×Sess	Counselor		0.05 [-0.05,0.14]		0.06 [-0.08,0.22]
	Client				-0.03 [-0.09,0.03]
σ_{dyad}	Counselor	0.56 [0.27,0.93]	0.56 [0.28,0.94]	0.55 [0.27,0.92]	0.57 [0.28,0.97]
	Client	0.54 [0.27,0.89]	0.54 [0.28,0.90]	0.54 [0.28,0.89]	0.56 [0.29,0.94]
ρ_{dyad}		-0.03 [-0.72,0.63]	-0.03 [-0.71,0.67]	-0.00 [-0.65,0.71]	-0.06 [-0.75,0.62]
σ_{conv}	Counselor	0.73 [0.55,0.93]	0.74 [0.56,0.96]	0.65 [0.49,0.84]	0.66 [0.49,0.86]
	Client	0.42 [0.31,0.54]	0.43 [0.32,0.56]	0.39 [0.29,0.50]	0.38 [0.29,0.50]
ρ_{conv}		-0.68 [-0.86,-0.47]	-0.70 [-0.87,-0.48]	-0.62 [-0.83,-0.38]	-0.61 [-0.85,-0.36]

Table 6.6: Mixed-effect regression models predicting the proportion of time (logit-transformed) spent speaking. Posterior means and 95% credible intervals are shown. Bold values indicate a credible interval which excludes zero ($p < 0.05$)

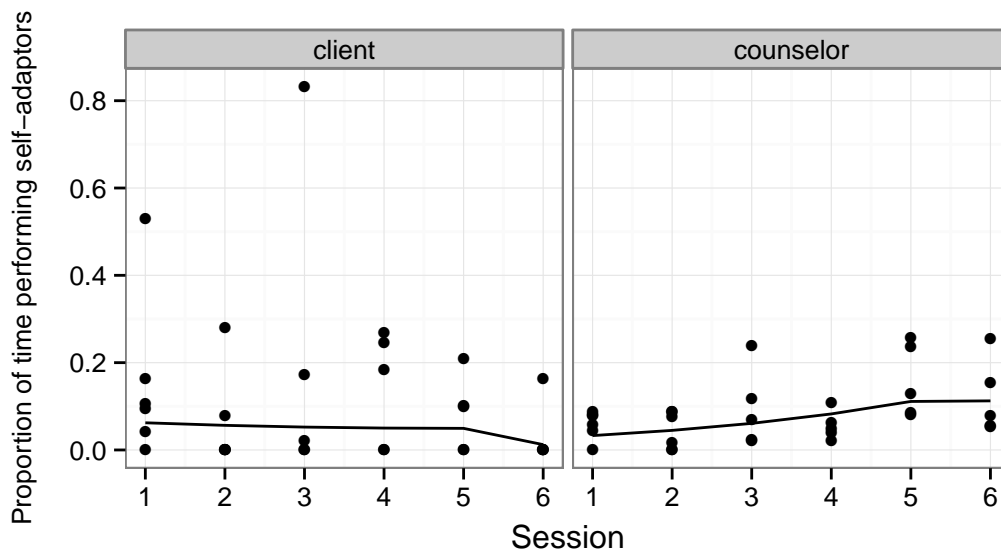


Figure 6.8: Occurrence of self-adaptors, by session and role. The lines are the model-based predictions for the estimated average dyad.

6.2.5 Self-Adaptors

The proportion of time spent using self-adaptors during speech was best predicted by a model which allowed different trends for the counselor and the clients (Table 6.7). The counselor used more self-adaptors in later sessions, but there was no significant trend for clients (Figure 6.8).

6.2.6 Hand Gesture

The proportion of time spent performing hand gestures (other than self-adaptors) was best predicted by a model which allowed different trends for the counselor and the clients (Table 6.8). The counselor performed fewer non-adaptor gestures in later conversations, but reversed this trend in the last session (Figure 6.9). There was no significant trend for clients.

6.2.7 Eyebrow Movement

The proportion of time spent performing eyebrow raises or lowering was predicted about equally well (or equally poorly) by every model (Table 6.9). No

	Model A	Model B	Model C	Model D	
DIC	104.27	105.36	98.15	105.36	
Intercept	Counselor	-2.97 [-3.73,-2.24]	-2.72 [-3.54,-1.89]	-3.24 [-3.97,-2.44]	-3.10 [-3.97,-2.23]
	Client	-3.42 [-4.67,-2.28]	-3.32 [-4.54,-2.01]	-2.78 [-4.14,-1.39]	-2.66 [-3.97,-1.41]
Session	Counselor	0.19 [-0.03,0.42]	0.14 [-0.08,0.38]	0.29 [0.04,0.54]	0.28 [0.01,0.56]
	Client			-0.06 [-0.40,0.33]	-0.26 [-0.81,0.20]
Last	Counselor	-0.57 [-1.62,0.42]	-0.61 [-1.67,0.34]	-0.31 [-1.47,0.90]	-0.27 [-1.51,1.03]
	Client			-1.15 [-2.86,0.60]	-0.79 [-2.77,1.02]
Alliance	Counselor		0.29 [-0.09,0.65]		0.13 [-0.27,0.52]
	Client				0.49 [-0.19,1.16]
All×Sess	Counselor		-0.08 [-0.18,0.03]		-0.01 [-0.13,0.10]
	Client				0.04 [-0.06,0.14]
σ_{dyad}	Counselor	0.56 [0.26,0.94]	0.58 [0.28,0.98]	0.56 [0.27,0.94]	0.59 [0.27,0.99]
	Client	1.14 [0.46,1.99]	1.13 [0.41,1.98]	1.21 [0.48,2.07]	0.98 [0.34,1.85]
ρ_{dyad}		0.04 [-0.71,0.79]	0.04 [-0.76,0.77]	0.03 [-0.72,0.79]	0.00 [-0.77,0.76]
σ_{conv}	Counselor	0.97 [0.71,1.26]	0.99 [0.73,1.28]	0.93 [0.71,1.19]	0.95 [0.71,1.22]
	Client	1.47 [1.07,1.94]	1.41 [1.03,1.85]	1.35 [0.99,1.75]	1.40 [1.00,1.80]
ρ_{conv}		0.01 [-0.34,0.38]	-0.04 [-0.42,0.35]	0.10 [-0.26,0.46]	-0.01 [-0.39,0.37]

Table 6.7: Mixed-effect regression models predicting the proportion of time (logit-transformed) spent performing self-adaptors during speech. Posterior means and 95% credible intervals are shown. Bold values indicate a credible interval which excludes zero ($p \leq 0.05$)

		Model A		Model B		Model C		Model D	
DIC		86.26		85.68		83.27		87.16	
Intercept	Counselor	-3.27	[-4.07,-2.58]	-3.26	[-4.05,-2.44]	-2.85	[-3.68,-2.12]	-3.07	[-3.93,-2.16]
	Client	-2.14	[-3.03,-1.30]	-2.25	[-3.12,-1.39]	-2.75	[-3.80,-1.80]	-2.70	[-3.74,-1.67]
Session	Counselor	-0.16	[-0.36,0.02]	-0.17	[-0.37,0.00]	-0.37	[-0.63,-0.13]	-0.32	[-0.58,-0.04]
	Client					0.14	[-0.17,0.45]	-0.02	[-0.38,0.37]
Last	Counselor	0.83	[0.01,1.61]	0.94	[0.09,1.77]	1.20	[-0.01,2.37]	1.44	[0.13,2.58]
	Client					0.28	[-1.21,1.69]	0.33	[-1.18,1.78]
Alliance	Counselor			-0.32	[-0.92,0.35]			-0.54	[-1.35,0.26]
	Client							0.60	[-0.42,1.55]
All×Sess	Counselor			0.16	[-0.03,0.33]			0.19	[-0.07,0.42]
	Client							-0.05	[-0.22,0.10]
σ_{dyad}	Counselor	0.64	[0.27,1.09]	0.68	[0.30,1.13]	0.64	[0.30,1.08]	0.66	[0.29,1.12]
	Client	0.79	[0.34,1.39]	0.79	[0.31,1.42]	0.82	[0.33,1.43]	0.86	[0.35,1.47]
ρ_{dyad}		0.01	[-0.71,0.80]	-0.06	[-0.78,0.72]	-0.03	[-0.77,0.71]	-0.03	[-0.77,0.76]
σ_{conv}	Counselor	0.95	[0.71,1.22]	0.92	[0.69,1.18]	0.92	[0.69,1.19]	0.93	[0.67,1.22]
	Client	1.17	[0.87,1.52]	1.17	[0.85,1.50]	1.11	[0.81,1.42]	1.11	[0.82,1.45]
ρ_{conv}		-0.26	[-0.60,0.08]	-0.22	[-0.56,0.16]	-0.19	[-0.55,0.16]	-0.26	[-0.63,0.16]

Table 6.8: Mixed-effect regression models predicting the proportion of time (logit-transformed) spent performing hand gestures (other than self-adaptors) during speech. Posterior means and 95% credible intervals are shown. Bold values indicate a credible interval which excludes zero ($p <= 0.05$)

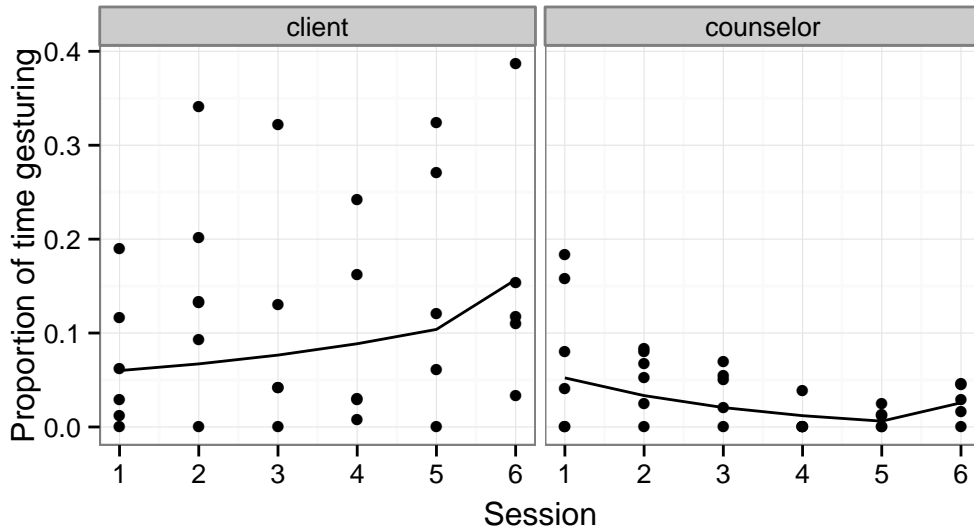


Figure 6.9: Occurrence of hand gestures, by session and role. The lines are the model-based predictions for the estimated average dyad.

systematic trends were observed consistently across all models, although in Models C and D (which include separate trends for the counselor and the client), there was a significant decrease in the use of eyebrow movement by the counselor in later stages.

6.3 Discussion

I show significant longitudinal differences in most behaviors investigated. In later conversations, gaze-aways during speech are more common, and less time is spent smiling and frowning. Participants nod more when they report lower therapeutic alliance. The counselor decreases speaking time while the client increases. Most trends reverse in the last conversation.

The results presented here do not fully agree with previous work. Prior (cross-sectional) studies have reported that friends share speaking time more equally than strangers, whereas here the counselor speaks less initially and further decreases her speaking time in later conversations. This may be due to the nature of the conversational task, which is focused on the client's attitudes

Model A		Model B		Model C		Model D	
DIC		95.70	95.37	94.44	95.86		
Intercept	Counselor	-3.95 [-4.70,-3.20]	-4.23 [-5.06,-3.42]	-3.48 [-4.38,-2.66]	-3.66 [-4.64,-2.73]		
	Client	-2.26 [-3.05,-1.47]	-2.28 [-3.06,-1.50]	-2.86 [-3.80,-1.88]	-2.89 [-3.87,-1.91]		
Session	Counselor	-0.13 [-0.33,0.04]	-0.09 [-0.27,0.10]	-0.34 [-0.64,-0.06]	-0.34 [-0.66,-0.06]		
	Client			0.14 [-0.21,0.44]	0.33 [-0.11,0.71]		
Last	Counselor	0.51 [-0.34,1.30]	0.44 [-0.31,1.28]	0.65 [-0.65,1.98]	0.52 [-0.90,1.90]		
	Client			0.28 [-1.39,1.69]	0.24 [-1.26,1.83]		
Alliance	Counselor		-0.59 [-1.25,0.08]		-0.27 [-1.08,0.53]		
	Client				-0.82 [-1.85,0.17]		
All×Sess	Counselor		0.11 [-0.08,0.31]		-0.03 [-0.27,0.22]		
	Client				0.05 [-0.12,0.21]		
σ_{dyad}	Counselor	0.60 [0.28,1.01]	0.62 [0.28,1.09]	0.62 [0.27,1.08]	0.66 [0.29,1.17]		
	Client	0.65 [0.28,1.16]	0.65 [0.28,1.12]	0.70 [0.30,1.29]	0.75 [0.31,1.35]		
ρ_{dyad}		-0.02 [-0.75,0.72]	-0.03 [-0.77,0.72]	-0.10 [-0.84,0.67]	-0.04 [-0.80,0.73]		
σ_{conv}	Counselor	1.10 [0.84,1.42]	1.11 [0.81,1.41]	1.06 [0.80,1.38]	1.05 [0.80,1.37]		
	Client	1.26 [0.93,1.60]	1.23 [0.90,1.54]	1.21 [0.90,1.56]	1.18 [0.86,1.51]		
ρ_{conv}		-0.39 [-0.68,-0.08]	-0.42 [-0.69,-0.11]	-0.33 [-0.66,-0.01]	-0.38 [-0.68,-0.02]		

Table 6.9: Mixed-effect regression models predicting the proportion of time (logit-transformed) spent with raised or lowered eyebrows during speech. Posterior means and 95% credible intervals are shown. Bold values indicate a credible interval which excludes zero ($p <= 0.05$)

and behavior.

The finding of less smiling and frowning over time is broadly in agreement with Tickle-Degnen’s model of rapport [162]: positivity is more important in early conversations. In an informal examination of the corpus (during coding) we noted that few of the smiles coded are Duchenne smiles, which involve the use of facial muscles that raise the cheeks (in addition to those which raise the corners of the mouth), and are thought to indicate felt emotion [56]. I conjecture that “performing” appropriately-valenced facial expressions as an indication of empathy may be important for establishing rapport in early conversations.

Cassell et al. report that friends use fewer nods for acknowledgments than strangers [40]. I report an effect of relationship strength: stronger self-reported relationship is associated with less nodding, primarily in early conversations. I give two conjectures — which are not mutually exclusive — that attempt to explain this difference.

First, Cassell et al.’s study, which is cross-sectional, cannot easily distinguish differences between dyads from changes that occur over time: their finding of less nodding among friends could indicate either that less nodding indicates that a dyad is more likely to become friends, or that dyads initially nod more and this behavior tends to decrease with continued interaction. My results are more consistent with the first explanation: nodding is associated with stronger self-reported working relationships, rather than a longer interaction history.

Second, Cassell et al.’s study presents a novel conversational task (a direction-giving task) to all dyads. Task familiarity is uniformly low. In this study, task familiarity and interaction history are confounded: in later sessions, dyads always have both increased task familiarity and increased interaction history. I conjecture that in an interaction with low task familiarity — which includes both Cassell et al.’s study and the early sessions of the corpus — less nodding and acknowledgment behavior is associated with a stronger working relation-

ship. This is consistent both with these results, which rely on self-reported working relationship, and with Cassell et al.'s results, under the assumption that friends are likely to have a stronger working relationship than strangers.

The results I report on the use of hand gesture, including both self-adaptors and other gestures, are difficult to generalize. We see significant change for the counselor only, and these results may blend general and idiosyncratic factors: they may be useful for developing models of this particular counselor rather than more general models of human behavior.

Across nearly all behaviors, I report a pattern where the observed change over time reverses in the last conversation. I note that participants were always aware that the sixth conversation was their last, and all dyads had an explicit discussion about the end of their relationship. I conjecture that a final interaction, like an initial interaction, has increased uncertainty about the participants' interpersonal relationship, and this uncertainty is associated with changes in behavior. The final interaction is an abrupt and major change in the nature of the participants' interpersonal relationship. In very long-term interaction, other major changes may occur: for example, a shift from a professional and impersonal relationship to a friendship. This suggests a future research question: Do other changes in the nature of an interpersonal relationship produce similar effects (i.e., verbal and nonverbal behavior similar to a first interaction)?

CHAPTER 7

Rhythm: Behavior Generation for Long-Term Interaction

Thus far, this dissertation has presented a survey of results that show that verbal and nonverbal behavior in face-to-face conversation may vary in long-term repeated interaction. These results have examined human-human interaction, and have not involved conversational agents or other computer interfaces.

In this chapter, I examine the problem of constructing an embodied conversational agent (ECA) which incorporates these results, and displays conversational behavior that varies in the context of a changing user-agent interpersonal relationship, and varies over time both within and across conversations. I present “Rhythm,” an implementation of verbal and nonverbal behavior generation for ECAs, based partially on the findings in this dissertation.

7.1 Introduction

An ECA is a complex interface capable of multimodal output (and often multimodal input, although I do not focus on input here) including both the audio of its utterances, and synchronized accompanying nonverbal behavior, either in visual form (with an animated character) or through physical movement (with a robotic character). Producing behavior involves multiple related tasks, including generating the semantic content of utterances, choosing the

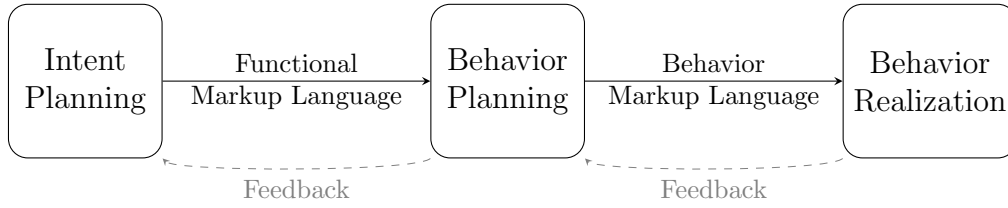


Figure 7.1: The SAIBA framework [89, 167] for multimodal behavior generation.

surface (lexical) form of utterances, choosing the timing and form of nonverbal behaviors to accompany an utterance, producing audio with appropriate prosodic and intonational features, rendering the nonverbal behaviors in an animated representation (or other embodiment), and synchronizing the verbal and nonverbal communication channels.

To address the breadth of issues required, most ECA implementations take a modular approach to behavior generation. The SAIBA framework (Situation, Agent, Intention, Behavior, Animation) [89, 167] currently used in several implementations, attempts to codify best practices learned from previous systems, while specifying standards intended to allow modules developed independently to interoperate with each other.

SAIBA is based on a pipelined generation model (Figure 7.1), consisting of three modules, although each may contain submodules; the framework treats each as a black box, and avoids specifying the internal structure. An intent planning module is responsible for generating the communicative and expressive intent of an utterance, in a form that avoids reference to any physical behavior. A behavior planning module takes this as input, and generates a high-level description of the verbal and nonverbal behaviors to be performed; this is high-level in that it specifies behaviors to be performed (e.g., a head nod) while avoiding details tied to a particular embodiment (e.g., specific joint angles). Finally, an embodiment-specific behavior realization module is responsible for generating the agent behavior. SAIBA specifies two standardized markup languages to be used to communicate between these modules:

Functional Markup Language (FML) to describe communicative content, and Behavior Markup Language (BML) to describe a communicative behavior.

Within this framework, Rhythm is a behavior planning module. It consumes a description of the communicative intent of an utterance, which is operationalized as the text of the utterance with some (optional) additional annotations that specify interpersonal and emotional contexts. From this, it produces a behavior description that can be consumed by a behavior realization module.

7.2 Background and Related Work

Multiple approaches to nonverbal behavior generation have been explored. Rule-based systems, such as BEAT [43] and NVBG [100], take as input the text of an utterance (possibly annotated with additional contextual information), and generate annotations for cooccurring nonverbal behavior using a manually authored set of rules. In both of these systems, behavior rules are applied following a shallow linguistic analysis that, either heuristically or stochastically, attempts to identify syntactic (e.g., parts of speech and phrase structure) and pragmatic (e.g., given and new information) that will be used as input features by the rules. The rules are typically based on prior observational work (either qualitative or quantitative) of human interaction. While the set of rules must be pre-specified, these systems can allow for some variability in behavior by including stochastic rules, which generate a probability distribution over nonverbal behavior annotations: for example, generating the probability of occurrence of a behavior at a particular point in an utterance.

An alternative approach, sometimes labelled as a “data-driven” approach, avoids the use of hand-written rules by directly applying a statistical or machine learning model which was trained on annotations of observations of human behavior. For example, Neff et al. [120] trained speaker-specific models of gesture generation by fitting a set of Hidden Markov Models to annotated

video samples. The resulting models produced nonverbal behavior annotations (primarily gestures) given an annotated text as input, and an empirical study showed that observers could distinguish a particular speaker’s “style” in behavior generated using a model trained on that speaker.

Either rule-based or data-driven approaches can be driven by inputs other than the text of an utterance. For example, a series of “Rapport Agents” were developed that generate backchannel nodding behavior for a listener agent, using features of the speaker’s communication as input and applying either hand-authored rules [67] or a data-driven model [81] created by training a Conditional Random Field [92] model on observational data.

All of these approaches generate nonverbal behavior either as an addition to existing content of an utterance or, in the case of the Rapport Agent, independently without any other accompanying behavior. It is also possible to generate the content of an utterance and its cooccurring nonverbal behavior jointly, from the same semantic content. Kopp et al. [90] give an example of grammar-based natural language generation extended to generating hand gestures.

7.3 Implementing Longitudinal Behavior Changes

Rhythm is a rule-based nonverbal behavior generation system, most similar in design to BEAT [43]. However, the approach I outline for generating long-term dynamical behavior is intended to be more generally applicable, and able to combine with other underlying approaches to nonverbal behavior generation. I require that the underlying behavior system be able to output generation probabilities for each behavior event (rather than simply the occurrence of an event), and model long-term changes in behavior through the use of rules that make adjustments to these probabilities.

The results in previous chapters are not sufficient by themselves for a full

computational model of nonverbal behavior generation in the context of a complex multi-conversation interaction. I consider only a subset of verbal and nonverbal behavior, and do not model all features of these behaviors: for example, when examining head nods (Chapter 6), I constructed a model that predicts only the occurrence of nods, and not duration, velocity, or other features. I also considered a subset of possible predictors of behavior, and have not considered some predictors known to be significant from prior work: for example, I have not explicitly included turn-taking structure when modeling gaze [42]. Finally, the corpus is limited to six short conversations per dyad, and several of the results are limited to an examination of the first minute of conversation.

In order to implement these findings, it is necessary to make some additional assumptions. The assumptions are chosen pragmatically, and intended to produce the simplest computational model of conversational behavior that both accounts for the results in this dissertation and can be used to produce an implementation of nonverbal behavior generation. I do not argue that they necessarily produce realistic behavior; that claim requires further validation.

First, I assume that all results which are found by examining the first minute of conversation hold constant throughout a conversation. Based on this assumption, the models developed Chapter 6 are used to construct rules that apply to any point in a conversation.

Second, I assume that these results combine additively with other predictors of behavior; that is, that no interaction effects exist between the predictors I have modeled and those I have not modeled. This assumption allows combining the findings in this dissertation with those given by prior work, treating these findings as adjustments to baseline probabilities, rates, or durations determined from prior work, and ignoring how those baselines are generated.

My results are based on generalized linear mixed effect models, with response variables drawn from a transformed normal distribution (for articulation rates, mouth movements, and head nods), a binomial distribution (for

posture shifts), or a Poisson distribution (for gaze). If one ignores random effects, which model participant-specific variation, and focus on modeling the estimated population-average, then a response variable y (a normalized duration for articulation rate, and a generation probability for others) can be modeled as:

$$g(y) = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$$

where g is a monotonic, invertible and differentiable function (log or logit in the models), $x_1 \dots x_n$ are predictors (e.g., the number of previous sessions), and $\beta_0 \dots \beta_n$ are model coefficients, with β_0 (the intercept) representing the population-average baseline.

Given a duration or probability y produced by some underlying nonverbal behavior generation system, and under the assumption that these results combine additively with those that produced y , we simply substitute $g(y)$ as the baseline and can give an “adjusted” duration or probability y' as:

$$y' = g^{-1}(g(y) + \beta_1 x_1 + \dots + \beta_n x_n)$$

7.4 Architecture

Architecturally, Rhythm is similar to BEAT [43], with a pipeline architecture consisting of four modules (Figure 7.2): (1) an analysis module performs a shallow linguistic analysis to identify relevant syntactic and pragmatic features; (2) a behavior generation module generates baseline behavior annotations; (3) a behavior adjustment module modifies these annotations to account for the long-term changes documented here; and (4) a filtering module modifies annotations to resolve conflicts and changes stochastic annotations into deterministic ones.

An input utterance is tokenized as a preliminary step. Utterances can be supplied with optional annotations at input, which provide additional in-

formation about the discourse and interpersonal context above the level of a single utterance. Input annotations currently used by Rhythm include:

- Turn-taking context: whether an utterance starts or ends the speaker's turn.
- Within-conversation discourse context: the amount of dialogue since the conversation began, in time and in number of dialogue turns.
- Affective context: the emotional state the speaker intends to communicate.
- Interpersonal context: the speaker's beliefs about the speaker-listener interpersonal relationship, expressed as an estimated working alliance score.

Each module is responsible for adding or modifying a set of annotations on the sequence of tokens in an utterance. Annotations include contextual information (as above), linguistic information (e.g., part of speech tags), and suggested nonverbal behaviors. Annotations can either apply to the entire utterance (as in the interpersonal context annotations above), or to a contiguous sub-sequence of tokens (as in an annotation identifying syntactic information).

Relative to the architecture used in BEAT, the adjustment module is novel, while both the generation and filtering modules are modified to work with it.

7.4.1 Analysis

OpenNLP¹ is used to perform part-of-speech tagging, and to chunk sentences into phrases, both using stochastic natural language processing techniques. Wordnet [111], via the MIT Java Wordnet Interface², is used to determine the lemma for each word, and to find antonyms.

¹OpenNLP 1.5.2, <http://opennlp.apache.org/>

²JWI 2.2.2, <http://projects.csail.mit.edu/jwi/>

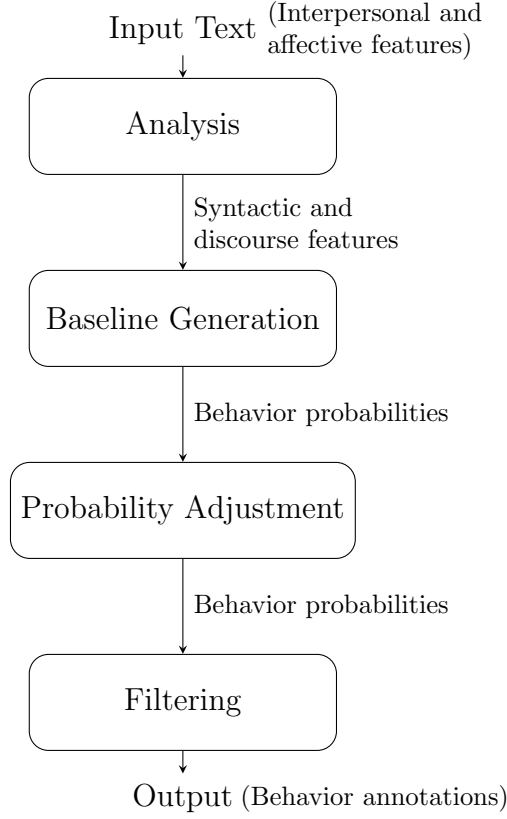


Figure 7.2: The Rhythm behavior generation pipeline.

For each clause of an utterance, the theme and rheme are identified using heuristic rules given by Hiyakumoto et al. [76]. These are units of the information structure of a clause, with the theme defined as the portion of the clause that provides continuity with previous clauses or utterances, and the rheme as the portion that contributes new information [72]. Determining the information structure of an utterance is useful, as this structure is associated with differences in verbal and nonverbal behavior: for example, theme and rheme are often marked by intonational patterns [131, 158], and gestures tend to occur in the rheme [38].

7.4.2 Behavior Generation

Behavior generation consists of a set of hand-written rules that operate on the annotated sequence of tokens. Each rule, if triggered, adds a behavior annotation. While the set of features varies depending on the particular behavior, all include timing information (word-aligned start and end positions), a priority, and a generation probability (which may be 1, indicating certain generation). Stochastic rules, where a behavior is generated with some probability, are implemented with a rule that *always* adds a behavior annotation, but attaches an appropriate generation probability.

Note that all behavior annotations are generated independently, without examining annotations added by other behavior generation rules. This may lead to the generation of conflicting annotation (i.e., involving the same bodily degrees of freedom), which will be resolved during a later filtering step.

Rhythm currently implements the following behavior generation rules, which are a subset of those implemented by BEAT and NVBG. A novel feature is the handling of stochastic rules.

- Intonation accents are generated on new objects within the theme (an L+H* accent) and the rheme (H*), following [131].
- Intonation boundary tones are generated at the end of a theme (L-H%) or a rheme (L-L%), also following [131].
- Beat gestures are generated on all new objects within a rheme, as they are the most common hand gesture [108], and gestures tend to occur on new information and in the rheme [38].
- Eyebrow raises are generated on new objects within a rheme, as this can be a signal of new information [124].
- Head nods are added with high probability on words that mark acknowledgment and grounding (e.g., “ok”) and which come immediately after a conversation partner’s utterance.

- Posture shifts are added at clause boundaries, with high probability in the presence of a topic shift and with low probability in the absence of a topic shift, following [41].
- Gaze-aways are generated as a function of information structure and turn-taking [163]: gaze-aways are generated on the theme of a clause with certainty if occurring at the beginning of a dialogue turn, and with lower probability if elsewhere.
- Affective facial expressions (e.g., smiles and frowns) are generated based on annotations identifying the affective content of an utterance, with generation probability based on the intensity of the affect.

7.4.3 Behavior Adjustment

Behavior adjustment consists of a set of hand-written rules, based on findings in the preceding chapters. Each rule operates on the generation probabilities of behavior annotations, and modifies them as a function of annotations on the utterance that specify its discourse context (i.e., position within the larger conversation) and interpersonal context (i.e., interaction history and interpersonal relationship of the speaker and listener).

The probability of generating a smile or frown changes as a function of the interaction history, where s is the number of prior conversations, and f is 1 if the agent believes this is a final conversation and 0 otherwise:

$$p' = \text{logit}^{-1}(\text{logit}(p) - 0.16s + 0.97f)$$

The probability of generating a gaze-away changes as a function of interaction history:

$$p' = 1 - (1 - p)^{\exp(0.2s - 0.8f)}$$

The probability of generating a headnod changes as a function of interaction history and the agent's beliefs about the strength of the user-agent

relationship, where a is therapeutic alliance, standardized to a z -score:

$$p' = \text{logit}^{-1}(\text{logit}(p) + 0.06s - 0.44f - 0.28a + 0.06sa)$$

The probability of generating a posture shift changes as a function of interaction history and minutes from conversation start (m):³

$$p' = \text{logit}^{-1}(\text{logit}(p) + 0.16s - 0.03m - 0.02sm)$$

The duration (d) of a word annotated to be a discourse marker or an acknowledgment changes as a function of interaction history and of the number of dialogue turns since conversation start (t):

$$d' = \exp(\log(d) - 0.015s - 0.045 \cdot (t/109.4))$$

7.4.4 Behavior Filtering

The behavior filtering module has two steps: conflict resolution and realizing stochastic behavior annotations. Conflict resolution consists of identifying behavior annotations which use the same degrees of freedom in the embodiment (e.g., two gestures using the same hand, or a simultaneous smile and frown). A conflict is resolved by removing the lower-priority behavior, or (in case of a tie) by removing a behavior by random choice.

Once conflicts are removed, any stochastic behavior annotations — those with a generation probability $p < 1$ — are randomly removed with probability $1 - p$.

7.5 Examples

Figure 7.3 shows an example of behavior generated by Rhythm for the utterance “Okay, great. And how is your exercise going?,” when accompanied by

³ This rule was based on previously published work [148] which used a variant of the model presented in Chapter 5, in which parameters were estimated by maximum likelihood rather than Bayesian methods. Due to this, the coefficients in this rule are slightly different from those given in Table 5.1.

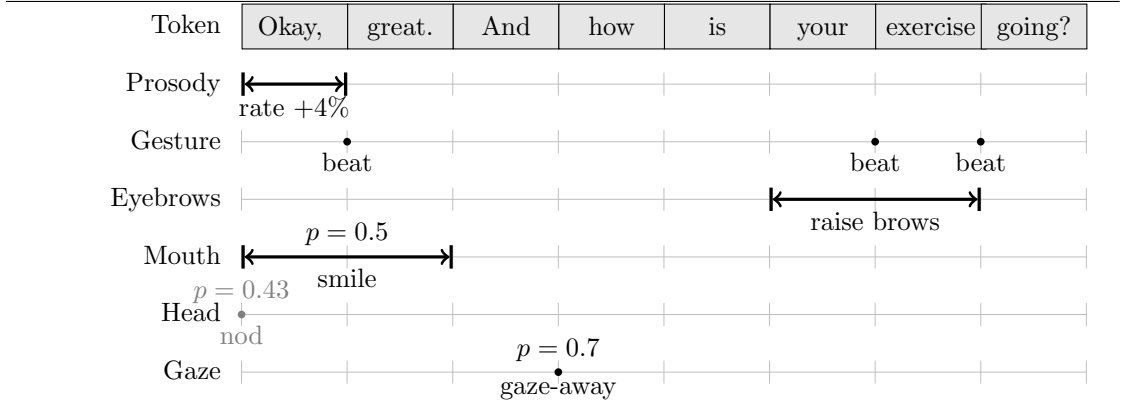


Figure 7.3: Sample behavior generation for a first conversation with high therapeutic alliance

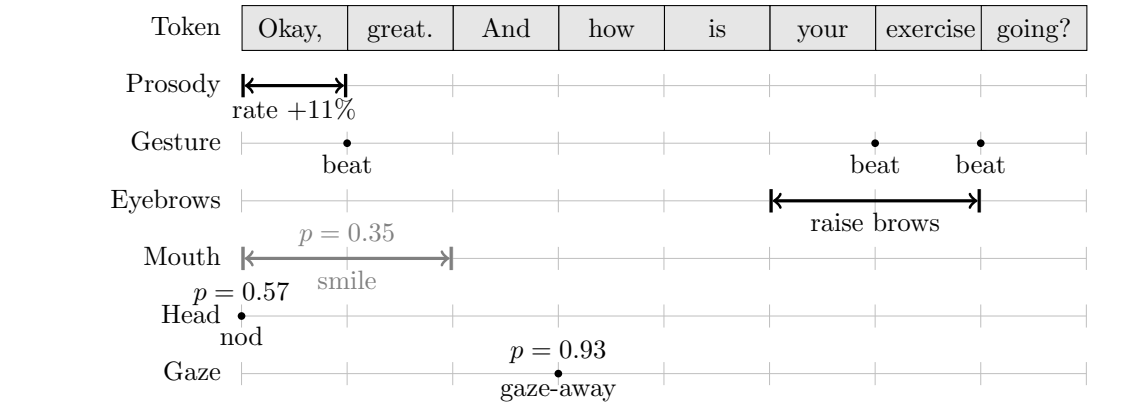


Figure 7.4: Sample behavior generation for a fifth conversation with low therapeutic alliance

annotations indicating that the speaker and listener have no previous interactions, and that the speaker believes there is high therapeutic alliance (+1 standard deviation). Figure 7.4 shows behavior generated for the same utterance, but accompanied instead by annotations indicating that the speaker and listener have had 4 previous conversations and that the speaker believes there is low therapeutic alliance (-1 standard deviation).

Aside from annotations related to interaction history and interpersonal relationship, both samples were generated using the same contextual annotations: they were annotated as both starting and ending the speaker’s dialogue turn, as occurring at 5 minutes (and 100 dialogue turns) from the start of

conversation, and as having positive affective content (intensity 0.5).

Articulation Rate The phrase “Okay” at the beginning of the utterance is identified as an acknowledgment, and the duration (d) is modified as a function of the number of previous sessions (s) and the number of previous dialogue turns within the conversation (t). For the first conversation, they take the values $s = 0$ and $t = 100$, giving $d' = 0.96d$ (approximately a 4% increase in articulation rate), while in the fifth conversation, they take the values $s = 4$ and $t = 100$, giving $d' = 0.90d$ (approximately a 11% increase in articulation rate).

Facial Expression The sentence “Okay, great” is annotated as having positive affective content. Rhythm’s default rules give a baseline probability $p = 0.5$ of generating a smile. This is modified as a function of the number of previous sessions (s) and whether it is a final session ($f \in 0, 1$). In the first conversation ($s = 0, f = 0$), the generation probability is unmodified, while in the fifth ($s = 4, f = 0$) it is decreased to $p' = \text{logit}^{-1}(\text{logit}(p) - 0.64) = 0.35$.

Head Movement Rhythm’s default rules give a baseline probability $p = 0.5$ of a head nod before the start of the utterance. This is modified as a function of the number of previous sessions (s), whether it is a final session ($f \in 0, 1$), and the speaker’s working alliance (a). For the first session, with high alliance ($s = 0, f = 0, a = 1$), the generation probability is modified to $p' = \text{logit}^{-1}(\text{logit}(p) - 0.28) = 0.43$. In the fifth session, with low alliance ($s = 4, f = 0, a = -1$), the generation probability is modified to $p' = \text{logit}^{-1}(\text{logit}(p) + 0.24 + 0.28 - 0.24) = 0.57$.

Gaze The tokens “how is...” were heuristically identified as beginning the theme of a clause in the utterance, and Rhythm’s default rules give a baseline probability $p = 0.7$ of a gaze-away. This is modified as a function of the number of previous sessions (s) and whether it is a final session ($f \in 0, 1$). For the first session ($s = 0, f = 0$), this probability is unmodified, while for

the fifth session ($s = 4, f = 0$), it is modified to $p' = 1 - (1 - p)^{\exp(0.8)} = 0.93$.

CHAPTER 8

Evaluation

I now present the final contribution of this dissertation: a longitudinal randomized controlled trial of long-term conversational behavior adjustments in an Embodied Conversational Agent which acts as a physical activity counselor, and a test of the effects of those adjustments on users' engagement, on the users' attitudinal change toward physical activity, and on the perceived behavioral realism of the agent.

This trial is intended first as a test of the thesis statement of this dissertation: that including long-term dynamic changes in verbal and nonverbal behavior of an Embodied Conversational Agent — where those changes are modeled on changes observed in human-human interaction — will improve user engagement with the agent, and cause the agent to be perceived as more realistic.

A second motivation is to provide some additional empirical support for the observational results obtained from examining the Exercise Counseling Corpus in earlier chapters (Chapter 4, Chapter 5, and Chapter 6). The corpus includes a small ($N = 6$) number of clients and only a single counselor. There is a concern that the results in Chapters 4, 5, and 6 may be idiosyncratic patterns of behavior of one particular individual (the counselor) or a small group of individuals (the clients) that may not generalize to the larger population. While I have tried to mitigate this concern by interpreting only patterns of behavior which are observed in both the counselor and the clients,

this evaluation study will also provide partial validation that these patterns of behavior are also perceived as realistic by a larger population — albeit with a virtual rather than a human counselor.

Aside from concerns of the validity of the observational results, I also wish to test whether these patterns of behavior are perceived as realistic and have the desired effect when implemented in an Embodied Conversational Agent, rather than performed by a human speaker. An ECA has neither the visual nor behavioral realism of a human speaker; this is a limitation of both current state-of-the-art agents and of the particular ECA used in this study (described in more detail below) and the behavior generation used with it (Section 7.3). Communication between a user and the agent is also limited; in this particular case, the user is limited to multiple-choice input (avoiding problems of natural language understanding) with strict turn-taking determined by the system. An evaluation — with agent-human rather than human-human interaction — is therefore required before we can conclude that these patterns of behavior are realistic in a particular ECA.

The evaluation study presented here is a three-arm between-subjects trial. In all three conditions (described in greater detail below), participants interacted for six weeks (up to once per week) with an ECA that acted as a virtual health behavior change counselor that promoted regular exercise in the form of planned bouts of brisk walking. The appearance of the counselor, and the content of the user-agent conversations were both constant across conversations, while the agent’s verbal and nonverbal behavior varied: in one condition (**Dynamic**), the agent’s behavior followed the models developed in previous chapters, while in others the agent’s behavior either did not change across conversations (**Static**) or changed at a greater rate than predicted (**Exaggerated**).

8.1 An ECA for Long-Term Interaction

As a proof-of-concept of an embodied conversational agent supporting realistic conversational behavior in long-term interaction, I embedded the Rhythm conversational behavior annotator in the Litebody [14] web-based ECA framework. The combined system produces a functional ECA, which is used in this evaluation study.

Litebody is an open-source¹ framework intended for the creation and deployment of very lightweight ECAs. Litebody-based ECAs can be shown in most standard web browsers without the requirement of installing additional software or plugins.² The ECAs are shown from the shoulders up only, and conduct system-initiated conversation with restricted, multiple-choice user input. It can deliver synthesized speech with prosodic variation and synchronized nonverbal behavior including eye gaze, facial expressions (smiles and frowns), eyebrow movement, head movement, and postural shifts.

Litebody uses a client-server architecture with an HTTP-based protocol: the server is responsible for generating an audio file for each agent utterance, along with an animation script giving the timing of any associated nonverbal behaviors. The Litebody framework itself does not determine the content of utterances or perform any other dialogue management functions; it is middleware, responsible for managing the details of the client-server protocol. A modular design allows for various dialogue managers to be used; a dialogue manager is responsible for delivering the text of agent utterances, annotated with nonverbal behavior and prosodic information.

In this study, Litebody is used with a dialogue manager that executes scripts written in a custom hierarchical transition network-based scripting language. States in this language contain agent utterances, while state transitions are user input choices. The scripting language specifies the lexical form of ut-

¹<http://litebodysuite.sourceforge.net>

²Litebody requires Adobe Flash, which is currently installed on a large majority of desktop-based web browsers.

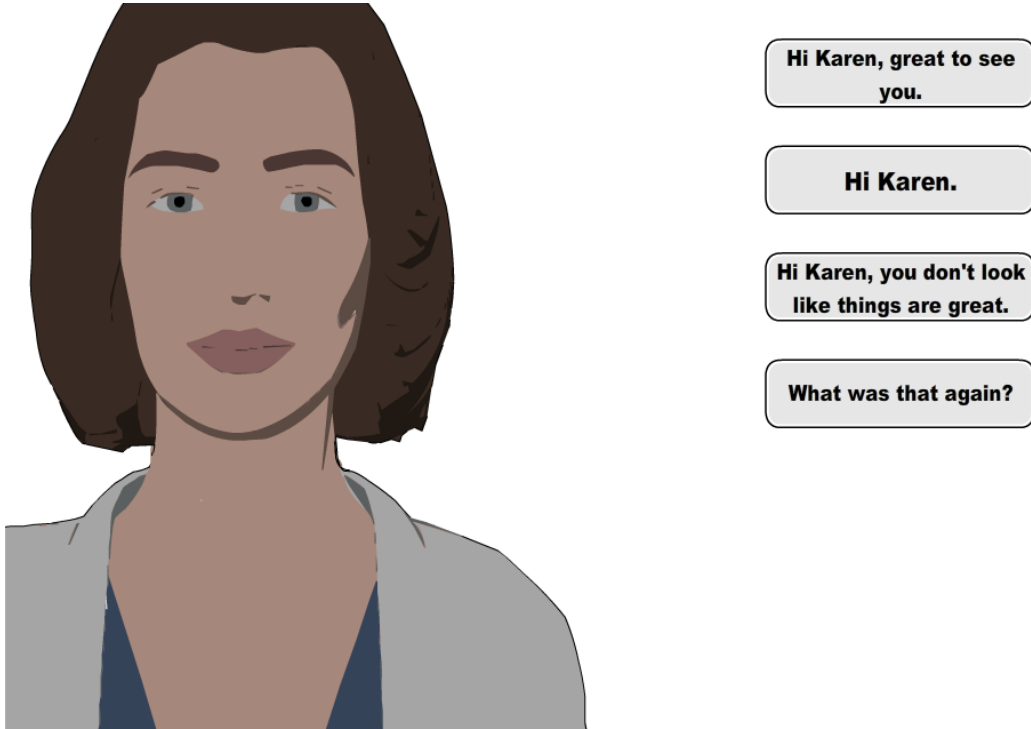


Figure 8.1: The Litebody web-based ECA.

terances and some additional annotations, while Rhythm is used at the time of utterance generation to add nonverbal and prosodic annotations. The dialogue manager is also responsible for tracking within-conversation contextual information used by Rhythm — primarily the temporal offset of an utterance within the conversation. Affective context is specified by manual annotation within the dialogue scripts, indicating whether utterances are positively-valenced, negatively-valenced, or neutral.

8.2 A Preliminary Study of the Effect of Behavior Changes

Many of the changes predicated by the model are quite subtle: for example, after five conversations, the average speaker increased their articulation rate approximately 8% (and only on specific words). It is possible that such changes are not perceptible or have no measurable effect on users. It was desirable to

have a preliminary test of the perceptibility and effect of some of the changes before conducting a more extensive evaluation study.

I conducted a preliminary evaluation in order to test whether the model-predicted differences in articulation rates, when incorporated in a conversational agent’s speech, were perceptible to users, and whether they had any measurable effect on attitudes toward the agent. In contrast to the full evaluation study described below, the preliminary evaluation was performed in a laboratory setting, and participants took part in a single experimental session. Participants had two similar conversations, with two similar agents, which differed in articulation rates: In the SLOW condition, the articulation rate of the agent’s speech was left unchanged, while in FAST, the articulation rate of acknowledgments and discourse markers was increased by the amount the model predicted would occur after five conversations (8%), and also increased at the predicated rate within a conversation (to a total of approximately 13%).

8.2.1 Apparatus and Measures

The two agents were chosen to have a similar appearance, and both used synthesized speech with synchronized nonverbal behavior. Participants used multiple-choice spoken input, with up to 6 utterance choices displayed by the agent at each turn. However, the agents were controlled via a Wizard-of-Oz setup [48], in order to eliminate any possible effects of speech recognition errors.

Both dialogues consisted of social dialogue only. The dialogues were designed to be approximately the same length (about 40 turns, varying slightly based on participant choices), and contained similar (but not identical) topics. Topics with a low intimacy level were used, such as weather, local sports, and features of the experiment location. Dialogues were manually tagged to identify acknowledgments and discourse markers that should increase in articulation rate when in the FAST condition.³

³Although presented here for clarity, the preliminary evaluation was performed prior to

Perceived rapport was assessed with the bond subscale of the Working Alliance Inventory [79] following each conversation. Participant introversion/extroversion was assessed using a 16-item subset of the Interpersonal Adjective Scales [171].

8.2.2 Procedure

The order of conditions, agents, and dialogues were randomly assigned. Following a demographics questionnaire, participants received brief instructions on how to interact with the agent. Participants were told they would be interacting with two different agents, but were not informed of differences in articulation rates, or any other specific differences. The experimenter left the room during the conversations, and returned to administer questionnaires afterward.

8.2.3 Participants

8 participants (5 female, mean age 34.6, age range 23–63) were recruited via a contact list of potential participants who had expressed interest in previous studies conducted by colleagues but had not participated. All reported high levels of computer proficiency, and all but one were college graduates.

8.2.4 Results

No significant difference was observed in perceived rapport (Working Alliance Inventory) between the SLOW and FAST speech (paired $t(7) = -0.296$, $p = 0.78$). However, given results by Nass and Lee [116], I also analyzed the effect of the participant's extroversion. A linear regression showed that extroversion predicted the difference in perceived rapport between SLOW and

the development of the Rhythm conversational behavior generator. Therefore, automatic annotation was not available and manual adjustments to articulation rates were made instead.

FAST ($R^2=0.55$, $F(1, 6)=7.39$, $p=0.035$). Participants who were more extroverted were more likely to report a higher perceived rapport in the FAST condition.

Only one participant reported noticing a difference in speaking rate. When participants were asked to “guess” which agent spoke faster, 5 of 8 identified the correct agent; this is not significantly different from chance ($\chi^2(1)=0.5$, $p=0.48$). Therefore, I cannot conclude that participants consciously distinguish this difference in speaking rates in conversation, although the effect on rapport (moderated by participant extroversion) is weak evidence for an effect of the difference in behaviors even in the absence of perceptibility.

8.2.5 Discussion

I show some preliminary evidence that changes in articulation rates of a speaker may have a measurable effect, even though listeners may not necessarily be able to consciously perceive the changes. However, the characteristics of the listener may be equally important: Extroverted listeners may prefer a speaker that “jumps right in” with a speaking style that indicates greater familiarity.

This preliminary study is limited by a small number of participants. These results require additional study to determine whether they generalize across people, languages or dialects, or cultural backgrounds.

8.3 A Long-Term Interaction Scenario

The full evaluation is designed around a long-term interaction scenario broadly similar to the scenario used to collect the Exercise Counseling Corpus: a six-week interaction during which an ECA, acting as a virtual behavior change counselor, attempts to persuade the user to change his or her attitudes toward regular physical activity and to perform more regular physical activity.

8.3.1 Behavior Change Intervention

The focus of this evaluation study is on improving engagement and voluntary usage of an ECA-based behavior change intervention, and the study is not primarily intended to explore variations in the content of a behavior change intervention. The content of the interactions is therefore based on previous health behavior change interventions which used ECAs and targeted increases in regular physical activity, or related behaviors. Most heavily (although not exclusively) I have adapted content from the “Health Behavior Change Ontology” project⁴ [149, 17]. Modifications have been made as needed to combine content from separate interventions, and to adapt content originally developed in the context of a different frequency of use (up-to-daily interaction rather than weekly).

The Transtheoretical Model of Behavior Change [132] theorizes that persons attempting to change a health behavior, such as increasing their physical activity, pass through several Stages of Change, and predicts that different change processes characterize these stages. Following this model, the dialogue content delivered to a participant is tailored based on an initial stage of change assessment (made in the first conversation). Participants in later stages of change (preparation, action, or maintenance) receive dialogue (Figure 8.3) consisting of goal-setting, self-monitoring, and problem solving: each week the participant is asked to set a weekly walking goal, which is reviewed in the next conversation, where the agent can show a visualization of the participant’s progress and provide suggestions for common problems and barriers to behavior change.

Participants in earlier stages of change (precontemplation and contemplation), distinguished from later stages by a lack of intention to change their behavior in the immediate future, do not receive goal-setting dialogue immediately. Instead, they first received dialogue content inspired by techniques

⁴<http://hbco.sourceforge.net>

- (1) **Agent:**
Okay. How motivated are you to do more physical activity? Can you give it a number, on a scale from 0 to 10?
- (2) **User:**
3
- (3) **Agent:**
It sounds like you have some reasons to feel motivated to get more physical activity, since you answered 3 instead of zero. What would it take to move you up to 7? What stops you from feeling more motivated?
- (4) **User:**
I just haven't thought about it.
Physical activity won't help me right now.
Other things are more important.
Physical activity won't help me.
I'll never manage to do it.
Actually, I DO feel very motivated.
I'd rather not talk about this now.
- (5) **Agent:**
In that case, let's talk about some of the benefits of physical activity. One of the great things about physical activity is that it helps you both now and in the future. Which would you like to talk about?
- (6) **User:**
Benefits now.
Benefits later.
Neither.
- (7) **Agent:**
Sure. Besides helping you to get and stay healthy and fit, physical activity also can give you more energy and reduce stress. Many people find it helps to keep up a positive outlook on life.
- (8) **User:**
OK.
- (9) **Agent:**
Thanks for talking about this with me. How are you feeling about physical activity after our discussion?
- (10) **User:**
I'm not sure.
About the same.
I feel more motivated.
- (11) **Agent:**
Okay. We can talk about this again another time.

Figure 8.2: A fragment of a sample user-agent conversation, showing Motivational Interviewing-inspired content. User input was multiple choice, with the chosen item indicated by italics.

- (1) **Agent:**
So, your goal from last time was to walk for 20 minutes, 3 days a week. How did it go?
- (2) **User:**
I didn't walk at all.
I only walked once.
I walked less than that.
I did it!
I walked more!
- (3) **Agent:**
I'm sorry to hear that. But it sounds like even when you had a rough week, you still tried to do some walking. Would you like some tips that may help with weeks like that?
- (4) **User:**
Yes.
No.
- (5) **Agent:**
Here's a plot of the walking you've been doing.
- (6) **User:**
OK.
Thanks.
- (7) **Agent:**
So, it's time to set a goal for next week. Can you try again to do 20 minutes, 3 times a week?
- (8) **User:**
I can do that.
I'm not sure.
I don't think I can.
I might have some problems.
I can walk more!
- (9) **Agent:**
Great! I'm sure you'll do it.
- (10) **User:**
OK.
Thanks.

Figure 8.3: A fragment of a sample user-agent conversation, showing goal-setting content. User input was multiple choice, with the chosen item indicated by italics.

from Motivational Interviewing (MI), a client-centered counseling method for enhancing motivation and confidence in behavior change by prompting clients to explore and resolve ambivalence toward change [112]. MI has been applied to physical activity promotion [74] (among numerous other health behaviors), and has been suggested as an effective counseling method for precontemplators and contemplators [50]. The intervention dialogue included three MI-inspired conversations: a discussion of the pros and cons of regular physical activity (used in the first conversation), a discussion (Figure 8.2) of reasons to be motivated to engage in more physical activity (used in later conversations, possibly repeatedly, until a participant indicated strong motivation), and a discussion of reasons to be confident in successfully engaging in physical activity (used until a participant indicates strong confidence). All were based on eliciting statements from the user through questions with the surface form of an open-ended question (e.g., “When I talk about physical activity, what kinds of things come to mind?”), with participant input restricted to a set of common answers, to which the ECA gave reflective and reframing responses. Participants who indicated strong motivation and confidence in these discussions were assumed to have progressed to a later stage of change, and received the goal-setting content in subsequent conversations.

8.3.2 Interaction Schedule

I designed a long-term interaction intended to consist of six weekly conversations. The scenario does not assume every user will participate in the full set of six conversations (although that is a goal), and the interaction is intended to remain coherent in the presence of missing sessions.

Initial Interaction The first weekly conversation began with an introduction to the agent, and an introduction to the mechanics of conversations, including the use and limitations of the agent’s multiple-choice input. This was followed by a general introduction to the topic of physical activity and

the specific target behavior (brisk walking), and an in-conversation stage of change assessment.

Depending on the assessed stage of change, participants received either dialogue negotiating an initial walking goal (if in preparation, action or maintenance), or the MI-based discussion of pros and cons of regular physical activity (if in precontemplation or contemplation). At the end of the MI dialogue, participants who asserted a strong intention to begin physical activity were given an opportunity to begin goal-setting, followed by the negotiation of a walking goal; other early-stage participants were given an opportunity to immediately proceed to the MI-based discussion of reasons to be motivated for physical activity, or could choose to postpone the discussion to subsequent weeks.

The conversation concluded with a discussion about scheduling subsequent sessions, during which the agent attempted to convince the participant to state a commitment to continued interaction. This was followed by ritual social dialogue and a farewell.

Routine Interactions The second, third, and fourth weekly conversations follow the same pattern. The conversation begins with a greeting and ritual social dialogue. Subsequent dialogue was tailored depending on whether the participant was currently receiving a goal-setting intervention or a Motivational Interviewing-based intervention.

Participants receiving a goal-setting intervention were given a reminder of their previous negotiated walking goal (if any), and asked to self-report their walking behavior in the past week. Depending on the reported behavior and their goal, participants received either positive reinforcement (e.g., “Great job!”) or problem solving. This was followed by a visualization (a bar chart) of walking behavior, and a negotiation of the walking goal for the subsequent week.

Participants receiving a Motivational Interviewing-based intervention re-

ceived a discussion about either reasons for motivation or confidence in regular physical activity, the latter delivered for participants who had already reported high motivation. At the end of each discussion, participants were asked whether they had increased motivation or confidence, respectively; upon receiving a positive answer, the agent would advance to either the confidence discussion (when high motivation was reported) or the goal-setting intervention (when high confidence was reported) in subsequent weeks.

As with the initial conversation, the conversation concluded with an attempt to elicit commitment to continued interaction, followed by ritual social dialogue and farewell.

Pre-Closing Interaction The fifth weekly conversation followed the same pattern as the second through fourth, with the exception of the closing discussion of scheduling and the farewell dialogue. Following patterns observed consistently throughout the Exercise Counseling Corpus, the agent explicitly introduced the concept that the scheduled interactions would be ending (e.g., “Next week is the last session on our schedule, so we’ll try to wrap up everything we have been talking about.”).

Closing Interaction The sixth weekly conversation began with a modified greeting, which stated that it was the last scheduled session (e.g., “So, this is our last weekly session. I’ve really enjoyed talking with you.”). Participants who were receiving a goal-setting intervention reviewed their recent walking behavior and were given a visualization of all their walking behavior across the duration of the study; compared to the discussion in earlier weeks, this primarily omitted the negotiation of a new walking goal. Participants who were receiving a Motivational Interviewing-based intervention instead had a repeated discussion about the pros and cons of regular physical activity, followed by positive reinforcement.

All participants were then offered a final problem-solving discussion (e.g., “Since this may be our last conversation, I thought it would be a good idea

to talk one last time about some tips for walking.”). This was followed by a meta-conversation asking participants to give an overall assessment of the user-agent relationship (e.g., “Now that it’s been several weeks, how do you feel about working with me?”), and a farewell. The agent stated that future conversations were possible, but did not otherwise discuss future interaction or attempt to elicit a commitment.

Encore Interactions Participants who chose to log-in to the system for the two week period after completing the six-week intervention (participating in this period was optional and not compensated in any way) received a modified and abbreviated conversation. Participants who had been receiving a Motivational Interviewing-based intervention received a brief assessment of their current motivation and confidence levels, and were given the option of having a discussion about either topic; these were similar to the discussion in earlier weeks. Participants who had been receiving a goal-setting intervention were given an opportunity to report their recent walking behavior, to see a visualization, and to request problem-solving content.

These conversations concluded with a simple farewell. As with the closing interaction, the agent made no attempt to elicit a commitment to future interaction.

8.4 Methods

The evaluation study was a 6-session longitudinal randomized controlled trial with a 3-group between-subjects design:

1. **Dynamic:** The ECA’s conversational behavior was generated by Rhythm according to the models given in Chapter 7, with the exception of a modified baseline gaze model, as described below.
2. **Static:** The ECA’s conversational behavior was generated by the same models, but without change across conversations: Rhythm always gen-

erated conversational behavior as if it was the first conversation.

3. **Exaggerated:** The ECA’s conversational behavior was generated as in **Dynamic**, but using a modification of the rules given in Section 7.4.3 in which all coefficients were multiplied by 3; this had the effect of giving highly exaggerated changes across conversations.

Preliminary pilot testers noted that the number of gaze-aways appeared unrealistically high in all conditions, and did not indicate any perceptible differences in the rate of gaze-aways across conversations. I modified Rhythm’s baseline rules for gaze-away generation, which were based on rules published by Torres et al. [163], so that gaze-aways were instead generated at approximately the overall rate observed in the corpus: A gaze-away was generated with probability $p = 0.3$ at the beginning of a theme when at the start of a turn, and with $p = 0.2$ at the beginning of a theme otherwise. Rhythm’s adjustment rules for changes in the rate of gaze-away across conversation were unchanged.

In all groups, Rhythm’s input was annotated to indicate that the agent’s working alliance was high (1 standard deviation above the population mean). The affective content of dialogue was manually annotated, with a bias toward high emotional expressivity: positive and negative affective annotation were added wherever plausible.

8.4.1 Measures

The short revised Working Alliance Inventory (WAI-SR) [75] was administered after each conversation. The WAI-SR, as originally designed, refers specifically to psychotherapy; to make it more appropriate for counseling in general, I modified it slightly by replacing the word “therapy” with “counseling”. I also replaced generic references to “the therapist” with “Karen”, the name of the ECA.⁵

⁵With the exception of the reference to the ECA’s name, this questionnaire is identical to the modified WAI-SR collected with the Exercise Counseling Corpus (Section 3.2.2).

Prompt	Low	High
The counselor looks at me...	Too little	Too much
The counselor changes her body posture...	Too little	Too much
The counselor smiles at me...	Too little	Too much
The counselor nods her head...	Too little	Too much
The counselor's speech is...	Too slow	Too fast

Table 8.1: Semantic differential items used to assess behavioral realism after each conversation.

Perceived nonverbal behavior was assessed with the Self-Report of Immediacy Behaviors (SRIB) questionnaire [139], following each conversation. This and related measures have been used in multiple studies examining the relationship between perceived nonverbal behavior and outcomes, although primarily within education. The SRIB was slightly modified, changing instructions that refer to “supervisor” to refer to “counselor.”. I removed two items related to the use of touch, as the counselor has no physical presence or ability to touch. I removed one item related to the use of hands for gesture, as the counselor’s hands are never visible.

Perceived gestalt (overall) impression of behavioral realism was assessed with a set of 7-point Likert-scale items used by Guadango et al. [69], slightly modified to replace the term “virtual person” with the more specific “virtual counselor,” and to remove an additional item (“I felt that the movement of the virtual person was controlled by a real person”) not relevant to this study. The original items had acceptable reliability ($\alpha = 0.72$) and were successfully used as a manipulation check in the cited study.

To assess the perceived behavioral realism of specific conversational behavior changes, a set of single-item 5-point rating scales was administered after each conversation, asking about the counselor’s gaze, posture shifts, facial expressions, head nods, and speech rate (Table 8.1). Participants were asked, for each behavior, to compare the frequency or prominence of the ECA’s conversational behavior to a human counselor.

Attitudes toward physical activity were assessed based on constructs from

the Transtheoretical Model of Behavior Change [132, 134, 133]. The same measures used when collecting the Exercise Counseling Corpus (Section 3.2.4) were applied here: Stage of Change was assessed at screening and after the final conversation, using a widely-used questionnaire [103]. Decisional balance (i.e., the subjective importance placed on the pros and cons of a behavior change) was assessed prior to the first conversation (although after screening and consent), and following the final conversation, also using a widely-used questionnaire [121].

8.4.2 Procedure

The study was designed to be entirely web-based and heavily automated, including intake, screening, and consent: no direct communication between researchers and participants occurred, and I did not manually intervene at any point except in the case of technical problems (e.g., software version incompatibilities). However, researchers were not technically blinded to the assignment of participants to conditions.

Participants were recruited through flyers and through classified ads (at <http://boston.craigslist.org>), both of which directed potential participants to a web page for intake. Potential participants were shown basic information about the study, including required tasks (interacting with the ECA and answering questionnaires), duration, eligibility requirements, and reimbursement. After clicking a button to indicate they were “interested” in the study, potential participants were administered a screening questionnaire to check whether they met eligibility requirements: (a) 18 years of age or older; (b) a native English speaker; (c) able to regularly access the study website with a computer and web browser capable of displaying the ECA; (d) able to safely begin moderate physical activity, according to the Physical Activity Readiness Questionnaire (PAR-Q) [159]; and (e) not currently in the Action or Maintenance stage of change with regards to regular exercise (defined as

20-60 minutes, 3-5 times per week [103]).

Potential participants who passed the screening questionnaire were asked to read and agree to a consent form, and those who agreed were asked to supply their given name (used in conversation with the ECA) and an email address and password. Upon responding to a confirmation email, participants were considered to be enrolled in the study, were randomly assigned (by simple randomization, with independent random draws) to one of the three conditions, and immediately received a demographics questionnaire and the initial decisional balance assessment, followed by the first conversation with the ECA. After the conversation, participants received the SRIB, WAI-SR, and behavioral realism questionnaire, and the first session was concluded with a message asking participants to return the following week.

For the remaining five weeks of the study, participants were sent a reminder email at the beginning of each week asking them to log in to the study website for their weekly session. Participants were free to complete their sessions at any day or time during the week, although a second reminder email was sent two days prior to the end of the week if a participant had not yet completed a session. Each reminder email included a link which could be used to indicate a desire to withdraw from the study; participants who did so would no longer receive reminder emails. Participants who did not visit the study website for two consecutive weeks were also considered to have withdrawn from the study, and did not receive further reminder emails. In both cases, however, a participant who chose to visit the study website within the six-week duration of their participation was allowed to resume their participation in the study, including eligibility for reimbursement (detailed below).

Each weekly session after the first week consisted of a conversation with the ECA, followed by the SRIB, WAI-SR, and behavioral realism questionnaires. The final (sixth) session additionally included the Stage of Change and decisional balance questionnaires.

Each week, all participants who completed a weekly session were entered in

a random drawing, from which a single participant was selected (participants were informed of the drawing during recruitment and intake). Each drawing was independent, and participants were eligible to win on multiple weeks, although the odds of winning varied depending on the number of participants in the study. The winner received a \$50 gift card to Amazon.com via email, along with a congratulatory message and a reminder to continue participating to remain eligible for the prize drawing.

8.5 Hypotheses

Following the main thesis of this dissertation, I predicted that modeling realistic behavior changes in the ECA would promote user engagement, rapport, voluntary system usage, and voluntary retention in an agent-based intervention. Therefore, I predicted positive effects of the **Dynamic** condition relative to **Static** on the frequency of voluntary system usage and on the bond component of therapeutic alliance (a measure of engagement and rapport). As the conversational behavior changes performed by the agent were gradual, I specifically predicted a more positive rate of change; this also reflects the increase over time in self-reported therapeutic alliance observed in the corpus (Section 3.4.2).

However, I provide no specific hypotheses regarding the effect of the **Exaggerated** condition on engagement or usage relative to either **Static** or **Dynamic**, and consider it to be exploratory rather than confirmatory for the purposes of this study. I also provide no specific hypotheses regarding the task or goal components of therapeutic alliance.

H1 Participants in the **Dynamic** condition will have a more positive rate of change in the frequency of their voluntary system usage than those in the **Static** condition.

H2 Participants in the **Dynamic** condition will have a more positive rate of change in the bond subscale of their self-reported therapeutic alliance (WAI-SR) scores.

Most of the changes in behavior that the models predict will occur in later sessions correspond to a decrease in the prevalence of behaviors associated with nonverbal immediacy [5], including fewer postural shifts, fewer smiles and other facial expressions, and less gaze by the speaker toward the listener. As the SRIB questionnaire used in this study asks participants to judge the prevalence of behaviors associated with nonverbal immediacy (without interpreting those behaviors), I treated the perceived nonverbal immediacy of the agent as a test of the ability of users to reliably perceive and report these changes in behavior, and predicted that users would report decreases in immediacy over time in the conditions with behavior changes.

H3 Participants will report lower immediacy (SRIB) scores over time, in (a) the **Dynamic** condition as compared to **Static**, and (b) in the **Exaggerated** condition as compared to **Dynamic**.

I also predicted, as part of the thesis of this dissertation, that conversational behavior that shows long-term changes would be perceived as more realistic. The **Dynamic** condition was designed to more closely match observed behavior in human interaction than the comparison conditions, and both measures of behavioral realism used in the study ask participants to compare the ECA's behavior to their expectations of human behavior in a similar situation. I hypothesized that this would hold both with gestalt or overall realism, and with the perceived realism of specific aspects of behavior.

H4 Participants will report higher gestalt behavioral realism in the **Dynamic** condition compared to the **Static** or **Exaggerated** conditions.

H5 Participants will be more likely to report that the agent's conversational behaviors are "about right" (at the midpoint) in the **Dynamic** condition

compared to the **Static** or **Exaggerated** conditions.

Following previous findings on the importance of engagement to behavior change, I hypothesized that the **Dynamic** condition would show positive effects on attitudes toward physical activity, which is operationalized in terms of Stage of Change assessments. As above, I provide no specific hypotheses regarding the **Exaggerated** condition.

H6 Participants in the **Dynamic** condition, as compared to the **Static** condition, will be more likely to advance in stage of change from the baseline assessment at intake to the final assessment.

Finally, and also following previous findings on the importance of engagement to behavior change, I hypothesized that participants in the **Dynamic** condition would report more walking behavior.

H7 Participants in the **Dynamic** condition, as compared to the **Static** condition, will report performing more scheduled bouts of walking.

8.6 Results

8.6.1 Screening and Intake

While intake for the study was open, a total of $N = 216$ potential participants visited the study website and completed the screening questionnaire. Of these, 193 (89.4%) were eligible. Of the remaining 23 (10.6%) ineligible visitors, 8 (34.8%) were not native speakers of English, 13 (56.5%) did not have regular access to a suitable computer and/or internet connection, and 3 (13.0%) answered affirmatively to one or more of the PAR-Q questions, indicating that he or she should not begin an exercise question without consulting a doctor.⁶

Following the screening questionnaire, 88 participants (40.7% of the original 216) agreed to the consent form. Eighty three (83, 38.4%) signed up for the

⁶The sum of the percentages exceeds 100% due to some visitors being ineligible for multiple reasons

		Static	Dynamic	Exaggerated	Total
N		25	20	17	62
Age		35.2 (12.6)	38.1 (16.4)	41.8 (9.0)	37.9 (13.2)
Gender	Female	21 (84.0%)	16 (80.0%)	13 (76.5%)	50 (80.6%)
	Male	4 (16.0%)	4 (20.0%)	4 (23.5%)	12 (19.4%)
Education	High School	0 (0.0%)	2 (10.0%)	1 (5.9%)	3 (4.8%)
	Tech./Voc.	1 (4.0%)	0 (0.0%)	0 (0.0%)	1 (1.6%)
	Some College	9 (36.0%)	4 (20.0%)	6 (35.3%)	19 (30.6%)
	College Grad.	10 (40.0%)	10 (50.0%)	7 (41.2%)	27 (43.5%)
	Adv. Degree	5 (20.0%)	4 (20.0%)	3 (17.6%)	12 (19.4%)
Ethnicity	Am. Indian	1 (4.0%)	0 (0.0%)	0 (0.0%)	1 (1.6%)
	Asian	1 (4.0%)	2 (10.0%)	1 (5.9%)	4 (6.5%)
	Black	2 (8.0%)	4 (20.0%)	2 (11.8%)	8 (12.9%)
	White	19 (76.0%)	14 (70.0%)	13 (76.5%)	46 (74.2%)
	Hispanic	2 (8.0%)	0 (0.0%)	1 (5.9%)	3 (4.8%)
Marital Status	Single	16 (64.0%)	13 (65.0%)	6 (35.3%)	35 (56.5%)
	Married	4 (16.0%)	1 (5.0%)	8 (47.1%)	13 (21.0%)
	Divorced	4 (16.0%)	5 (25.0%)	2 (11.8%)	11 (17.7%)
	Other	1 (4.0%)	1 (5.0%)	1 (5.9%)	3 (4.8%)

Table 8.2: Demographics of participants, by condition and overall.

study, providing a given name and email address. Seventy two (72, 33.3%) responded to an email request for confirmation. Finally, 62 (28.7%) completed initial demographics questionnaires and began the first weekly session; this last group was considered as recruited into the study.

8.6.2 Participants and Demographics

The total of $N = 62$ were randomized into one of the three study conditions. The group of participants was slightly unbalanced across conditions: participants were randomized after consent into balanced groups, but a fraction of participants who consented to the study failed to respond to confirmation messages or begin the first study session. The **Static** condition was the largest sub-group ($n = 25$), followed by **Dynamic** ($n = 20$) and **Exaggerated** ($n = 17$). Since participants who did not return after consent did so before receiving any intervention content or any interaction with the agent, these participants are ignored in all subsequent analysis.

	Static	Dynamic	Exaggerated	Total
Precontemplation	2 (8.0%)	2 (10.0%)	1 (5.9%)	5 (8.1%)
Contemplation	7 (28.0%)	9 (45.0%)	6 (35.3%)	22 (35.5%)
Preparation	16 (64.0%)	9 (45.0%)	10 (58.8%)	35 (56.5%)

Table 8.3: Stage of Change, at intake.

Table 8.2 gives demographic information for participants. The participants were largely female (80.6%), and most either were college graduates (62.9%, including those with advanced degrees) or currently in college (30.6%). The mean age was 37.9 years (SD 13.2) and the majority (56.5%) had never been married.

There were no significant differences observed between conditions in participants' gender ($\chi^2(2) = 0.375$, $p = 0.83$), ethnic background ($\chi^2(8) = 4.97$, $p = 0.76$), or education level ($\chi^2(8) = 5.24$, $p = 0.73$). There was a near-significant difference in age (Kruskal-Wallis $\chi^2(2) = 4.68$, $p = 0.10$); participants in the **Exaggerated** condition tended to be older. There was a near-significant difference in marital status ($\chi^2(6) = 11.16$, $p = 0.08$); fewer participants in the **Exaggerated** condition were single, and more were married than in other conditions.

8.6.3 Baseline Assessments

At the time of intake, the majority of participants were in the preparation stage of change (Table 8.3). Few participants were in precontemplation. There were no significant differences observed between groups ($\chi^2(4) = 1.83$, $p = 0.77$).

Participants reported high scores on the decisional balance pros (mean 3.90, SD 0.70 on a 1–5 scale) and low scores on the cons (mean 1.91, SD 0.78) at intake. There were no significant differences observed between groups (Figure 8.4) for either the pros (Kruskal-Wallis $\chi^2(2) = 2.20$, $p = 0.33$) or the cons (Kruskal-Wallis $\chi^2(2) = 0.43$, $p = 0.80$).

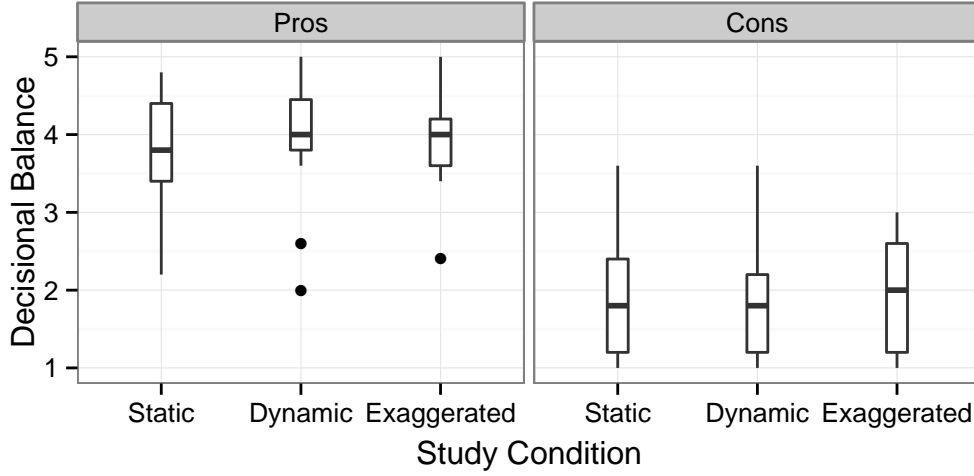


Figure 8.4: Decisional Balance at intake.

8.6.4 System Usage

The amount of voluntary system usage is operationalized as a binary outcome, per participant and per week, of whether or not a weekly conversation with the agent was completed. I exclude the first conversation as it immediately followed study intake and did not require participants to make a separate decision to log in to the study website; failure to complete the first conversation, once begun, was rare ($N = 2$) and appeared to be due to technical problems (e.g., a software version incompatibility) in both cases.

Each participant had five subsequent opportunities to log in and complete a weekly conversation with the agent. Out of the total $N = 310$ opportunities for a conversation, $N = 155$ (50%) were completed.

Figure 8.5 shows the proportion of completed weekly conversations, separately by study week and by study condition. Participants in the **Exaggerated** condition had the lowest overall proportion of completed conversations, completing 32.9% (28 of 85) opportunities, compared to 55.2% (69 of 125) for the **Static** condition and 58.0% (58 of 100) for the **Dynamic** condition. Comparing weeks, the highest proportion of participants completed the second week, which was the first voluntary conversation (37 of 62, 59.7%), and

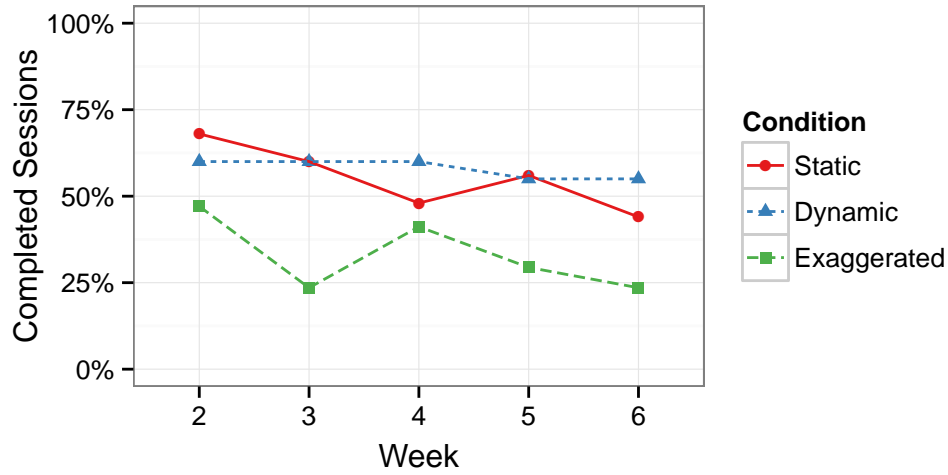


Figure 8.5: Proportion of participants completing a weekly conversation, by week and by condition.

Parameter		Estimate	95% CI	<i>p</i>
Fixed Effects	Intercept	2.061	[0.063, 4.866]	0.043*
	Dynamic	-0.968	[-4.457, 2.375]	0.570
	Exaggerated	-3.559	[-8.288, -0.529]	0.020*
	Week	-0.656	[-1.182, -0.216]	0.002**
	Week · Dynamic	0.485	[-0.166, 1.191]	0.138
	Week · Exaggerated	0.206	[-0.566, 0.970]	0.575
Random Intercept (SD)		3.942	[2.505, 5.606]	

Table 8.4: Random-intercept logistic regression fit to the probability of completing a session, with a linear trend over time. Confidence intervals are estimated from a parametric bootstrap (5000 replicates). * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

the lowest proportion completed the sixth and final week (26 of 62, 41.9%).

To test the effects of study condition on system usage, we fit a random-intercept mixed effect logistic regression model (Table 8.4), including fixed effects of study condition, study week (indexed so that week 2 was the “first” week), and interaction terms, with the latter allowing different rates of change in the probability of completing a conversation across study conditions. Taking the **Static** condition as the reference category, there is a significantly lower probability of participants in the **Exaggerated** condition completing a con-

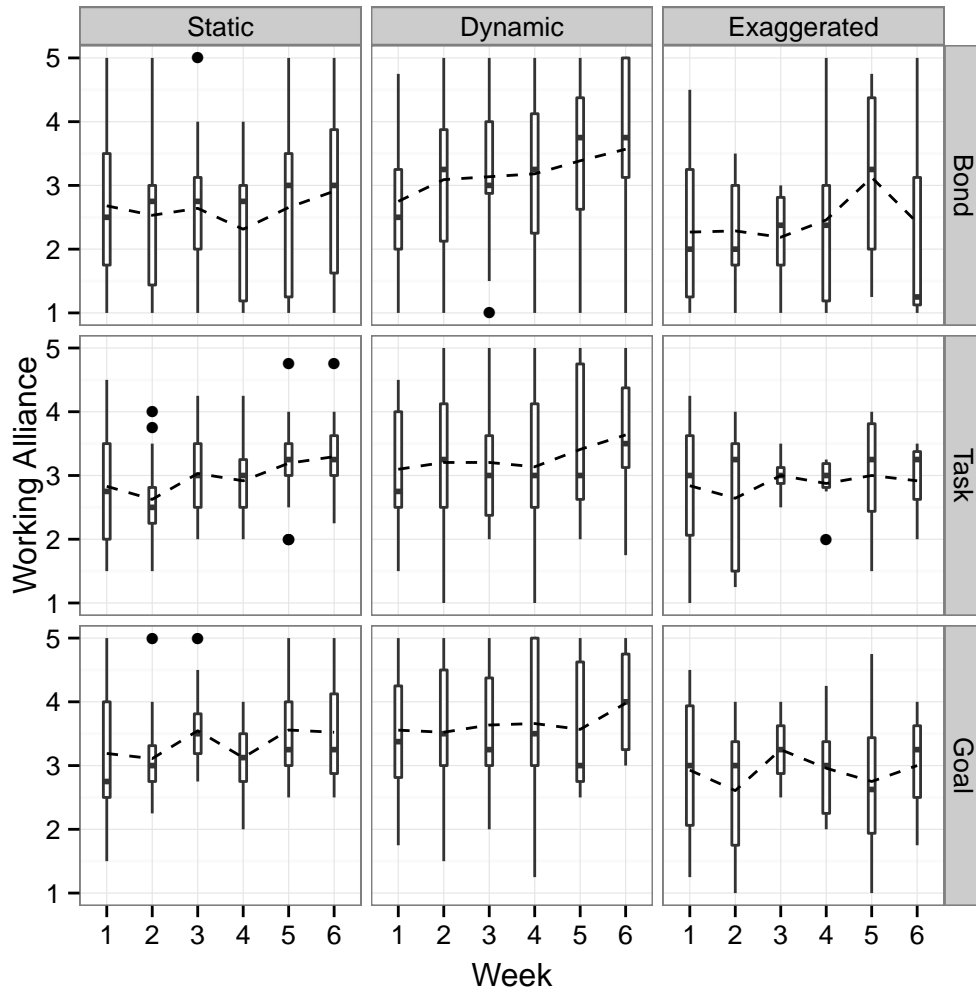


Figure 8.6: Working Alliance Inventory (Short Revised) scores, by week and study condition. The dotted line indicates group means.

version ($\beta = -3.559$, 95% CI $[-8.288, -0.529]$, $p = 0.020$). There is a significant decrease in the probability of completing a conversation in later weeks (linear trend, $\beta = -0.656$, 95% CI $[-1.182, -0.216]$, $p = 0.002$). No significant differences were found in the rate of this decrease in the **Dynamic** or **Exaggerated** conditions, as compared to **Static**.

8.6.5 Therapeutic Alliance

Figure 8.6 shows self-reported weekly WAI-SR scores, aggregated by week, by study condition, and by the components of therapeutic alliance (bond, task,

Parameter		Estimate	95% CI	<i>p</i>
Fixed Effects	Intercept	2.613	[2.179, 3.045]	< 0.001 ***
	Dynamic	0.212	[-0.431, 0.861]	0.521
	Exaggerated	-0.276	[-0.986, 0.427]	0.453
	Week	0.014	[-0.053, 0.079]	0.655
	Week · Dynamic	0.111	[0.015, 0.207]	0.027 *
	Week · Exaggerated	0.018	[-0.113, 0.152]	0.789
Random Intercept (SD)		1.017	[0.809, 1.201]	
Residual (SD)		0.512	[0.444, 0.575]	

Table 8.5: Random-intercept regression model fit to a linear trend on the working alliance bond subscale. Confidence intervals are estimated from a residual-resampling bootstrap (5000 replicates). * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

and goal). Overall (i.e., across all three study conditions), self-reported bond increased from the first conversation (mean 2.60, SD 1.08 on a 1–5 scale) to the last (mean 3.14, SD 1.50). The task (mean 2.92, SD 0.91 to mean 3.40, SD 0.88) and goal (mean 3.24, SD 1.00 to mean 3.66, SD 0.90) components also increased.

To test the effect of study condition on changes in the therapeutic alliance bond component, I fit a random-intercept mixed-effect linear regression model to the WAI-SR bond subscale scores (Table 8.5), including fixed effects of study condition, study week, and interaction terms, with the latter allowing different rates of change per study condition. Taking the **Static** condition as the reference category, I find a significantly greater rate of increase in WAI-SR bond scores for the **Dynamic** condition ($\beta = 0.111$, 95% CI [0.015, 0.207], $p = 0.027$). No significant difference was found in the rate of increase for the **Exaggerated** condition ($\beta = 0.018$, 95% CI [-0.113, 0.152], $p = 0.789$). There was a large amount of between-participant variability (SD=1.017, 95% CI [0.809, 1.201]) and a large amount of residual variability (SD=0.512, 95% CI [0.444, 0.575]).

Similar regression models were fit to the task and goal subscales (Table 8.6 and Table 8.7). There was a significant increase in scores on the task subscale

Parameter		Estimate	95% CI	<i>p</i>
Fixed Effects	Intercept	2.757	[2.429, 3.089]	< 0.001 ***
	Dynamic	0.355	[-0.144, 0.863]	0.186
	Exaggerated	0.056	[-0.500, 0.598]	0.847
	Week	0.094	[0.029, 0.159]	0.005 **
	Week · Dynamic	-0.008	[-0.103, 0.089]	0.886
	Week · Exaggerated	-0.085	[-0.215, 0.048]	0.215
Random Intercept (SD)		0.731	[0.574, 0.867]	
Residual (SD)		0.515	[0.438, 0.590]	

Table 8.6: Random-intercept regression model fit to a linear trend on the working alliance task subscale. Confidence intervals are estimated from a residual-resampling bootstrap (5000 replicates). * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

Parameter		Estimate	95% CI	<i>p</i>
Fixed Effects	Intercept	3.158	[2.812, 3.499]	< 0.001 ***
	Dynamic	0.341	[-0.189, 0.878]	0.195
	Exaggerated	-0.245	[-0.822, 0.322]	0.403
	Week	0.061	[-0.011, 0.092]	0.824
	Week · Dynamic	-0.012	[-0.012, 0.092]	0.824
	Week · Exaggerated	-0.052	[-0.197, 0.099]	0.476
Random Intercept (SD)		0.750	[0.587, 0.887]	
Residual (SD)		0.564	[0.494, 0.633]	

Table 8.7: Random-intercept regression model fit to a linear trend on the working alliance goal subscale. Confidence intervals are estimated from a residual-resampling bootstrap (5000 replicates). * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

over time ($\beta = 0.094$, 95% CI [0.029, 0.159], $p = 0.005$). There was no significant difference between study conditions. On the goal subscale, there was no significant change over time and no significant difference between study conditions.

8.6.6 Perceived Behavior and Immediacy

Figure 8.7 shows self-reported weekly immediacy (SRIB) scores, aggregated by week and by study condition. Overall, participants report a midrange amount

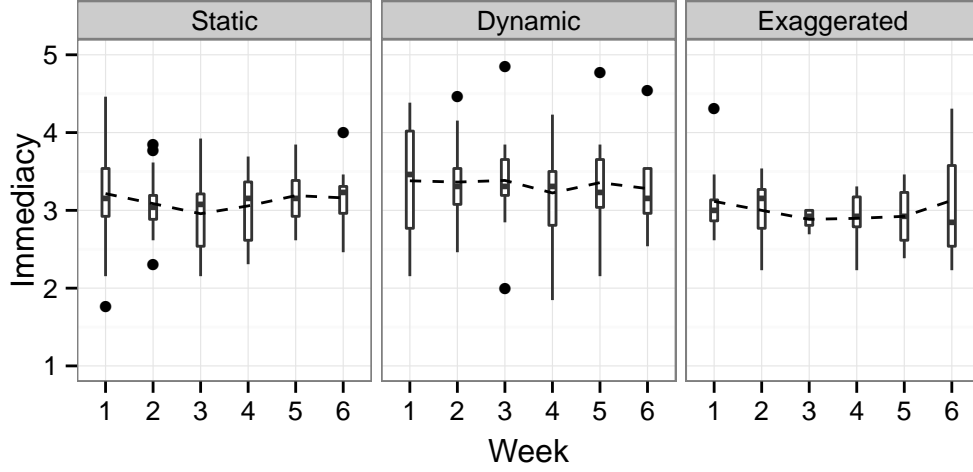


Figure 8.7: Self-Report of Immediacy Behavior (SRIB) scores, by week and by study condition. The dotted line indicates group means.

Parameter		Estimate	95% CI	<i>p</i>
Fixed Effects	Intercept	3.144	[2.949, 3.345]	<0.001***
	Dynamic	0.241	[-0.064, 0.544]	0.125
	Exaggerated	-0.050	[-0.388, 0.302]	0.762
	Week	-0.014	[-0.053, 0.023]	0.456
	Week · Dynamic	-0.024	[-0.079, 0.034]	0.423
	Week · Exaggerated	-0.039	[-0.116, 0.037]	0.318
Random Intercept (SD)		0.452	[0.330, 0.557]	
Residual (SD)		0.295	[0.256, 0.332]	

Table 8.8: Random-intercept regression model fit to a linear trend on self-reported immediacy (SRIB). Confidence intervals are estimated from a residual-resampling bootstrap (5000 replicates). * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

of immediacy in the agent’s behaviors (mean 3.18, SD 0.54 on a 1–5 scale). The reports were similar across conditions and across weeks.

To test the effect of study condition on perceived immediacy, and on the change in perceived immediacy over time, I fit a random-intercept mixed-effect linear regression model to the SRIB scores (Table 8.8), including fixed effects of study condition, study week, and interaction terms. As the agent’s models of nonverbal behavior included differential behavior in the final session, I also

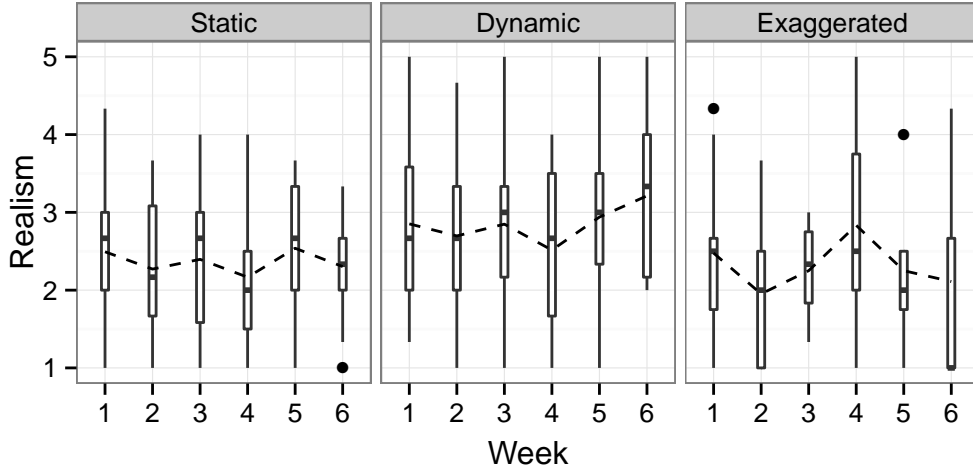


Figure 8.8: Gestalt behavioral realism scores, by study condition and by week. The dotted line indicates group means.

fit a variant model (not shown) which included a fixed effect of the last week. However, a likelihood ratio test (estimated by residual-resampling bootstrap, 5000 replicates) did not indicate a significantly better fit to the data for the more complex model ($\chi^2(3) = 4.92$, $p = 0.18$), and the more complex model gave similar parameter estimates and confidence bounds. In both models, no significant changes were found over time in SRIB scores, and no significant differences emerged between conditions, either in baseline reports (intercept) or in rates of change.

8.6.7 Behavioral Realism (Gestalt)

Figure 8.8 gives gestalt behavioral realism scores, aggregated by study condition and by week. Overall, participants reported the agent had low to medium realism (mean 2.54, SD 1.02 on a 1–5 scale). Of the three conditions, the highest realism was reported in the **Dynamic** condition (mean 2.84, SD 1.07) compared to both the **Static** (mean 2.38, SD 0.87) and **Exaggerated** (mean 2.36, SD 1.13) conditions.

To test the effect of study condition on perceived realism, I fit a random-intercept mixed effect regression model to the reported behavioral realism

Parameter		Estimate	95% CI	<i>p</i>
Fixed Effects	Intercept	2.455	[2.094, 2.821]	<0.001***
	Dynamic	0.326	[-0.227, 0.905]	0.263
	Exaggerated	0.008	[-0.603, 0.626]	0.996
	Week	-0.016	[-0.083, 0.050]	0.630
	Week · Dynamic	0.022	[-0.080, 0.120]	0.654
	Week · Exaggerated	-0.010	[-0.145, 0.125]	0.880
Random Intercept (SD)		0.842	[0.651, 1.006]	
Residual (SD)		0.517	[0.423, 0.614]	

Table 8.9: Random-intercept regression model fit to a linear trend on gestalt behavioral realism scores. Confidence intervals are estimated from a residual-resampling bootstrap (5000 replicates). * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

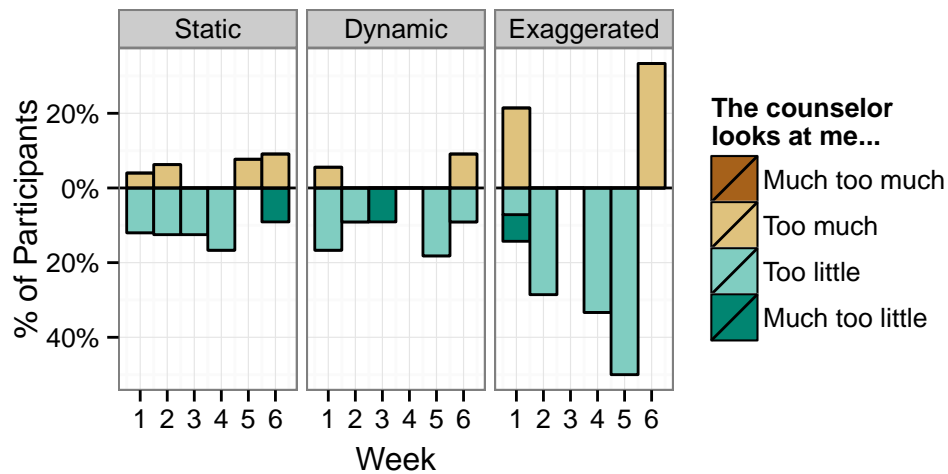


Figure 8.9: The proportion of participants reporting the agent gazes at them “too little” or “too much,” by study condition and week.

scores (Table 8.9), including fixed effects of study condition, week, and interaction terms. No significant differences between conditions were found, and no significant changes over time (linear trend) were found.

Parameter		Estimate	95% CI	<i>p</i>
Threshold	1—2	-6.065	$[-\infty, -4.793]$	< 0.001 ***
	2—3	3.136	$[-8.891, -2.107]$	< 0.001 ***
	3—4	4.753	$[3.593, 10.837]$	< 0.001 ***
Fixed Effects	Dynamic	0.189	$[-2.178, 2.500]$	0.882
	Exaggerated	0.681	$[-1.679, 3.544]$	0.509
	Week	0.037	$[-0.574, 0.641]$	0.902
	Week · Dynamic	-0.026	$[-1.011, 0.898]$	0.940
	Week · Exaggerated	-0.641	$[-2.240, 0.405]$	0.169
	Final	0.082	$[-2.912, 3.372]$	0.961
	Final · Dynamic	0.780	$[-3.418, 6.511]$	0.655
	Final · Exaggerated	6.970	$[2.028, 36.596]$	0.010 **
Random Intercept (SD)		2.253	$[1.482, 9.140]$	

Table 8.10: Random-intercept ordinal regression model fit to participant’s reports of the prevalence of the agent’s gaze, with linear trends over weeks and an effect of the final week. Confidence intervals are estimated from a parametric bootstrap (2000 replicates). * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

8.6.8 Behavioral Realism (Specific Behaviors)

8.6.8.1 Gaze

Figure 8.9 shows the proportion of participants (by condition and week) reporting the agent gazed at them “too little” or “too much”; the agent’s actual behavior was to gaze away from the participant more frequently in later sessions in the **Dynamic** and **Exaggerated** conditions, except for the last week. Overall, participants reported that the agent’s gaze was “about right” in 82.4% of reports (168 of 204), and participants in the **Exaggerated** condition were least likely to do so (68.4%, 26 of 38) compared to both the **Static** (84.9%, 79 of 93) and **Dynamic** (86.3%, 63 of 73) conditions.

To test for systematic changes in perceived gaze behavior over time and between study conditions, I fit a random-intercept mixed-effect ordinal logistic regression model (Table 8.10), including fixed effects of study condition, study week, whether it was the final week (“Final” in Table 8.10), and interactions of study condition with both of the latter variables. The “final” variable was

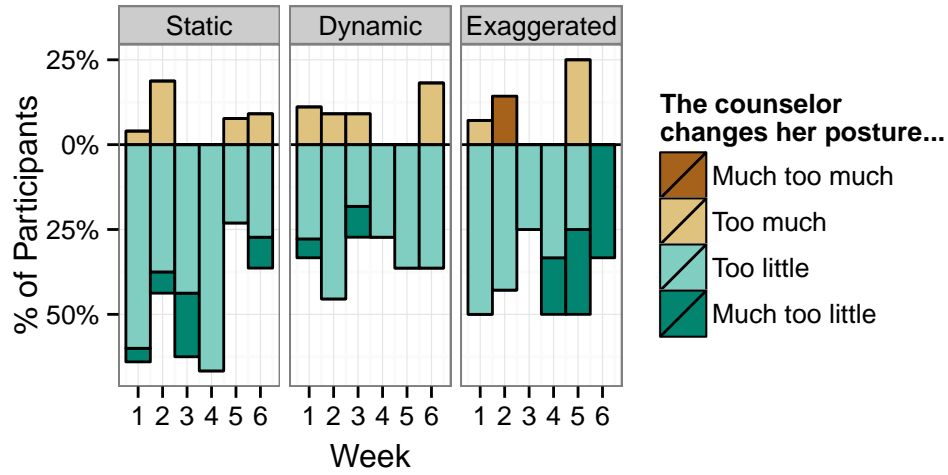


Figure 8.10: The proportion of participants reporting the agent shifts her posture “too little” or “too much,” by study condition and week.

added to allow for differing perceptions of agent behavior in the final week, as the agent’s actual behavior changed in this week.

There is a trend toward reporting less gaze in later sessions in the **Exaggerated** condition (matching the agent’s actual behavior), however it is not statistically significant ($\beta = -0.641$, 95% CI $[-2.240, 0.450]$, $p = 0.169$). Participants in the **Exaggerated** condition reported significantly more gaze in the last week relative to the overall trend ($\beta = 6.970$, 95% CI $[2.028, 36.596]$, $p = 0.010$). This also matches the agent’s actual behavior, but I note that it is based only on reports of the small proportion of participants in that condition who completed the final session. No other significant effects were observed.

8.6.8.2 Posture Shifts

Figure 8.10 shows the proportion of participants, by study condition and week, reporting that the agent changes her posture “too little” or “too much”; the agent’s actual behavior was to change her posture less frequently in later weeks in the **Dynamic** and **Exaggerated** conditions. Aggregating all reports together, participants most commonly reported that the agent’s frequency of posture changes was “about right” (48.5%, 99 of 204) or “too little” (38.7%,

Parameter		Estimate	95% CI	p
Threshold	1—2	-3.551	[-5.213, -2.413]	< 0.001 ***
	2—3	0.492	[-0.601, 1.637]	0.359
	3—4	4.837	[3.599, 6.811]	< 0.001 ***
	4—5	8.175	[6.438, $+\infty$]	< 0.001 ***
Fixed Effects	Dynamic	1.425	[-0.320, 3.303]	0.104
	Exaggerated	1.176	[-0.656, 3.217]	0.197
	Week	0.319	[0.055, 0.618]	0.023 *
	Week · Dynamic	-0.284	[-0.722, 0.110]	0.149
	Week · Exaggerated	-0.511	[-1.094, 0.001]	0.051
Random Intercept (SD)		2.092	[1.361, 3.021]	

Table 8.11: Random-intercept ordinal regression model fit to participant’s reports of the prevalence of the agent’s posture changes, with linear trends over weeks. Confidence intervals are estimated from a parametric bootstrap (2000 replicates). * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

79 of 204). Participants in the **Dynamic** condition more frequently reported that the agent’s posture changes were “about right” (57.5%, 42 of 73) than those in the **Static** (41.9%, 39 of 93) or **Exaggerated** (47.4%, 18 of 37) conditions.

To test for systematic changes in perceived posture shifts over time and between study conditions, I fit a random-intercept mixed-effect ordinal logistic regression model (Table 8.11), with fixed effects of study week and study condition, and interaction terms. Note that unlike the model used above to examine gaze, I do not include a separate fixed effect from the last week, as the agent’s rate of posture shifts did not depart from the overall trend in the last week.

In the **Static** condition (taken as the reference category in this model), participants reported significantly more perceived posture shifts relative to expectations in later weeks ($\beta = 0.319$, 95% CI [0.055, 0.618], $p = 0.023$). For participants in the **Exaggerated** condition, this trend was in the opposite direction, with borderline significance ($\beta = -0.511$, 95% CI [-1.094, 0.001], $p = 0.051$). There was a trend (borderline near-significant) for participants in

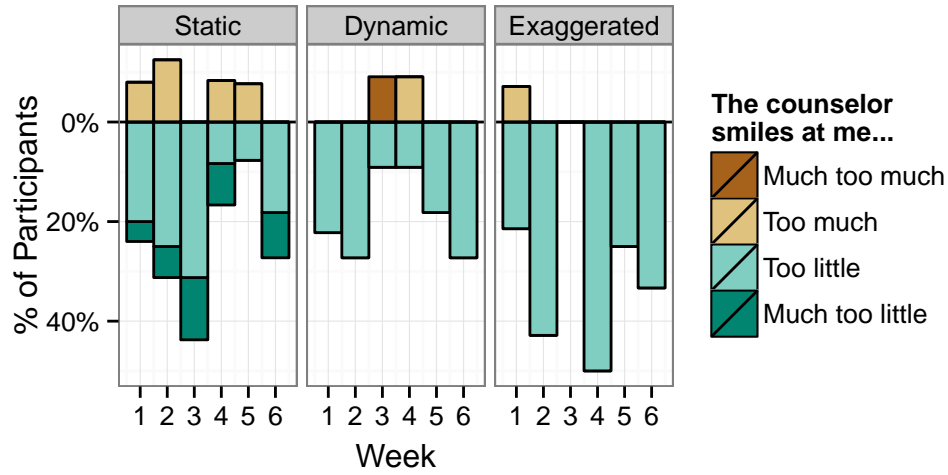


Figure 8.11: The proportion of participants reporting that the agent smiles “too little” or “too much,” by study condition and week.

the **Dynamic** condition to report more perceived posture shifts ($\beta = 1.425$, 95% CI $[-0.320, 3.303]$, $p = 0.104$).

8.6.8.3 Facial Expressions

Figure 8.11 shows the proportion of participants, by study condition and week, reporting that the agent smiled at them “too little” or “too much”; the agent’s actual behavior was to smile (and frown) less frequently in later weeks in the **Dynamic** and **Exaggerated** conditions, with the exception of the final week. Overall, participants most frequently reported that the agent’s smiles were “about right” (71.6%, 146 of 204) or “too little” (21.1%, 43 of 204). Participants in the **Dynamic** condition more frequently reported that the agent’s smiles were “about right” (78.1%, 57 of 73) than those in the **Static** (67.8%, 63 of 93) or **Exaggerated** (68.4%, 26 of 38) conditions.

To test for systematic changes in perceived facial expressions over time and between study conditions, I fit a random-intercept mixed-effect ordinal logistic regression model (Table 8.12), including fixed effects of study condition, study week, whether it was the final week, and interactions of study condition with both of the latter variables. Taking the **Static** condition as the reference

Parameter		Estimate	95% CI	p
Threshold	1—2	-6.491	$[-\infty, -4.759]$	$< \mathbf{0.001}^{***}$
	2—3	-1.654	$[-3.018, -0.354]$	$\mathbf{0.011}^*$
	3—4	5.179	$[3.618, 7.893]$	$< \mathbf{0.001}^{***}$
	4—5	8.106	$[6.007, +\infty]$	$< \mathbf{0.001}^{***}$
Fixed Effects	Dynamic	0.132	$[-1.792, 2.315]$	0.883
	Exaggerated	0.456	$[-1.555, 2.858]$	0.650
	Week	0.201	$[-0.242, 0.678]$	0.346
	Week · Dynamic	0.103	$[-0.600, 0.801]$	0.779
	Week · Exaggerated	-0.666	$[-1.577, 0.110]$	0.100
	Final	-1.266	$[-3.707, 1.030]$	0.229
	Final · Dynamic	-0.465	$[-3.896, 2.833]$	0.774
	Final · Exaggerated	2.285	$[-2.481, 6.970]$	0.293
Random Intercept (SD)		2.335	$[1.401, 3.270]$	

Table 8.12: Random-intercept ordinal regression model fit to participant’s reports of the prevalence of smiles by the agent, with linear trends over weeks and an effect of the final week. Confidence intervals are estimated from a parametric bootstrap (2000 replicates). $*p \leq 0.05$, $**p \leq 0.01$, $***p \leq 0.001$.

category, there is a near-significant trend for participants in the **Exaggerated** condition to report less smiling (relative to expectations) over time ($\beta = -0.666$, 95% CI $[-1.577, 0.110]$, $p = 0.100$); the direction of this trend matches the agent’s actual behavior.

8.6.8.4 Nodding

Figure 8.12 shows the proportion of participants reporting that the agent nods “too little” or “too much,” by study condition and week; the agent’s actual behavior was to nod more frequently in later weeks in the **Dynamic** and **Exaggerated** conditions. Overall, participants most frequently reported that the agent’s nodding was “about right” (62.7%, 128 of 204) compared to “too little” (12.3%, 25 of 204) and “too much” (22.5%, 46 of 204) occurred less frequently, and the extreme categories (“much too little,” “much too much”) were rare. Participants reported that the agent’s nodding was “about right” most frequently in the **Exaggerated** condition (68.4%, 26 of 38), followed by

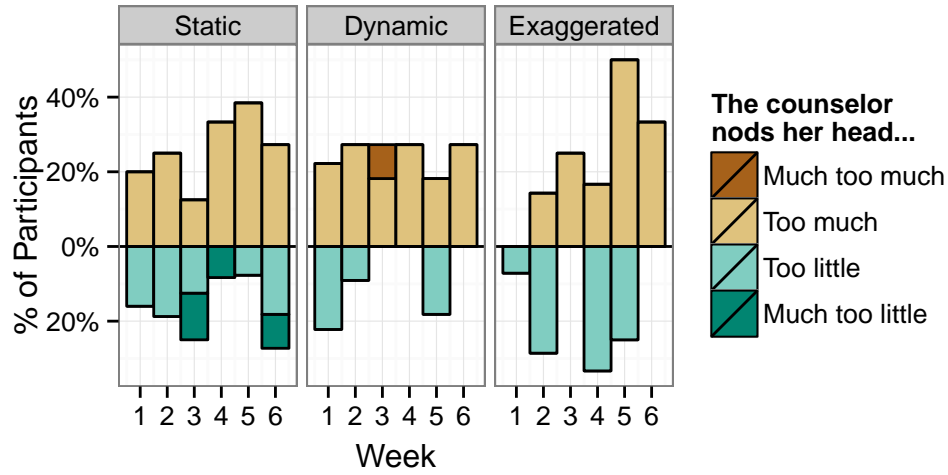


Figure 8.12: The proportion of participants reporting that the agent nods “too little” or “too much,” by study condition and week.

the **Dynamic** (65.8%, 48 of 73) and **Static** (58.1%, 54 of 93) conditions.

To test for systematic changes in perceived nodding over time and between study conditions, I fit a random-intercept mixed-effect ordinal logistic regression model (Table 8.13), including fixed effects of study condition, study week, whether it was the final week, and interactions of study condition with both of the latter variables. No significant differences over time or between conditions were observed.

8.6.8.5 Speech Rate

Figure 8.13 shows the proportion of participants reporting that the agent’s speech was “too fast” or “too slow,” by study condition and week; the agent’s actual behavior was to speak faster, although only on particular words, in later weeks in the **Dynamic** and **Exaggerated** conditions. No participants reported at any time that the agent spoke “too fast” or “much too fast.” Overall, participants most frequently reported that the agent’s speech was “about right” (61.8%, 126 of 204) or “too slow” (31.4%, 64 of 204), while reports of “much too slow” were rare (6.9%, 14 of 204). Participants in the **Dynamic** condition most frequently reported that the agent’s speaking rate

Parameter		Estimate	95% CI	<i>p</i>
Threshold	1—2	-5.963	$[-\infty, -4.644]$	$<0.001^{** *}$
	2—3	-2.926	$[-5.613, -1.822]$	$<0.001^{***}$
	3—4	2.825	$[1.680, 5.550]$	$<0.001^{***}$
	4—5	8.418	$[6.498, +\infty]$	$<0.001^{***}$
Fixed Effects	Dynamic	0.190	$[-1.872, 2.072]$	0.867
	Exaggerated	-0.392	$[-2.747, 1.673]$	0.734
	Week	0.290	$[-0.091, 0.772]$	0.137
	Week · Dynamic	-0.208	$[-0.890, 0.434]$	0.510
	Week · Exaggerated	-0.117	$[-0.951, 0.673]$	0.721
	Final	-1.468	$[-4.055, 0.428]$	0.122
	Final · Dynamic	2.237	$[-0.593, 5.838]$	0.119
	Final · Exaggerated	2.786	$[-1.214, 10.001]$	0.163
Random Intercept (SD)		2.462	$[1.653, 5.200]$	

Table 8.13: Random-intercept ordinal regression model fit to participant’s reports of the prevalence of nodding by the agent, with linear trends over weeks and an effect of the final week. Confidence intervals are estimated from a parametric bootstrap (2000 replicates). $*p \leq 0.05$, $**p \leq 0.01$, $***p \leq 0.001$.

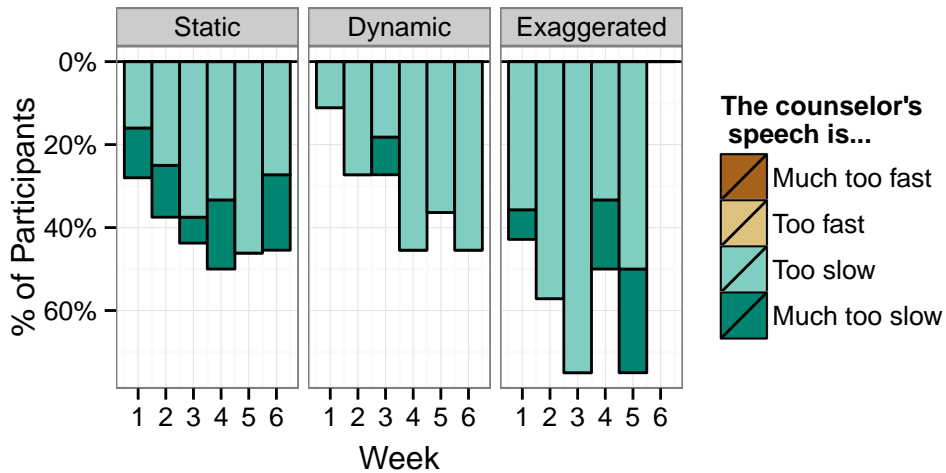


Figure 8.13: The proportion of participants reporting that the agent’s speech was “too fast” or “too slow,” by study condition and week.

Parameter		Estimate	95% CI	p
Threshold	1—2	-4.888	[-7.803, -3.396]	< 0.001 ***
	2—3	-1.244	[-3.344, 0.0433]	0.0580*
Fixed Effects	Dynamic	2.029	[-0.109, 5.004]	0.067
	Exaggerated	-0.764	[-3.198, 1.66]	0.485
	Week	-0.302	[-0.668, 0.010]	0.060
	Week · Dynamic	-0.135	[-0.716, 0.437]	0.621
	Week · Exaggerated	0.162	[-0.498, 0.843]	0.620
Random Intercept (SD)		2.513	[1.428, 4.138]	

Table 8.14: Random-intercept ordinal regression model fit to participant’s reports of the agent’s speaking rate, with linear trends over weeks. Confidence intervals are estimated from a parametric bootstrap (2000 replicates). * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

was “about right” (69.9%, 51 of 73), followed by the **Static** (60.2%, 56 of 93) and **Exaggerated** (50.0%, 19 of 38) conditions.

To test for systematic changes in perceived speech rate over time and between study conditions, I fit a random-intercept mixed-effect ordinal logistic regression model (Table 8.14), including fixed effects of study condition and study week, and interaction terms between them. Note that I do not include an effect for the final study week, as the agent’s speaking behavior did not deviate from the trend in the last week, unlike some other behaviors tested. There was a near-significant trend for participants to report speech as “too slow” more frequently in later weeks ($\beta = -0.302$, 95% CI [-0.668, 0.010], $p = 0.06$). Taking the **Static** condition as the reference category, there was a near-significant trend for participants in the **Dynamic** condition to report faster perceived speech.

8.6.9 Behavioral Realism (Combined Behaviors)

In the previous section, the results were primarily concerned with testing for systematic changes in perceptions of the agent’s conversational behaviors over time: for example, do participants report that the agent nods more or less in

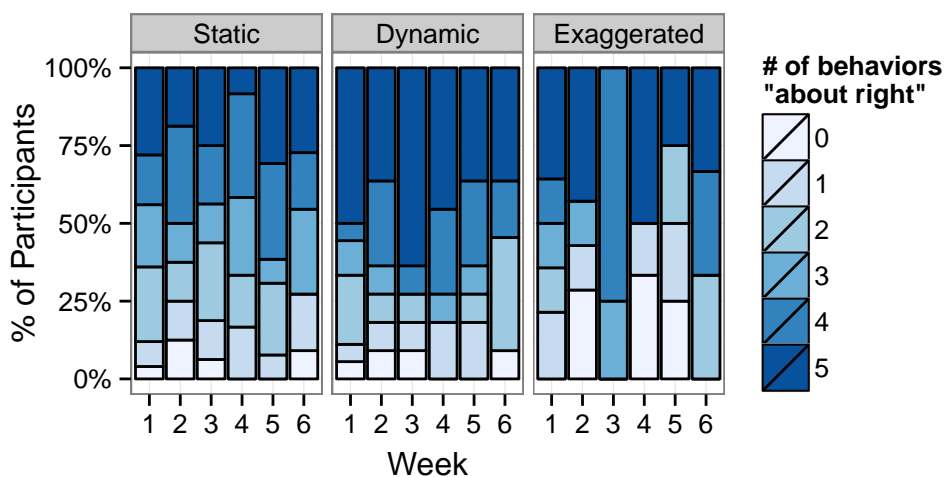


Figure 8.14: Distribution of behavior items rated as realistic, by condition and week.

later weeks, relative to their expectations?

In this section, I examine whether participants reported the agent’s behavior as realistic or unrealistic — that is, approximately matching the expected behavior of a human speaker — across all of the behaviors addressed in the previous section. Reports of realism are treated as a dichotomous outcome: a behavior is considered to be reported as realistic if a participant chose the midpoint (“about right”) rather than any level above or below on any of the relevant items (Table 8.1), and unrealistic otherwise, without concern for the direction of departure from realism (e.g. more or less prevalent).

Overall, participants reported that a behavior was “about right” 65.4% of the time (667 of 1020). Participants in the **Dynamic** condition most frequently reported that a behavior was “about right” (71.5%, 261 of 365), followed by the **Static** (62.6%, 291 of 465) and **Exaggerated** (60.5%, 115 of 190) conditions.

I fit a mixed-effect logistic regression model (Table 8.15), with random effects of the participant, the study session (nested within participants) and of the item (crossed with participants and sessions).⁷ These random effects

⁷As an alternative description, using terminology from Item Response Theory, this

	Parameter	Estimate	95% CI	<i>p</i>
Fixed Effects	Intercept	1.990	[0.568, 3.552]	0.005**
	Static	-0.960	[-2.682, 0.599]	0.230
	Exaggerated	-0.352	[-2.190, 1.507]	0.710
	Week	0.104	[-0.225, 0.428]	0.522
	Week · Static	-0.025	[-0.416, 0.362]	0.921
	Week · Exaggerated	-0.648	[-1.185, -0.121]	0.013*
	Final	-1.308	[-2.824, 0.179]	0.089
	Final · Static	0.905	[-1.005, 2.911]	0.389
	Final · Exaggerated	3.792	[0.947, 21.476]	0.006**
Random Effects (SD)	Participant	2.209	[1.526, 2.728]	
	Session	0.753	[0.000, 0.956]	
	Item	0.876	[0.019, 1.318]	

Table 8.15: Mixed-effect logistic regression model fit to behavioral realism items, with linear trends over weeks and an effect of the final week. Confidence intervals are estimated from a parametric bootstrap (2000 replicates). * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

allow for correlation among answers from, respectively, the same participant, the same sessions within participants, and on the same item. To test the effects of study condition and to test for systematic changes over time, I include fixed effects of study condition, study week, whether it was the final week, and interactions of study condition with both of the latter variables.

Taking the **Dynamic** condition as the reference category⁸ participants in the **Exaggerated** condition report realistic behaviors significantly less frequently in later weeks ($\beta = -0.648$, 95% CI [-1.185, -0.121], $p = 0.013$), but more frequently in the final week ($\beta = 3.792$, 95% CI [0.947, 21.476], $p = 0.006$). There is an overall trend (near-significant) for participants to report less realism in the final week ($\beta = -1.308$, 95% CI [-2.824, 0.179], $p = 0.089$).

model is equivalent to a multilevel Rasch model.

⁸Taking **Static** as the reference category (as done elsewhere) would yield an equivalent model; the use of **Dynamic** here was chosen for easier comparison between the **Dynamic** and **Exaggerated** conditions. Note that the difference between **Static** and **Exaggerated** is not significant.

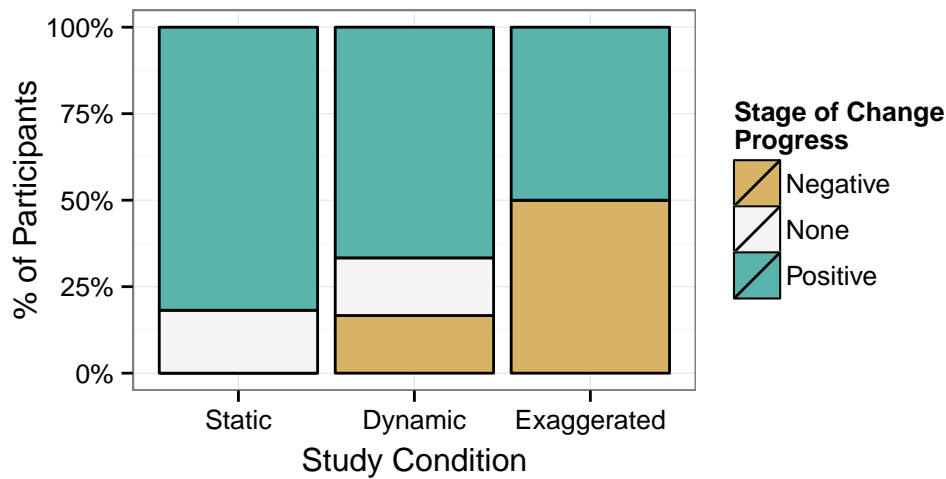


Figure 8.15: The proportion of participants (of those completing the study) who report negative, positive, or no progress on the Stages of Change toward beginning regular physical activity.

Parameter		Estimate	95% CI	<i>p</i>
Threshold	Neg.—None	-2.562	[-18.053, -1.737]	<0.001***
	None—Pos.	-1.594	[-16.527, -0.397]	0.009**
Condition	Dynamic	-0.911	[-16.223, 1.414]	0.342
	Exaggerated	-2.078	[-19.080, 14.316]	0.258

Table 8.16: Ordinal logistic regression model to the outcome of showing negative, positive, or no progress on the Stages of Changes toward beginning regular physical activity. Confidence intervals are estimated from a nonparametric bootstrap (2000 replicates). * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

8.6.10 Attitudes toward Physical Activity

I examined changes in attitudes toward physical activity in terms of progress on the Stages of Change from the assessment at intake to the final assessment following the last conversation (for those participants who completed the last conversation). Progress was taken as an ordinal outcome: positive if a participant moved from an earlier to a later stage (e.g., from contemplation to preparation), negative if a participant moved from later to earlier, and neutral if there was no change.

Figure 8.15 shows Stage of Change progress, by study condition. Overall,

		Final Stage					
		PC	C	P	A	M	None
Initial Stage	PC	1	0	1	1	0	2
	C	1	1	2	3	2	13
	P	0	3	2	6	4	20

Table 8.17: Counts of participants reporting each Stage of Change at intake and at the final assessment. PC=precontemplation, C=contemplation, P=preparation, A=action, M=maintenance, and None=no final assessment completed.

70.4% (19 of 27) of participants who remained in the study and gave the final assessment reported advancing at least one stage of change, 14.8% (4 of 27) reported no change, and 14.8% reported falling back to an earlier stage. Participants in the **Exaggerated** condition were more likely to report negative progress, however only a small fraction of participants in that condition ($n = 4$) remained in the study to give the final assessment.

Table 8.17 gives the number of participants (including all study conditions) reporting progress to each of the Stages of Change at the time of final assessment, contingent on the initial Stage of Change assessment. Of the subset of participants who completed the final assessment, the majority of participants initially in contemplation (5 of 9) or preparation (10 of 15) reached either the action or maintenance stages.

To test the effect of study condition on Stage of Change progress, I fit an ordinal logistic regression, with study condition as the predictor (Table 8.16). No significant effect of study condition was found.

8.6.11 Self-Reported Walking

In a subset of sessions, participants negotiated walking goals: either in the first session, for participants who indicated they were in the preparation, action, or maintenance Stages of Change, or in subsequent sessions, following a statement of increased motivation and confidence. Walking goals were nego-

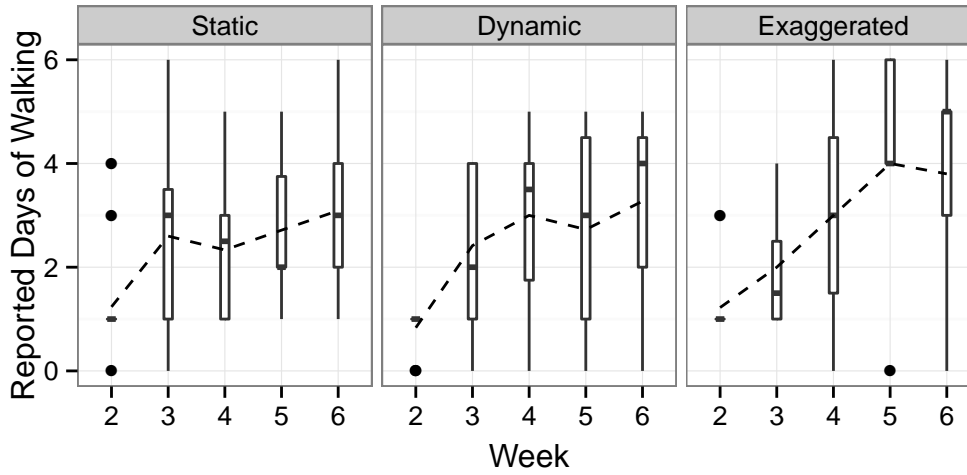


Figure 8.16: Self-reported days of planned brisk walking per week, aggregated by study condition. The dotted line indicates group means.

tiated in terms of a number of days per week to engage in a scheduled bout of brisk walking, without specifying the days, time, or place, of walking. For each negotiated goal, the ECA asked participants, in the following session, to self-report the number of days actually walked in the previous week. Note that self-reported days of walking were never available before the second week of the study.

Figure 8.16 shows the number of days of walking reported per week, aggregated by study condition. Participants reported walking on average 2.42 days per week (SD 1.70). The average number of days reported increased from the second week (mean 1.11, SD 0.73) to the sixth (mean 3.30, SD 1.79). Participants in the **Exaggerated** condition reported the highest number of days walked (mean 2.63, SD 2.04), followed by the **Dynamic** (mean 2.43, SD 1.74) and **Static** (mean 2.32, SD 1.52) conditions.

To test for systematic changes in reported walking over time and between study conditions, I fit a random-intercept mixed-effect linear regression model, including fixed effects of study condition and study week, and interaction terms between them (Table 8.18). Participants reported significantly more days of walking in later weeks ($\beta = 0.443$, 95% CI [0.232, 0.648], $p < 0.001$).

Parameter		Estimate	95% CI	<i>p</i>
Fixed Effects	Intercept	1.557	[0.877, 2.231]	<0.001***
	Dynamic	-0.261	[-1.243, 0.767]	0.626
	Exaggerated	-0.279	[-1.450, 0.899]	0.632
	Week	0.443	[0.232, 0.648]	<0.001***
	Week · Dynamic	0.122	[-0.185, 0.429]	0.450
	Week · Exaggerated	0.264	[-0.105, 0.639]	0.170
Random Intercept (SD)		1.045	[0.696, 1.332]	
Residual (SD)		1.192	[1.021, 1.347]	

Table 8.18: Random-intercept regression model fit to a linear trend on self-reported days of walking per week. Confidence intervals are estimated from a residual-resampling bootstrap (5000 replicates). * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

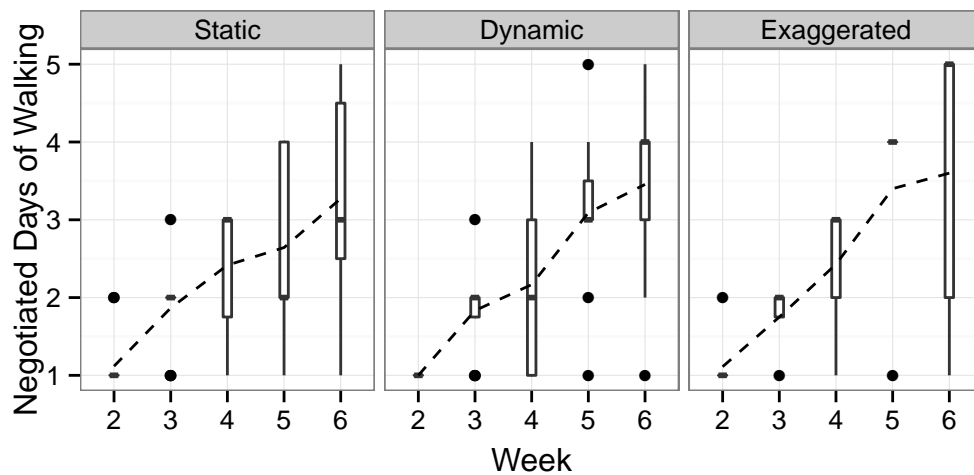


Figure 8.17: Negotiated goals for days of walking per week, aggregated by study condition. The dotted line indicates group means.

Taking the **Static** condition as the reference category, the estimated rate of increase was slightly higher in both the **Dynamic** and **Exaggerated** conditions, but the difference was not significant in either case.

Figure 8.17 shows the number of days per week of walking negotiated as a goal between participants and the ECA, aggregated by study condition. Overall, participants agreed to an average of 2.23 days of walking per week (SD 1.22). Higher goals were negotiated in later weeks, increasing from a

Parameter		Estimate	95% CI	<i>p</i>
Fixed Effects	Intercept	1.213	[0.814, 1.590]	< 0.001 ***
	Dynamic	-0.239	[-0.838, 0.358]	0.416
	Exaggerated	-0.127	[-0.802, 0.545]	0.830
	Week	0.518	[0.409, 0.632]	< 0.001 ***
	Week · Dynamic	0.121	[-0.042, 0.284]	0.151
	Week · Exaggerated	0.139	[-0.067, 0.340]	0.174
Random Intercept (SD)		0.640	[0.442, 0.799]	
Residual (SD)		0.647	[0.558, 0.732]	

Table 8.19: Random-intercept regression model fit to a linear trend on negotiated goals for days of walking per week. Confidence intervals are estimated from a residual-resampling bootstrap (5000 replicates). * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

mean of 1.08 days (SD 0.27) in the first week to 3.41 days (SD 1.42) in the last week. Participants in the **Exaggerated** condition negotiated the highest goals (mean 2.30, SD 1.39) followed by the **Dynamic** (mean 2.28, SD 1.23) and **Static** (2.16, SD 1.16) conditions.

To test for systematic changes in negotiated walking goals over time and across study conditions, I fit a mixed-effect linear regression model, as above (Table 8.19). Participants negotiated significantly higher goals in later weeks ($\beta = 0.518$, 95% CI [0.409, 0.632], $p < 0.001$). As above, taking the **Static** condition as the reference category, the estimated rate of change was higher in the **Dynamic** and **Exaggerated** conditions, but the difference was not significant in either case.

Figure 8.18 shows the number of days above or below the negotiated goal which participants reported walking per week, aggregated by study condition. Overall, participants reported walking approximately the same number of days as their goal (mean 0.19, SD 1.33). There was no consistent pattern of change over time. Participants in the **Exaggerated** condition reported walking on more days relative to their goals (mean 0.33, SD 1.03) than either the **Static** (mean 0.16, SD 1.09) or **Dynamic** (mean 0.16, SD 1.24) conditions.

To test for systematic changes in reported days of walking per week rela-

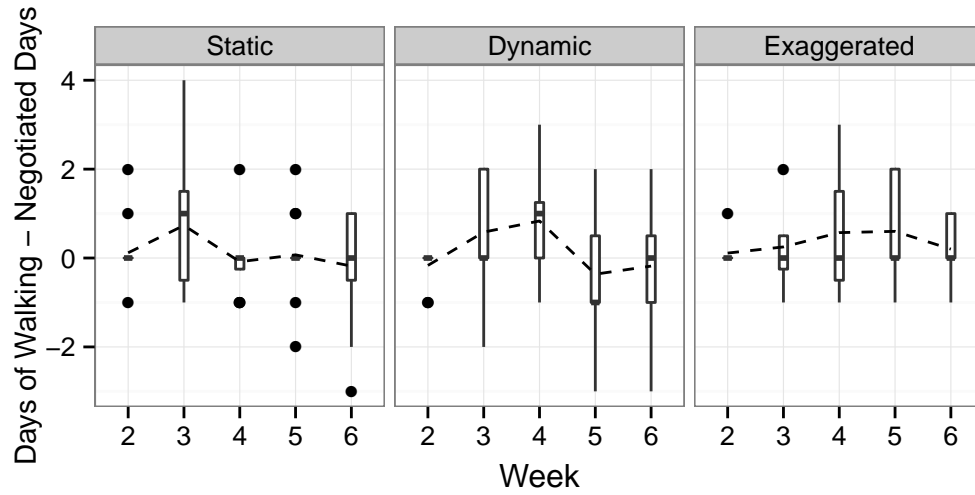


Figure 8.18: The difference between self-reported days of walking and negotiated goals for days of walking per week, aggregated by study condition. The dotted line indicates group means.

Parameter		Estimate	95% CI	<i>p</i>
Fixed Effects	Intercept	0.361	[-0.060, 0.826]	0.090
	Dynamic	-0.033	[-0.714, 0.635]	0.913
	Exaggerated	0.157	[-0.975, 0.633]	0.679
	Week	-0.104	[-0.311, 0.080]	0.227
	Week · Dynamic	0.018	[-0.273, 0.297]	0.866
	Week · Exaggerated	0.167	[-0.154, 0.531]	0.289
Random Intercept (SD)		0.359	[0.00, 0.421]	
Residual (SD)		1.087	[0.968, 1.279]	

Table 8.20: Random-intercept regression model fit to a linear trend on the difference between self-reported days of walking and negotiated walking goals per week. Confidence intervals are estimated from a residual-resampling bootstrap (5000 replicates). * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

tive to negotiated goals, I fit a mixed-effect linear regression model, as above (Table 8.20). Participants tended, in a near-significant trend, to report walking more days than their negotiated goals ($\beta = 0.361$, 95% CI $[-0.060, 0.826]$, $p = 0.090$). No significant change over time was found, and no significant differences between conditions were found.

8.7 Discussion

Of the hypotheses that examined the effect of the agent’s behavior on users’ engagement and interpersonal bond, I find mixed results, but partial support for the overall thesis that realistic long-term conversational behavior (or behavior intended to be realistic) will have positive effects on the user-agent interpersonal relationship. Hypothesis **H1** was not supported: I found no significant difference in the frequency of voluntary system usage between the **Static** and **Dynamic** conditions. However, hypothesis **H2** was supported: participants in the **Dynamic** condition reported significantly greater bond with the agent in later weeks.

I did not find evidence that the changes in the agent’s behavior in the **Dynamic** or **Exaggerated** condition are perceptible to participants and can be reliably reported. Hypothesis **H3** was not supported: no significant difference was found across conditions in ratings of nonverbal immediacy. Note that in contrast to the behavioral realism measure discussed below (**H5**), which asked participants to judge the prevalence of various behaviors relative to the expected prevalence of those behaviors when interacting with a human conversation partner, this measure asked participants to judge the prevalence of behaviors without reference to an external standard.

Of the hypotheses that examine the effect of the agent’s behavior on the perceived behavioral realism of the agent, I find mixed results, but again partial support for the overall thesis that realistic long-term behavior will cause the agent to be perceived as more human-like and realistic. Hypothesis

H4 was not supported: no significant difference was found across conditions in reports of overall or gestalt behavioral realism. However, hypothesis **H5** was partially supported: participants in the **Exaggerated** condition were significantly less likely to report that specific behaviors were realistic.

I report no effect of the agent’s behavior on attitudes toward physical activity. Hypothesis **H6** was not supported: there was no significant difference across conditions in progress along the Stages of Changes toward uptake of regular physical activity. Similarly, I report no effect of the agent’s behavior on self-reported walking behavior: Hypothesis **H7** was not supported.

8.7.1 Agent Behavior and User Engagement

As noted above, I find partial support for the hypotheses that the (assumed) realistic agent behavior in the **Dynamic** condition will promote increased user engagement and a stronger user-agent interpersonal bond: there is a significant difference in the therapeutic alliance bond construct, but not in system usage (taken as a measure of engagement).

The therapeutic alliance construct (and the bond subscale of it) is known to be meaningful and predictive of outcomes in counseling [104], and is associated with other constructs relevant to interpersonal relationship (e.g., adult attachment style [51, 52]). I argue that significant improvements over time in this construct is encouraging evidence that the conversational behavior in the **Dynamic** condition is producing meaningful improvements in the user-agent interpersonal relationship. Of the three study conditions, only the **Dynamic** condition produced a pattern of increasing interpersonal bond over time similar to that observed in the corpus (Section 3.4.2).

The difference between these two findings — a significant difference over time in self-reported trust and bond, with a lack of significant difference in observed system usage — merits some discussion. I first note that a lack of significant difference is not a lack of a difference, and it is possible that a study

with more participants or otherwise greater statistical power would produce a difference: indeed, there is a trend for participants in the **Dynamic** condition to have relatively higher system usage over time (Table 8.4), although it does not approach significance ($p = 0.138$). The choice of whether or not to have a conversation at all is, while certainly a meaningful measure, also a rather low-precision measure of engagement, and may only be capable of differentiating a very weak user-agent interpersonal relationship from others.

The increased interpersonal bond in the **Dynamic** condition does not occur at baseline (i.e., immediately in the first session), but rather is a significant difference in the *rate* of change in interpersonal bond relative to the **Static** condition. The difference in participant's reports of interpersonal bond with the agent between the two conditions are larger in later weeks. This is reasonable, since the changes in behavior in the **Dynamic** condition occur gradually across several conversation, and leads to a conjecture: if the systematic changes in behavior introduced in the **Dynamic** condition are able to produce a measurable difference in system usage, it may require more than six interactions (and likely a longer duration of interaction than six weeks) before this difference is apparent. Future studies in this area should consider including longer-term interaction; and indeed, subtle behavior changes have been able to produce measurable differences in voluntary system usage with a similar counseling agent, when measured over much longer periods of interaction [15].

8.7.2 Exaggerated Behavior and Realism

The **Exaggerated** condition, in which the agent demonstrated the same behavior changes as in the **Dynamic** condition but with a greatly increased rate of change, did not produce improved reports of interpersonal bond. Participants in the **Exaggerated** condition also had significantly less actual usage of the system than other conditions. I argue that this is evidence that more extreme change in behavior *can* have a measurable effects on participation in

long-term interactions with an agent, or on participation in long-term behavior change, even within a relatively short period of six weeks.

The effect of the **Exaggerated** condition is toward less engagement and system usage: broadly speaking (as they are significant on different outcomes) this is in the opposite direction of the **Dynamic** condition, although the same systematic behavior changes were used (with different rates of change). Moreover, the negative effect on system usage occurs quickly: participants in the **Exaggerated** condition show a lower rate of completed conversations as early as the second week of the study. The changes in the agent's behavior between consecutive conversations are quite pronounced in this study condition. The change, for example, of the rate of the agent's gaze-aways between the first and second conversations in the **Exaggerated** condition approximately doubles (equivalent to the change between the first and fourth conversations in the **Dynamic** condition). I conjecture that, while realistic changes in verbal and nonverbal behavior across conversations may have positive effects on the user-agent interaction, the **Exaggerated** condition begins to identify the points at which such changes become too extreme, may be perceived as unrealistic, and can have negative effects.

Supporting this conjecture, I find that participants in the **Exaggerated** condition reported decreasing perceived behavioral realism of the agent over time, relative to other conditions (Section 8.6.9); that is, participants in this condition were less likely to say, in later weeks, that the agent approximately matched the participant's expectations of a human speaker's conversational behavior.

8.7.3 Engagement and Behavior Change

Finally, I note that while there was no significant effect of study condition on participants' Stage of Change toward regular physical activity, the likelihood of a participant reporting an advance of at least one stage of change was

high; estimated at 83.1% (95% CI [59.8%, 99.9%]) for the **Static** condition and 66.4% (95% CI [38.6%, 91.7%]) for the **Dynamic** condition. Similarly, while there was no significant effect of study condition on self-reported walking behavior, participants' in all three groups significantly increased their walking behavior during the study (assuming the accuracy of their self-reporting did not change over time).

Since only participants who completed the final weekly session and the final assessment questionnaires are included in these proportions, this gives some evidence that maintaining user engagement and participation in an agent-based behavioral intervention can result in behavior change for many users — although not strong evidence, as I have not compared with the subgroup of participants who did *not* participate in the study through the final session. Combined with the findings that realistic agent behavior can improve user-agent interpersonal bond, and that exaggerated behavior can reduce engagement, this suggests that the use of realistic behavior may be a promising approach to maintaining long-term engagement and promoting behavior change when interacting with ECAs.

CHAPTER 9

Conclusions

I now return to the main thesis statement of this dissertation, and reevaluate it in the context of the research contributions presented.

Thesis Statement When humans interact in multi-conversation discourse, their verbal and nonverbal behavior changes in systematic ways, both over time and in the context of changes to their interpersonal relationship. Modeling these behaviors and implementing them within an Embodied Conversational Agent designed for long-term interaction will increase user engagement with the agent, and will benefit the user’s perceptions of the agent as realistic and human-like.

This statement makes three major claims, each of which has been addressed in previous chapters.

First, I have demonstrated evidence of systematic changes in verbal and nonverbal conversational behavior in human-human multi-conversation discourse, through the collection and analysis of the Exercise Counseling Corpus (Chapter 3). Annotation and statistical modeling of this corpus revealed changes in the prevalence of multiple behaviors, predicted both by the cumulative amount of interaction (Chapters 4, 5, and 6), and by the strength of the participants’ interpersonal relationship (Chapter 6). While these results are limited by the size and scope of the corpus, and require additional validation

to be generalized, they establish a basic set of plausible long-term behavior changes which can be tested and further validated.

Second, I have shown that computational models of long-term conversational behavior can be created, and implemented within an Embodied Conversational Agent (ECA). I present a proof-of-concept implementation (Chapter 7), and describe a set of assumptions sufficient to give a computational model of conversational behavior generation based on the observational results which is fully-specified and implementable.

Third, I present an empirical test of the statement that implementing these behaviors within an ECA can benefit long-term user-agent engagement and interpersonal relationship, and improve the perceived realism of an agent. I conducted a longitudinal randomized controlled trial (Chapter 8) comparing agents with behavior that changed according to my findings, to agents with otherwise similar behavior that did not change, to agents with an exaggerated form of my findings that displayed rapid and abrupt changes between conversations. I show mixed, but encouraging results: there is evidence that dynamically changing agent behavior can cause positive changes in meaningful measures of interpersonal bond, and that changes in agent behavior are capable of a significant impact on user engagement, in the context of an agent that produces significant changes in self-reported user health behaviors.

9.1 Research Contributions

This work is interdisciplinary, and makes research contributions across several related fields, including computer science (specifically, human-computer interaction), health psychology, social psychology, and health informatics. I list contributions roughly in the order of presentation.

9.1.1 A Methodology for the Study of Conversational Behavior in Long-Term Interaction

I develop a methodology for producing realistic verbal and nonverbal behavior in an ECA that engages in long-term interaction with users. I extend an approach used extensively by prior work (e.g. [114, 41, 42, 163]): collecting a corpus of human-human interactions that is an example of the type of interaction intended for an ECA, annotating it for behaviors of interest, building statistical models of the occurrence and features of behaviors based on those annotations, and finally constructing heuristic rules based on those models, which can then be used as part of a behavior generation system for an ECA. The prior approach was typically based on a context of single isolated conversations (both in terms of the corpus of human-human interaction and in terms of the intended ECA interaction); I adapt this methodology to one appropriate for multi-conversation discourse and long-term human-ECA interaction, by collecting a longitudinal corpus of human-human interaction (containing multiple conversations per dyad), annotating it for behaviors of interest, and applying appropriate statistical models for the resulting longitudinal data. From this, I construct heuristic rules for behavior generation that include predictors which vary over multiple conversations, including interaction history and measures of interpersonal relationship.

As part of the the overall methodology, I develop a methodology and a set of requirements for the collection of a corpus suitable for the study of long-term interaction. The Exercise Counseling Corpus can serve as an example of such a corpus, and can motivate the design and methodology of corpus-collection efforts for future work involving long-term interaction. The relevant features of the corpus include: 1. a longitudinal design with multiple conversations per dyad; 2. a meaningful task — behavior change counseling — that both requires multiple conversations and in which the strength of the dyads' interpersonal relationship is meaningful; and 3. longitudinal, conversation by conversation

assessment, from both participants of appropriate constructs expected to vary over time and to be predictive of behavior — here, the therapeutic alliance construct.

9.1.2 Patterns of Conversational Behavior in Long-Term Interaction

I present a set of empirical results illustrating changes in verbal and nonverbal behavior across multiple conversations between the same dyad. Some of these results follow and are compatible with previous theoretical and empirical work (e.g., there is a decrease in the prevalence of smiles in later conversations, which is compatible with prior work arguing that the importance of behaviors associated with positivity will decrease over time [161, 162]), while others may refine previous work or suggest modification. All results are explicitly longitudinal, giving a predicted time course of changes, where most previous work only makes predictions about differences in behavior before and after a major change in interpersonal relationship (e.g., between strangers and friends). I identify systematic patterns of conversational behavior which are not well-described in previous work, most notably an abrupt change in multiple nonverbal behaviors in the final conversation between a dyad in which behaviors much more strongly resemble a first conversation than the immediately preceding conversation. All such results are based on a small corpus, limited to a single conversational task and set of conversational roles (a counselor and a client), and require future work to provide further validation and to test whether these results generalize to other types of long-term interactions. However, these results represent a substantial contribution to the literature and should provide direction for future investigations, both for research in realistic behavior for ECAs, and for research primarily focused on human-human interaction.

9.1.3 An ECA with Realistic Behavior in Long-Term

Interaction

I show, via Rhythm, a proof-of-concept implementation, how these observed patterns of conversational behavior may be incorporated into a system that generates verbal and nonverbal behavior for an ECA. The resulting system has been integrated into a functioning ECA, which has interacted with users through multiple conversations and displayed changing behavior.

I present preliminary validation of my empirical findings, and of the implementation of those findings, via a longitudinal randomized controlled evaluation study. The evaluation study provides multiple research contributions. It provides some partial validation of the empirical results found in the Exercise Counseling Corpus, beyond the small number of participants in the corpus. It provides validation that the implementation of those results in Rhythm and in the ECA captures at least some aspects (through the various assumptions and compromises required for implementation) of the human behavior observed. It provides evidence that these behavior patterns have a measurable effect on user engagement, user-agent interpersonal relationship, and on the perceived behavioral realism of the agent. Finally, being an experimental study, in which the agent's behavior is manipulated as an independent variable, the evaluation study provides some causal information: I show that changes in the agent's behavior cause changes in the user-agent interpersonal relationship, whereas the Exercise Counseling Corpus (being observational) can only show associations.

9.2 Future Research: Toward Rich Models of Behavior in Long-Term Interaction

I have presented a model of verbal and nonverbal behavior across multiple conversations in which behavior is predicted primarily by two aspects of the

interaction and of the participants in that interaction: interaction history (limited to the number of previous conversations), and the strength of the interpersonal bond. Long-term interaction is complex, and a richer model of verbal and nonverbal behavior could potentially include a large number of additional predictors. In this section, I sketch some future research questions in that direction.

Most simply, I am interested in extending these results to very long-term interaction: I observe roughly linear trends in changes in verbal and nonverbal behavior over a period of six weeks. However, many real-world scenarios of interest could easily involve interaction for much longer periods: months, years, or indefinitely. It is not plausible that the trends identified here continue linearly over very long-term interactions, as they would produce behavior well outside of conversational norms (either for an ECA or for a human interactant). I conjecture that trends level off over time: for example, that the 99th and 100th conversation between a dyad look far more similar than the first and second. Future research is required to confirm this prediction and give a more detailed time course of these changes. In very long-term interaction, it may also become increasingly less likely that a dyad follows a smooth pattern of behavior over time without disruption (whether external or internal). The role of ruptures in interpersonal relationship, or other significant deviations from systematic patterns of behavior change, should be studied.

In the Exercise Counseling Corpus, participants interacted once per week for six weeks, and the evaluation study was based on a similar scenario. In both cases, I operationalized the concept of interaction history as simply the number of previous conversations a dyad has participated in. Interaction history may be broadened to consider not only the number of previous conversations, but also the pattern of those conversations: for example, the regularity of interaction, and the delay in time between conversations. Considering varying patterns of interaction, in which participants may have conversations at different intervals (e.g., once per day, or several per day, or rarely), and may have

conversations at irregular intervals, leads to a group of new research questions.

First, does the duration of time spent between interactions predict any differences in behavior? I conjecture that participants who have a long duration between conversations will display a slower rate of change in their verbal and nonverbal behaviors across conversations. For example, the difference between a first and second conversation may be smaller for participants who have conversations one month apart than those who have conversations one week apart.

Second, as a refinement of the previous question, must conversations be separated by some period of time in order to show the observed effects? For example, if we predict a change in behavior after five weekly conversations, is it reasonable to predict the same change in behavior after five *daily* conversations, or even five conversations within the same day?

Third, if participants do *not* interact for an extended period of time, is there a “decay” in which their behavior begins to more closely resemble a first conversation (or an early conversation)? I conjecture that such a decay would be observed, given a sufficiently large delay between interactions. The observational results provide some support for this conjecture, by establishing that the changes in behavior that occur across conversations are not monotonic: they can and do reverse direction, as seen in the final conversations in the Exercise Counseling Corpus (Chapter 6).

The Exercise Counseling Corpus is an example of a corpus in which all conversations are intended to accomplish a portion of the same conversational task (behavior change counseling). It is an open research question whether a corpus built around a different conversational task — for example, education or tutoring, or a task with no clear “goal,” such as long-term companionship — would produce similar patterns of behavior. Other corpora may include or emphasize the use different relational behaviors, and would encourage the study of associated verbal and nonverbal behaviors.

In some interaction scenarios, there may be more variety in the types of

conversational tasks addressed by a single dyad. In the Exercise Counseling Corpus, there is a strong coupling between task familiarity and interaction history; a dyad who has had more previous conversations is also more familiar with the task of behavior change counseling. A different corpus, built around a different task, might partially decouple these two aspects of long-term interaction. For example, consider an educational agent designed for long-term interaction, in which occasionally a student completes one portion of a curriculum and begins a new task (with which he or she has much less familiarity). I raise a research question: what is the conversational behavior of a dyad with a long interaction history, but low task familiarity?

A related set of questions concerns the roles of participants in a conversation. In the Exercise Counseling Corpus, the roles of participants are asymmetric: one counselor and one client. Motivated by the small size of the corpus — which is, after all, limited to a single counselor — I have not addressed these questions, but it is highly plausible that varying conversational roles are reflected in verbal and nonverbal behavior. For example, I conjecture that there is a perceived power differential between the counselor and the client; politeness theory predicts that a speaker will use higher-politeness statements when interacting with a conversation partner they perceive to have power over them [23], and such politeness differences may be reflected in the details of verbal and nonverbal behavior [164]. A corpus including examples of dyads both with and without a similar power differential would allow the study of this potential research question.

A number of participant traits should be considered as predictors of conversational behavior; both traits of a speaker and the interaction between speaker and listener traits (e.g., concordance) are possible predictors of behavior. These traits include the frequently-studied Big 5 personality factors, but also traits such as attachment style[21], which is stable over time, strongly influence interpersonal relationship constructs including therapeutic alliance [52, 51], and may be associated with differences in nonverbal behavior in con-

versation [98, 165]. There are effects of gender and gender concordance, which may interact with various dimensions of interpersonal relationship to influence verbal and nonverbal behavior in conversation [71, 105]. Future work, based on a larger corpus with additional variation along any or all of these dimensions, may explore these differences.

9.3 Future Research: Toward Rich Models of Conversational Behavior in Health Behavior Change

The interactions studied here, and the intended applications of the ECA developed here, are within the context of an intervention for health behavior change. Health behavior change is itself a long-term process. The Transtheoretical Model of health behavior change [133] predicts that individuals attempting to change a health behavior will pass through several distinct Stages of Change (precontemplation, contemplation, preparation, action, and maintenance), and that different sets of Processes of Change (e.g., consciousness raising, dramatic relief, and several others) will be applied in different stages. Within a series of health behavior change counseling conversations, as studied here, the stage and processes represent another set of dimensions along which different conversations may vary over time.

There are at least two sets of future research questions raised by this. First, are there differences in verbal and nonverbal behavior in conversation predicted by changes in a client's stage of change, or other constructs in different theories of health behavior change, over time? Second, to what extent is conversational behavior predictive of changes in such constructs?

9.4 Future Research: Toward an Integrated Model of Verbal and Nonverbal Behavior

The efforts to model, predict, and reproduce verbal and nonverbal behavior which I have presented here have focused largely on the context of an utterance at the level of an entire conversation: I examine the state of participants in a conversation, and the overall position of an utterance within a conversation (e.g., near the beginning of a conversation versus the middle or end). By contrast, much prior work on nonverbal behavior generation for ECAs begins first with the observation that much nonverbal behavior (particularly hand gestures) is co-articulated with speech [108, 86] and predicts nonverbal behavior based on the content of the co-articulated speech, or the immediate context associated with that utterance, on a much smaller scale. Several examples have been referenced in earlier chapters: for example, an association between posture shifts and discourse segment boundaries [41], and associations between eye gaze and both turn-taking (i.e., whether a speaker yields the turn to the conversation partner at the end of an utterance) and information structure [42].

A future research problem is to produce accounts of verbal and nonverbal behavior that include both the immediate and conversation-scale context of an utterance. The model of posture shifts presented in Chapter 5 already includes some work in this direction, as the presence or absence of a topic shift — the discourse structure of a conversation at the immediate point where an utterance occurs — is included as a predictor. However, interactions between these different types of context have not yet been fully explored.

9.5 Closing

This thesis has presented a demonstration of the complex ways in which people behave when they interact and converse over long periods of time, an example

of reproducing those patterns of behavior within a computer interface, and a demonstration of the efficacy of such an interface for real-world, socially and practically meaningful, applications. The set of potential applications for an intelligent, engaging computer interface that can actively promote long-term engagement and voluntary use is broad, spanning counseling, education, and numerous fields within health care, and it is my hope that these results open new approaches toward these applications.

Bibliography

- [1] Andrea Abele. Functions of gaze in social interaction: Communication and monitoring. *Journal of Nonverbal Behavior*, 10(2):83–101, June 1986.
- [2] Jens Allwood, Loredana Cerrato, Kristiina Jokinen, Costanza Navarretta, and Patrizia Paggio. The MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena. *Language Resources and Evaluation*, 41(3):273–287, December 2007.
- [3] Irwin Altman and D. Taylor. *Social penetration: The development of interpersonal relationships*. Holt, Rinehart and Winston, 1973.
- [4] Nalini Ambady, Jasook Koo, Fiona Lee, and Robert Rosenthal. More than words: Linguistic and nonlinguistic politeness in two cultures. *Journal of Personality and Social Psychology*, 70(5):996–1011, 1996.
- [5] Peter A. Andersen. Nonverbal immediacy in interpersonal communication. In Aron W. Siegman and Stanley Feldstein, editors, *Multichannel Integrations of Nonverbal Behavior*, pages 1–36. Lawrence Erlbaum, Hillsdale, NJ, 1985.
- [6] Michael Argyle. *Bodily communication*. Routledge, second edition, 1988.
- [7] R. S. Aylett, S. Louchart, J. Dias, A. Paiva, and M. Vala. FearNot! - an experiment in emergent narrative. In Themis Panayiotopoulos, Jonathan Gratch, Ruth Aylett, Daniel Ballin, Patrick Olivier, and

- Thomas Rist, editors, *Intelligent Virtual Agents*, volume 3661, chapter 26, pages 305–316. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- [8] Jeremy N. Bailenson, Jim Blascovich, Andrew C. Beall, and Jack M. Loomis. Interpersonal distance in immersive virtual environments. *Pers Soc Psychol Bull*, 29(7):819–833, July 2003.
- [9] Roger Bakeman, Duncan McArthur, Vincenç Quera, and Byron F. Robinson. Detecting sequential patterns and determining their reliability with fallible observers. *Psychological Methods*, 2(4):357–370, 1997.
- [10] Douglas Bates, Martin Maechler, and Ben Bolker. *lme4: Linear mixed-effects models using S4 classes*, 2012. R package version 0.999999-0.
- [11] Amy L. Baylor and Yanghee Kim. Pedagogical agent design: The impact of agent realism, gender, ethnicity, and instructional role. In *Intelligent Tutoring Systems*, pages 268–270, 2004.
- [12] Frank J. Bernieri. Coordinated movement and rapport in teacher-student interactions. *Journal of Nonverbal Behavior*, 12(2):120–138, June 1988.
- [13] Timothy Bickmore. *Relational Agents: Effecting Change through Human-Computer Relationships*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 2003.
- [14] Timothy Bickmore, Daniel Schulman, and George Shaw. DTask and LiteBody: Open source, Standards-Based tools for building Web-Deployed embodied conversational agents. In Zsófia Ruttkay, Michael Kipp, Anton Nijholt, and Hannes Vilhjálmsón, editors, *Intelligent Virtual Agents*, volume 5773 of *Lecture Notes in Computer Science*, chapter 46, pages 425–431. Springer Berlin / Heidelberg, Berlin, Heidelberg, 2009.

- [15] Timothy Bickmore, Daniel Schulman, and Langxuan Yin. Maintaining engagement in long-term interventions with relational agents. *Applied Artificial Intelligence*, 24(6):648–666, July 2010.
- [16] Timothy W. Bickmore, Laura Pfeifer, Daniel Schulman, Sepalika Perera, Chaamari Senanayake, and Ishraque Nazmi. Public displays of affect: deploying relational agents in public spaces. In *CHI '08: CHI '08 extended abstracts on Human factors in computing systems*, pages 3297–3302, New York, NY, USA, 2008. ACM.
- [17] Timothy W. Bickmore, Daniel Schulman, and Candace L. Sidner. A reusable framework for health counseling dialogue systems based on a behavioral medicine ontology. *Journal of Biomedical Informatics*, January 2011.
- [18] J. M. Bland and D. G. Altman. Correlation, regression, and repeated data. *BMJ*, 308(6933), April 1994.
- [19] Edward S. Bordin. The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research, and Practice*, 16(3):252–260, 1979.
- [20] Edward S. Bordin. Theory and research on the therapeutic working alliance: New directions. In Adam O. Horvath and Leslie S. Greenberg, editors, *The Working Alliance: Theory, Research, and Practice*, chapter 1, pages 13–37. Wiley, New York, NY, 1994.
- [21] Kelly A. Brennan, Catherine L. Clark, and Phillip R. Shaver. Self-report measurement of adult romantic attachment: An integrative overview. In Jeffry A. Simpson and William S. Rholes, editors, *Attachment theory and close relationships*, chapter 3, pages 46–76. Guilford Press, New York, 1998.

-
- [22] N. E. Breslow. Extra-Poisson variation in Log-Linear models. *Applied Statistics*, 33(1):38+, 1984.
- [23] Penelope Brown and Stephen C. Levinson. *Politeness : Some Universals in Language Usage (Studies in Interactional Sociolinguistics)*. Cambridge University Press, February 1987.
- [24] Judee K. Burgoon, David B. Buller, Jerold L. Hale, and Mark A. de Turck. Relational messages associated with nonverbal behaviors. *Human Communication Research*, 10(3):351–378, 1984.
- [25] Judee K. Burgoon, David B. Buller, and William G. Woodall. *Nonverbal communication : the unspoken dialogue*. Harper & Row, 1989.
- [26] Judee K. Burgoon and Jerold L. Hale. The fundamental topoi of relational communication. *Communication Monographs*, 51(3):193–214, 1984.
- [27] Judee K. Burgoon and Jerold L. Hale. Validation and measurement of the fundamental themes of relational communication. *Communication Monographs*, 54(1):19–41, 1987.
- [28] Judee K. Burgoon and Beth A. Poire. Nonverbal cues and interpersonal judgments: Participant and observer perceptions of intimacy, dominance, composure, and formality. *Communication Monographs*, 66(2):105–124, 1999.
- [29] Hendrik Buschmeier, Kirsten Bergmann, and Stefan Kopp. An alignment-capable microplanner for natural language generation. In *ENLG '09: Proceedings of the 12th European Workshop on Natural Language Generation*, pages 82–89, Morristown, NJ, USA, 2009. Association for Computational Linguistics.

- [30] J. T. Cacioppo, W. von Hippel, and J. M. Ernst. Mapping cognitive structures and processes through verbal content: the thought-listing technique. *J Consult Clin Psychol*, 65(6):928–940, December 1997.
- [31] John T. Cacioppo and Richard E. Petty. Social psychological procedures for cognitive response assessment: The thought-listing technique. In Thomas V. Merluzzi, Carol R. Glass, and Myles Genest, editors, *Cognitive Assessment*, chapter 10, pages 309–342. Guilford Press, New York, 1981.
- [32] Joana Campos and Ana Paiva. MAY: My memories are yours. In Jan Allbeck, Norman Badler, Timothy Bickmore, Catherine Pelachaud, and Alla Safonova, editors, *Intelligent Virtual Agents*, volume 6356 of *Lecture Notes in Computer Science*, chapter 44, pages 406–412–412. Springer Berlin / Heidelberg, Berlin, Heidelberg, 2010.
- [33] Joseph N. Cappella. Dynamic coordination of vocal and kinesic behavior in dyadic interaction: Methods, problems, and interpersonal outcomes. In James H. Watt and C. Arthur VanLear, editors, *Dynamic Patterns in Communication Processes*, pages 353–386. Sage, Thousand Oaks, CA, 1996.
- [34] Joseph N. Cappella and Darren M. Schreiber. The interaction management function of nonverbal cues: Theory and research about mutual behavioral influence in Face-to-Face settings. In Valerie Manusov and Miles L. Patterson, editors, *The SAGE Handbook of Nonverbal Communication*, chapter 19, pages 361–379. Sage Publications, Thousand Oaks, CA, 2006.
- [35] Jean Carletta. Unleashing the killer corpus: experiences in creating the multi-everything AMI meeting corpus. *Language Resources and Evaluation*, 41(2):181–190, May 2007.

- [36] James R. Carpenter, Harvey Goldstein, and Jon Rasbash. A novel bootstrap procedure for assessing the relationship between class size and achievement. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(4):431–443, 2003.
- [37] Justine Cassell. *Embodied conversational agents*. MIT Press, April 2000.
- [38] Justine Cassell. Nudge nudge wink wink: Elements of Face-to-Face conversation for embodied conversational agents. In Justine Cassell, Joseph Sullivan, Scott Prevost, and Elizabeth Churchill, editors, *Embodied Conversational Agents*, pages 1–27. MIT Press, Cambridge, 2000.
- [39] Justine Cassell and Timothy Bickmore. Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User Modeling and User-Adapted Interaction*, 13(1):89–132, February 2003.
- [40] Justine Cassell, Alastair J. Gill, and Paul A. Tepper. Coordination in conversation and rapport. In *Workshop on Embodied Language Processing*, pages 41–50. Association for Computational Linguistics, June 2007.
- [41] Justine Cassell, Yukiko I. Nakano, Timothy W. Bickmore, Candace L. Sidner, and Charles Rich. Non-verbal cues for discourse structure. In *ACL '01: Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, pages 114–123, Morristown, NJ, USA, 2001. Association for Computational Linguistics.
- [42] Justine Cassell, Obed E. Torres, and Scott Prevost. Turn taking vs. discourse structure: How best to model multimodal conversation. In Yorick Wilks, editor, *Machine Conversations*, pages 143–154. Kluwer Academic Publishers, Norwell, MA, USA, 1999.
- [43] Justine Cassell, Hannes H. Vilhjálmsón, and Timothy Bickmore. BEAT: the behavior expression animation toolkit. In *Proceedings of*

- the 28th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '01, pages 477–486, New York, NY, USA, 2001. ACM.
- [44] Marc Cavazza, Fred Charles, and Steven J. Mead. Interacting with virtual characters in interactive storytelling. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1*, AAMAS '02, pages 318–325, New York, NY, USA, 2002. ACM.
- [45] Tanya L. Chartrand and John A. Bargh. The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6):893–910, 1999.
- [46] Herbert H. Clark. *Using language*. Cambridge University Press, Cambridge, England, 1996.
- [47] Mark G. Core and James F. Allen. Coding dialogs with the DAMSL annotation scheme. In *Proceedings of the Working Notes of the AAAI Fall Symposium on Communicative Action in Humans and Machines*, Cambridge, MA, November 1997.
- [48] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. Wizard of oz studies. In *IUI '93: Proceedings of the 1st international conference on Intelligent user interfaces*, pages 193–200, New York, NY, USA, 1993. ACM.
- [49] Doris M. Dehn and Suzanne van Mulken. The impact of animated interface agents: a review of empirical research. *International Journal of Human-Computer Studies*, 52(1):1–22, January 2000.
- [50] Carlo C. DiClemente and Mary M. Velasquez. Motivational interviewing and the stages of change. In William R. Miller and Stephen Rollnick, editors, *Motivational Interviewing: Preparing People for Change*, pages 201–216. Guilford Press, New York, 2002.

- [51] Marc J. Diener, Mark J. Hilsenroth, and Joel Weinberger. A primer on meta-analysis of correlation coefficients: The relationship between patient-reported therapeutic alliance and adult attachment style as an illustration. *Psychotherapy Research*, 19(4-5):519–526, July 2009.
- [52] Marc J. Diener and Joel M. Monroe. The relationship between adult attachment style and therapeutic alliance in individual psychotherapy: a meta-analytic review. *Psychotherapy (Chicago, Ill.)*, 48(3):237–248, September 2011.
- [53] J. M. Digman. Personality structure: Emergence of the Five-Factor model. *Annual Review of Psychology*, 41(1):417–440, 1990.
- [54] Steve Duck. Hypertext in the key of g: Three types of "history" as influences on conversational structure and flow. *Communication Theory*, 12(1):41–62, 2002.
- [55] I. Eibl-Eibesfeldt. Similarities and differences between cultures in expressive movement. In R. A. Hinde, editor, *Non-Verbal Communication*, chapter 11, pages 297–314. Cambridge University Press, 1972.
- [56] Paul Ekman, Richard J. Davidson, and Wallace V. Friesen. The Duchenne smile: Emotional expression and brain physiology II. *Journal of Personality and Social Psychology*, 58(2):342–353, 1990.
- [57] Paul Ekman, Wallace V. Friesen, and Phoebe Ellsworth. *Emotion in the Human Face*. Oxford University Press, 1972.
- [58] Raymond Firth. Verbal and bodily rituals of greeting and parting. In Jean S. La Fontaine, editor, *The Interpretation of Ritual: Essays in Honour of A.I. Richards*, pages 1–38. Tavistock Publications, London, 1972.

- [59] Katherine M. Flegal. Prevalence and trends in obesity among US adults, 1999-2008. *JAMA: The Journal of the American Medical Association*, 303(3):235–241, January 2010.
- [60] R. C. Fraley, N. G. Waller, and K. A. Brennan. An item response theory analysis of self-report measures of adult attachment. *Journal of personality and social psychology*, 78(2):350–365, February 2000.
- [61] Patrick Gebhard. ALMA: a layered model of affect. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, AAMAS '05, pages 29–36, New York, NY, USA, 2005. ACM.
- [62] Andrew Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1:1–19, 2006.
- [63] R. Gockley, A. Bruce, J. Forlizzi, M. Michalowski, A. Mundell, S. Rosenthal, B. Sellner, R. Simmons, K. Snipes, A. C. Schultz, and Jue Wang. Designing robots for long-term social interaction. In *Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 1338–1343, 2005.
- [64] Rachel Gockley, Jodi Forlizzi, and Reid Simmons. Interactions with a moody robot. In *HRI '06: Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 186–193, New York, NY, USA, 2006. ACM.
- [65] Erving Goffman. *Relations in public; microstudies of the public order*. Basic Books, 1971.
- [66] Jonathan Gratch and Stacy Marsella. Tears and fears: modeling emotions and emotional behaviors in synthetic agents. In *AGENTS '01: Proceedings of the fifth international conference on Autonomous agents*, pages 278–285, New York, NY, USA, 2001. ACM.

- [67] Jonathan Gratch, Anna Okhmatovskaia, Francois Lamothe, Stacy Marsella, Mathieu Morales, R. van der Werf, and Louis-Philippe Morency. Virtual rapport. In Jonathan Gratch, Michael Young, Ruth Aylett, Daniel Ballin, and Patrick Olivier, editors, *Intelligent Virtual Agents*, volume 4133, chapter 2, pages 14–27. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.
- [68] Barbara J. Grosz and Candace L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204, 1986.
- [69] Rosanna E. Guadagno, Jim Blascovich, Jeremy N. Bailenson, and Cade Mccall. Virtual humans and persuasion: The effects of agency and behavioral realism. *Media Psychology*, 10(1):1–22, 2007.
- [70] Jarrod Hadfield. MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, 33(2):1–22, December 2009.
- [71] Judith A. Hall. Gender effects in decoding nonverbal cues. *Psychological Bulletin*, 85(4):845–857, 1978.
- [72] M. A. K. Halliday. Notes on transitivity and theme in english: Part 2. *Journal of Linguistics*, 3(2):199–246, October 1967.
- [73] Anthony Hamlett, Louise Ryan, Paulina Serrano-Trespalcacios, and Russ Wolfinger. Mixed models for assessing correlation in the presence of replication. *Journal of the Air & Waste Management Association (1995)*, 53(4):442–450, April 2003.
- [74] Jane Harland, Martin White, Chris Drinkwater, David Chinn, Lorna Farr, and Denise Howel. The newcastle exercise project: a randomised controlled trial of methods to promote physical activity in primary care. *BMJ*, 319(7213):828–832, September 1999.

- [75] Robert L. Hatcher and Arthur J. Gillaspay. Development and validation of a revised short version of the working alliance inventory. *Psychotherapy Research*, 16(1):12–25, January 2006.
- [76] Laurie Hiyakumoto, Scott Prevost, and Justine Cassell. Semantic and discourse information for Text-to-Speech intonation. In *ACL Workshop on Concept-to-Speech Technology*, pages 47–56, 1997.
- [77] Laura Hoffmann, Nicole Krämer, Anh Lam-chi, and Stefan Kopp. Media equation revisited: Do users show polite reactions towards an embodied agent? In Zsófia Ruttkay, Michael Kipp, Anton Nijholt, and Hannes H. Vilhjálmsón, editors, *Intelligent Virtual Agents*, volume 5773, chapter 19, pages 159–165. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
- [78] Kate Hone. Empathic agents to reduce user frustration: The effects of varying agent characteristics. *Interacting with Computers*, 18(2):227–245, March 2006.
- [79] Adam O. Horvath and Leslie S. Greenberg. Development and validation of the working alliance inventory. *Journal of Counseling Psychology*, 36(2):223–233, April 1989.
- [80] Adam O. Horvath and Dianne B. Symonds. Relation between working alliance and outcome in psychotherapy: A Meta-Analysis. *Journal of Counseling Psychology*, 38(2):139–149, 1991.
- [81] Lixing Huang, Louis-Philippe Morency, and Jonathan Gratch. Virtual rapport 2.0. In Hannes H. Vilhjálmsón, Stefan Kopp, Stacy Marsella, and Kristinn R. Thórisson, editors, *Intelligent Virtual Agents*, volume 6895 of *Lecture Notes in Computer Science*, chapter 8, pages 68–79. Springer Berlin / Heidelberg, Berlin, Heidelberg, 2011.

- [82] Amy Isard, Carsten Brockmann, and Jon Oberlander. Individuality and alignment in generated dialogues. In *INLG '06: Proceedings of the Fourth International Natural Language Generation Conference*, pages 25–32, Morristown, NJ, USA, 2006. Association for Computational Linguistics.
- [83] Oliver P. John and Sanjay Srivastava. The big five trait taxonomy: History, measurement, and theoretical perspectives. In Lawrence A. Pervin and Oliver P. John, editors, *Handbook of Personality: Theory and Research*, pages 102–138. Guilford Press, New York, second edition, 1999.
- [84] Lewis W. Johnson and Jeff W. Rickel. Animated pedagogical agents: face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, 11:47–78, 2000.
- [85] Zerrin Kasap, Maher B. Moussa, Parag Chaudhuri, and Nadia Magnenat-Thalmann. Making them RememberEmotional virtual characters with memory. *IEEE Computer Graphics and Applications*, 29(2):20–29, March 2009.
- [86] Adam Kendon. Some relationships between body motion and speech. In Aron Seigman and Benjamin Pope, editors, *Studies in Dyadic Communication*, pages 177–216. Pergamon Press, Elmsford, NY, June 1972.
- [87] Patrick Kenny, Thomas Parsons, Jonathan Gratch, Anton Leuski, and Albert Rizzo. Virtual patients for clinical therapist skills training. In Catherine Pelachaud, Jean-Claude Martin, Elisabeth André, Gérard Chollet, Kostas Karpouzis, and Danielle Pelé, editors, *Intelligent Virtual Agents*, volume 4722 of *Lecture Notes in Computer Science*, chapter 19, pages 197–210. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

- [88] Michael Kipp. ANVIL- a generic annotation tool for multimodal dialogue. *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech)*, pages 1367–1370, 2001.
- [89] Stefan Kopp, Brigitte Krenn, Stacy Marsella, Andrew Marshall, Catherine Pelachaud, Hannes Pirker, Kristinn Thórisson, and Hannes Vilhjálmsón. Towards a common framework for multimodal generation: The behavior markup language. *Intelligent Virtual Agents*, pages 205–217, 2006.
- [90] Stefan Kopp, Paul Tepper, and Justine Cassell. Towards integrated microplanning of language and iconic gesture for multimodal output. In *Proceedings of the 6th international conference on Multimodal interfaces, ICMI '04*, pages 97–104, New York, NY, USA, 2004. ACM.
- [91] Nicole C. Krämer. Social communicative effects of a virtual program guide. In Themis Panayiotopoulos, Jonathan Gratch, Ruth Aylett, Daniel Ballin, Patrick Olivier, and Thomas Rist, editors, *Intelligent Virtual Agents*, volume 3661 of *Lecture Notes in Computer Science*, chapter 37, pages 442–453. Springer Berlin / Heidelberg, Berlin, Heidelberg, 2005.
- [92] John Lafferty, Andrew McCallum, and Fernando Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. 18th International Conf. on Machine Learning*, pages 282–289. Morgan Kaufmann, San Francisco, CA, 2001.
- [93] Marianne LaFrance. Nonverbal synchrony and rapport: Analysis by the Cross-Lag panel technique. *Social Psychology Quarterly*, 42(1):66–70, 1979.
- [94] Marianne LaFrance and William Ickes. Posture mirroring and interactional involvement: Sex and sex typing effects. *Journal of Nonverbal Behavior*, 5(3):139–154–154, March 1981.

- [95] Jessica L. Lakin and Tanya L. Chartrand. Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science*, 14(4):334–339, July 2003.
- [96] J. Richard Landis and Gary G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, March 1977.
- [97] Debi Laplante and Nalini Ambady. On how things are said: Voice tone, voice intensity, verbal content, and perceptions of politeness. *Journal of Language and Social Psychology*, 22(4):434–441, 2003.
- [98] Beth A. Le Poire, Carolyn Shepard, and Ashley Duggan. Nonverbal involvement, expressiveness, and pleasantness as predicted by parental and partner attachment style. *Communication Monographs*, 66(4):293–311, 1999.
- [99] Eun J. Lee and Clifford Nass. Does the ethnicity of a computer agent matter? an experimental comparison of human-computer interaction and computer-mediated communication. In *Proceedings of the 1st Workshop of Embodied Conversational Characters (WECC'98)*, October 1998.
- [100] Jina Lee and Stacy Marsella. Nonverbal behavior generator for embodied conversational agents. In Jonathan Gratch, Michael Young, Ruth Aylett, Daniel Ballin, and Patrick Olivier, editors, *Intelligent Virtual Agents*, volume 4133, chapter 20, pages 243–255. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.
- [101] L. Luborsky, P. Crits-Christoph, L. Alexander, M. Margolis, and M. Cohen. Two helping alliance methods for predicting outcomes of psychotherapy. a counting signs vs. a global rating method. *The Journal of nervous and mental disease*, 171(8):480–491, August 1983.

- [102] Nadia Magnenat-Thalmann and Zerrin Kasap. Modelling socially intelligent virtual humans. In *VRCAI '09: Proceedings of the 8th International Conference on Virtual Reality Continuum and its Applications in Industry*, page 9, New York, NY, USA, 2009. ACM.
- [103] B. H. Marcus, V. C. Selby, R. S. Niaura, and J. S. Rossi. Self-efficacy and the stages of exercise behavior change. *Research Quarterly for Exercise and Sport*, 63(1):60–66, March 1992.
- [104] Daniel J. Martin, John P. Garske, and M. K. Davis. Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. *Journal of consulting and clinical psychology*, 68(3):438–450, 2000.
- [105] Marianne S. Mast and Judith A. Hall. When is dominance related to smiling? assigned dominance, dominance preference, trait dominance, and gender as moderators. *Sex Roles*, 50(5-6):387–399, 2004.
- [106] Michael Mateas and Andrew Stern. Façade: An experiment in building a Fully-Realized interactive drama. In *Game Developers Conference (GDC '03)*, March 2003.
- [107] Charles E. McCulloch and John M. Neuhaus. Generalized linear mixed models. In *Encyclopedia of Biostatistics*. John Wiley & Sons, Ltd, 2005.
- [108] David McNeill. *Hand and Mind: What Gestures Reveal about Thought*. University Of Chicago Press, August 1992.
- [109] Scott McQuiggan, Jonathan Rowe, Sunyoung Lee, and James Lester. Story-Based learning: The impact of narrative on learning experiences and outcomes. In Beverley Woolf, Esma Aïmeur, Roger Nkambou, and Susanne Lajoie, editors, *Intelligent Tutoring Systems*, volume 5091 of *Lecture Notes in Computer Science*, chapter 56, pages 530–539. Springer Berlin / Heidelberg, Berlin, Heidelberg, 2008.

- [110] Albert Mehrabian and Susan R. Ferris. Inference of attitudes from non-verbal communication in two channels. *Journal of consulting psychology*, 31(3):248–252, June 1967.
- [111] George A. Miller. WordNet: a lexical database for english. *Commun. ACM*, 38(11):39–41, November 1995.
- [112] William R. Miller and Stephen Rollnick. *Motivational Interviewing, Second Edition: Preparing People for Change*. The Guilford Press, second edition edition, April 2002.
- [113] Kristen N. Moreno, Natalie K. Person, Amy B. Adcock, Richard N. Van Eck, G. Tanner Jackson, and Johanna C. Marineau. Etiquette and efficacy in animated pedagogical agents: The role of stereotypes. In *AAAI Symposium on Personalized Agents*, 2002.
- [114] Yukiko I. Nakano, Gabe Reinstein, Tom Stocky, and Justine Cassell. Towards a model of face-to-face grounding. In *ACL '03: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, pages 553–561, Morristown, NJ, USA, 2003. Association for Computational Linguistics.
- [115] Lloyd H. Nakatani, Kathleen D. O'Connor, and Carletta H. Aston. Prosodic aspects of american english speech rhythm. *The Journal of the Acoustical Society of America*, 69(S1):S82, 1981.
- [116] Clifford Nass and Kwan M. Lee. Does computer-generated speech manifest personality? an experimental test of similarity-attraction. In *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 329–336, New York, NY, USA, 2000. ACM.
- [117] Clifford Nass and Youngme Moon. Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1):81–103, January 2000.

- [118] Clifford Nass, Youngme Moon, and Paul Carney. Are people polite to computers? responses to Computer-Based interviewing systems. *Journal of Applied Social Psychology*, 29(5):1093–1109, 1999.
- [119] Clifford I. Nass, Youngme Moon, John Morkes, Eun Y. Kim, and B. J. Fogg. Computers are social actors: a review of current research. In Batya Friedman, editor, *Human values and the design of computer technology*, pages 137–162. Center for the Study of Language and Information, Stanford, CA, USA, 1997.
- [120] Michael Neff, Michael Kipp, Irene Albrecht, and Hans P. Seidel. Gesture modeling and animation based on a probabilistic re-creation of speaker style. *ACM Trans. Graph.*, 27(1):1–24, 2008.
- [121] C. R. Nigg, J. S. Rossi, G. J. Norman, and S. V. Benisovich. Structure of decisional balance for exercise adoption. *Annals of Behavioral Medicine*, 20(S211), 1998.
- [122] Maria J. O’Leary and Cynthia Gallois. The last ten turns: Behavior and sequencing in friends’ and strangers’ conversational findings. *Journal of Nonverbal Behavior*, 9(1):8–27, March 1985.
- [123] Andrew Ortony. Value and emotion. In W. Kessen, A. Ortony, and F. Craik, editors, *Memories, thoughts, and emotions: Essays in the honor of George Mandler*, pages 337–353. Hillsdale, NJ: Erlbaum, 1991.
- [124] Catherine Pelachaud, Norman I. Badler, and Mark Steedman. Generating facial expressions for speech. *Cognitive Science*, 20(1):1–46, March 1996.
- [125] Martin J. Pickering and Simon Garrod. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(02):169–190, 2004.

- [126] Jose Pinheiro, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, and R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*, 2012. R package version 3.1-104.
- [127] Hugo Pinto, Yorick Wilks, Roberta Catizone, and Alexiei Dingli. The senior companion multiagent dialogue system. In *AAMAS '08: Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems*, pages 1245–1248, Richland, SC, 2008. International Foundation for Autonomous Agents and Multiagent Systems.
- [128] Sally Planalp. Friends' and acquaintances' conversations II: Coded differences. *Journal of Social and Personal Relationships*, 10(3):339–354, August 1993.
- [129] Sally Planalp and Anne Benson. Friends' and acquaintances' conversations I: Perceived differences. *Journal of Social and Personal Relationships*, 9(4):483–506, November 1992.
- [130] Helmut Prendinger and Mitsuru Ishizuka. Evolving social relationships with animate characters. In *Proceedings of the AISB-02 Symposium on Animating Expressive Characters for Social Interactions*, pages 73–78, 2002.
- [131] Scott Prevost and Mark Steedman. Specifying intonation from context for speech synthesis. *Speech Communication*, 15(1-2):139–153, October 1994.
- [132] J. Prochaska and B. Marcus. The transtheoretical model: applications to exercise. In Rod Dishman, editor, *Advances in Exercise Adherence*, pages 161–180. Human Kinetics, Champaign, Illinois, 1994.
- [133] J. O. Prochaska and W. F. Velicer. The transtheoretical model of health behavior change. *American journal of health promotion : AJHP*, 12(1):38–48, 1997.

-
- [134] James O. Prochaska, Wayne F. Velicer, Joseph S. Rossi, Michael G. Goldstein, Bess H. Marcus, William Rakowski, Christine Fiore, Lisa Harlow, Colleen Redding, Deena Rosenbloom, and Susan R. Rossi. Stages of change and decisional balance for 12 problem behaviors. *Health Psychology*, 13(1):39–46, 1994.
- [135] Hugo Quené. Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *The Journal of the Acoustical Society of America*, 123(2):1104–1113, 2008.
- [136] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.
- [137] Byron Reeves and Clifford Nass. *The Media Equation : How People Treat Computers, Television, and New Media Like Real People and Places*. CSLI Lecture Notes. Center for the Study of Language and Inf, January 2003.
- [138] Debra Revere and Peter J. Dunbar. Review of computer-generated outpatient health behavior interventions. *Journal of the American Medical Informatics Association*, 8(1):62–79, January 2001.
- [139] Virginia P. Richmond and James C. McCroskey. *Nonverbal communication in interpersonal relations*. Allyn and Bacon, 2000.
- [140] Jeff Rickel and W. Lewis Johnson. Task-Oriented collaboration with embodied agents in virtual worlds. In Justine Cassell, Joseph Sullivan, and Scott Prevost, editors, *Embodied Conversational Agents*. MIT Press, 2000.
- [141] Raoul Rickenberg and Byron Reeves. The effects of animated characters on anxiety, task performance, and evaluations of user interfaces. In

- CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 49–56, New York, NY, USA, 2000. ACM.
- [142] Ronald E. Riggio, Howard S. Friedman, and M. Robin DiMatteo. Non-verbal greetings: effects of the situation and personality. *Personality and Social Psychology Bulletin*, 7(4):682–689, December 1981.
- [143] Daniela M. Romano, Gary Sheppard, James Hall, Adam Miller, and Zhinan Ma. BASIC: A believable adaptable socially intelligent character for social presence. In *PRESENCE 2005, The 8th Annual International Workshop on Presence*, September 2005.
- [144] Albert E. Schefflen. The significance of posture in communication systems. *Psychiatry*, 27:316–331, November 1964.
- [145] Emanuel A. Schegloff and Harvey Sacks. Opening up closings. *Semiotica*, 8:289–327, 1973.
- [146] Deborah Schiffrin. *Discourse Markers*. Cambridge University Press, 1987.
- [147] Klaus P. Schneider. *Small talk: Analyzing phatic discourse*. Hitzeroth, 1988.
- [148] Daniel Schulman and Timothy Bickmore. Posture, relationship, and discourse structure. In Hannes Vilhjálmsón, Stefan Kopp, Stacy Marsella, and Kristinn Thórisson, editors, *Intelligent Virtual Agents*, volume 6895 of *Lecture Notes in Computer Science*, pages 106–112, Berlin, Heidelberg, 2011. Springer Berlin / Heidelberg.
- [149] Daniel Schulman, Timothy Bickmore, and Candace Sidner. An intelligent conversational agent for promoting long-term health behavior change using motivational interviewing. In *AAAI Spring Symposium on AI and Health Communication*, 2011.

- [150] Nicole Shechtman. *Talking to People Versus Talking to Computers: Interpersonal Goals as a Distinguishing Factor*. PhD thesis, Stanford University, June 2002.
- [151] Nicole Shechtman and Leonard M. Horowitz. Media inequality in conversation: how people behave differently when interacting with computers and people. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '03, pages 281–288, New York, NY, USA, 2003. ACM.
- [152] Gregor Sieber and Brigitte Krenn. Towards an episodic memory for companion dialogue. In Jan Allbeck, Norman Badler, Timothy Bickmore, Catherine Pelachaud, and Alla Safonova, editors, *Intelligent Virtual Agents*, volume 6356 of *Lecture Notes in Computer Science*, chapter 33, pages 322–328–328. Springer Berlin / Heidelberg, Berlin, Heidelberg, 2010.
- [153] Bruce L. Smith, Bruce L. Brown, William J. Strong, and Alvin C. Rencher. Effects of speech rate on personality perception. *Language and Speech*, 18(2):145–152, April 1975.
- [154] Cameron Smith, Marc Cavazza, Daniel Charlton, Li Zhang, Markku Turunen, and Jaakko Hakulinen. Integrating planning and dialogue in a lifestyle agent. In Helmut Prendinger, James Lester, and Mitsuru Ishizuka, editors, *Intelligent Virtual Agents*, volume 5208 of *Lecture Notes in Computer Science*, chapter 15, pages 146–153. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.
- [155] Michael Smithson and Jay Verkuilen. A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, 11(1):54–71, March 2006.
- [156] David J. Spiegelhalter, Nicola G. Best, Bradley P. Carlin, and Angelika Van Der Linde. Bayesian measures of model complexity and fit. *Jour-*

- nal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, October 2002.
- [157] Sanjay Srivastava, Oliver P. John, Samuel D. Gosling, and Jeff Potter. Development of personality in early and middle adulthood: set like plaster or persistent change? *Journal of personality and social psychology*, 84(5):1041–1053, May 2003.
- [158] Mark Steedman. Information structure and the Syntax-Phonology interface. *Linguistic Inquiry*, 31:649–689, 2000.
- [159] S. Thomas, J. Reading, and R. J. Shephard. Revision of the physical activity readiness questionnaire (PAR-Q). *Canadian journal of sport sciences*, 17(4):338–345, December 1992.
- [160] Linda Tickle-Degnen and Elizabeth Gavett. Changes in nonverbal behavior during the development of therapeutic relationships. In Pierre Philippot, Robert S. Feldman, and Erik J. Coats, editors, *Nonverbal behavior in clinical settings*, chapter 4, pages 75–110. Oxford University Press, New York, NY, USA, 2003.
- [161] Linda Tickle-Degnen and Robert Rosenthal. Group rapport and nonverbal behavior. In Clyde Hendrick, editor, *Group processes and intergroup relations*, volume 9 of *Review of Personality and Social Psychology*, pages 113–136. Sage, Newbury Park, CA, 1987.
- [162] Linda Tickle-Degnen and Robert Rosenthal. The nature of rapport and its nonverbal correlates. *Psychological Inquiry: An International Journal for the Advancement of Psychological Theory*, 1(4):285–293, 1990.
- [163] Obed Torres, Justine Cassell, and Scott Prevost. Modeling gaze behavior as a function of discourse structure. In *In Proceedings of the First International Workshop on Human-Computer Conversations*, 1997.

- [164] April R. Trees and Valerie Manusov. Managing face concerns in criticism integrating nonverbal behaviors as a dimension of politeness in female friendship dyads. *Human Communication Research*, 24(4):564–583, 1998.
- [165] Joan Tucker and Sherry Anders. Adult attachment style and nonverbal closeness in dating couples. *Journal of Nonverbal Behavior*, 22(2):109–124, June 1998.
- [166] Geert Verbeke and Geert Molenberghs. *Linear Mixed Models for Longitudinal Data*. Springer, corrected edition, June 2000.
- [167] Hannes Vilhjálmsson, Nathan Cantelmo, Justine Cassell, Chafai, Michael Kipp, Stefan Kopp, Maurizio Mancini, Stacy Marsella, Andrew N. Marshall, Catherine Pelachaud, Zsofi Ruttkay, Kristinn R. Thórisson, Herwin Welbergen, and Rick J. Werf. The behavior markup language: Recent developments and challenges intelligent virtual agents. In Catherine Pelachaud, Jean-Claude Martin, Elisabeth André, Gérard Chollet, Kostas Karpouzis, and Danielle Pelé, editors, *Intelligent Virtual Agents*, volume 4722 of *Lecture Notes in Computer Science*, chapter 10, pages 99–111. Springer Berlin / Heidelberg, Berlin, Heidelberg, 2007.
- [168] Marilyn A. Walker, Janet E. Cahn, and Stephen J. Whittaker. Improvising linguistic style: social and affective bases for agent personality. In *AGENTS '97: Proceedings of the first international conference on Autonomous agents*, pages 96–105, New York, NY, USA, 1997. ACM.
- [169] Rebecca M. Warner. Speaker, partner and observer evaluations of affect during social interaction as a function of interaction tempo. *Journal of Language and Social Psychology*, 11(4):253–266, December 1992.
- [170] P. H. Waxer. Nonverbal cues for anxiety: an examination of emotional leakage. *Journal of abnormal psychology*, 86(3):306–314, June 1977.

-
- [171] Jerry S. Wiggins. A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology*, 37(3):395–412, March 1979.
- [172] Yorick Wilks. Artificial companions. In Samy Bengio and Hervé Bourlard, editors, *Machine Learning for Multimodal Interaction*, volume 3361, chapter 4, pages 36–45. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- [173] Nick Yee, Jeremy N. Bailenson, and Kathryn Rickertsen. A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 1–10, New York, NY, USA, 2007. ACM.
- [174] Jiahong Yuan, Mark Liberman, and Christopher Cieri. Towards an integrated understanding of speaking rate in conversation. In *International Conference on Spoken Language Processing (INTERSPEECH-2006)*, 2006.

APPENDIX A

Experimental Protocol for the Collection of the Exercise Counseling Corpus

The following sections give the detailed protocol for interaction between the researcher and the clients. Statements to be said verbatim to clients are in plain text, while other actions to be taken by the researcher are in bold text and enclosed in brackets, **[as in this example]**.

The protocol begins (for both the initial and subsequent sessions) with the arrival of the client, and assumes that the recording apparatus and experiment room have been prepared. Eligibility screening should have been performed by phone or email prior to scheduling the initial session. The counselor should be waiting in a separate room, out of view of the client.

A.1 Initial Session

[Ask the client to sit in the client’s chair, and sit in the counselor’s chair facing him/her].

Hi, welcome to Northeastern. We’re looking at how people talk about things like physical activity, to help us learn how to make computer systems that can help people get more physical activity and be healthier. We’re looking for participants to have several short conversations about physical activity with a counselor. if you agree to participate we’ll ask you to come back once

a week for six weeks, including today. Each of those sessions will take about half an hour of your time, and you'll be paid \$10 for each session, so that's \$60 if you come back every week.

If you participate, you'll have a short conversation — about 15 minutes — with the counselor each week. We'll videotape those conversations [**indicate cameras and microphone**], and after each conversations, we'll ask you to fill out some questionnaires to tell us how you felt about it. We'll also ask you to watch a videotape of each conversation while answering some more questions about it. Finally, we'll give you a pedometer [**indicate pedometer**], which is a small device you can wear on your belt or put in your pocket that records how many steps you take; this lets us measure how much physical activity you are doing.

Does this sound like something you would be interested in participating in? [**If not, thank and dismiss**]

Great! I just need to ask you a few questions before we begin. We asked you the same questions when you first signed up to come in today, but I'll ask them again now just to be sure.

Are you at least 18 years old? [**If not, thank and dismiss**]

Are you a native English speaker? Is English your first language? [**If not, thank and dismiss**]

Has your doctor ever said that you have a heart condition and that you should only do physical activity recommended by a doctor? [**If yes, thank and dismiss**]

Do you feel pain in your chest when you do physical activity? [**If yes, thank and dismiss**]

In the past month, have you had chest pain when you were not doing physical activity? [**If yes, thank and dismiss**]

Do you lose your balance because of dizziness, or do you ever lose consciousness? [**If yes, thank and dismiss**]

Do you have a bone or joint problem that could be made worse by a change

in your physical activity? **[If yes, thank and dismiss]**

Is your doctor currently prescribing drugs (for example, water pills) for your blood pressure or heart condition? **[If yes, thank and dismiss]**

Do you know of any other reason why you should not do physical activity? **[If yes, thank and dismiss]**

Regular Exercise is any planned physical activity, such as brisk walking, jogging, bicycling, or swimming, performed to increase physical fitness. Such activity should be performed 3 to 5 times per week for 20–60 minutes per session. Exercise does not have to be painful to be effective but should be done at a level that increases your breathing rate and causes you to break a sweat.

According to that definition — 3 times a week, 20 minutes a session — would you say that you are currently getting regular exercise? **[If yes:]** Have you been getting regular exercise for more than 6 months? **[If yes, thank and dismiss]**

Thanks. Now we have some paperwork. This is a consent form **[indicate consent form]** which we will go over together, and then I will ask you to sign. It says that you are agreeing to participate in a research study, and that we have informed you of the risks and benefits of the study, and that you understand your rights as a participant. **[Explain consent form, being sure to make the following points:]**

- You are participating of your own free will, and you may drop out any time without penalty.
- You have not committed to anything yet. Just being here right now does not require you to participate if you do not want to.
- You are being asked to come here and talk about physical activity with the counselor. The counselor may also ask you to do walking or other physical activity, but coming here for the session is all that's required: you'll still receive full payment even if you never do any physical activity.

- Your participation is voluntary at all times. You can stop participating at any time, for any reason. You will still be paid the full amount for any sessions you've completed.
- The risks of this study are low, only those associated with moderate exercise like brisk walking. However, as this is research and experimental, there may be negative outcomes or no benefit to you at all.
- Your participation in this study is considered private information. We will not release your name or other identifying information without permission.
- If you have any questions about your rights as a participant, either now or later, you can contact the Human Subjects Research Protection office here at Northeastern. The contact information is included on the consent form.

Please take as long as you want to read over the consent form, and if you're okay with everything, then sign on the third page. Let me know if you have any questions. **[Answer all questions. If he/she signs, then add your signature, otherwise thank and dismiss]**

Thanks. Now we have one additional consent form **[Indicate videotape consent form]**. This asks for permission to show very short clips of the videotape in educational contexts, such as explaining this research or showing students how to conduct a research project. You don't have to agree if you don't want to: if you say no, then nothing else changes, and you'll still get the full payment. **[If he/she does not sign, cross out and initial the signature line of the form]**

Do you want a copy of the consent forms for your records? **[If yes, make copies during initial questionnaires, below]**

Since this is the first session, before we get started and you meet the counselor, we have some questionnaires for you to fill out. These will tell us

about you and how you feel about physical activity. They'll take you a few minutes to get through; I'll leave to set everything up and come back when you're done. **[Give participant the sociodemographics, BFI, ECR-R, Stage of Change, and Decisional Balance questionnaires, and leave the room]**

[If requested, make copies of the consent forms. Fetch the counselor, but do not bring her into the experiment room yet. Check the video setup. When the participant has completed all questionnaires, enter the room.]

Thanks. There will be a few more questionnaires after the conversation, but not nearly as many. In a moment, I'll leave again and then the counselor will come in and start the first conversation. Please stay seated here when the conversation is done. Any questions? **[Answer questions]**

[Leave the room. Start recording. Send the counselor in, and close the door. Do not interrupt or interfere with the conversation except in case of drastic problems. When the conversation ends and the counselor leaves the room, stop recording.]

[Give the cognitive inventory and WAI-SR questionnaires to the counselor, and send her to a separate room to complete them.]

[Enter the room] Thanks! Now, I'll give you a couple short questionnaires about that conversation. **[Give the cognitive inventory instructions and response form to the client]** For this one, please write down your thoughts about the conversation and the counselor: whatever pops into your head first. Don't worry about spelling, grammar, or handwriting, but try and write each thought in a separate box. You have two minutes. **[Time two minutes, then collect the response form].**

Next, we have one more questionnaire. **[Give the WAI-SR to the client; while he/she is completing it, set up the equipment for the retrospective review]** Thanks! **[Collect the WAI-SR]**

The last thing I'll ask you to do today is to watch a video of the conversa-

tion you just had. While you're watching it, you'll be videotaped again, and we ask you to do two things. First, we're interested in how you were feeling during the conversation, so try and say out loud any thoughts you had about the conversation. Second, we want you to use this paper [**Place trust measure on the table**] to tell us how your trust in the counselor changed during the conversation. Just point to the paper while watching to show how much you trusted the counselor at that time. Any questions? [**Answer questions**]

[**Turn on the large-screen display. Leave the room and close the door. Start video recording, and then playback. When playback is complete, shut off recording and reenter the room.**]

That's everything for today. Before you go, I'll give you the pedometer we mentioned earlier. [**Indicate pedometer**] Try to wear it as much as possible. It can go either in your pocket or on a belt [**Demonstrate use**]. Make sure to bring it with you to the next sessions [**Give the pedometer to the client**]

Thanks! We'll see you next week. [**Give \$10 payment to the client, and schedule the next session**]

A.2 Subsequent Sessions

[**Ask the client to sit in the client's chair, and ask to take the pedometer. If it is forgotten or missing, note in the experiment log, otherwise ask the client to wait and download the step data; it should make a printout. Fetch the counselor, and give her the printout, but do not bring her into the room yet. Enter the room, and return the pedometer.**]

Thanks. We'll start this week's conversation now. Any questions? [**Answer questions**]

[**Leave the room. Start recording. Send the counselor in, and close the door. Do not interrupt or interfere with the conversation except in case of drastic problems. When the conversation ends and**

the counselor leaves the room, stop recording.]

[Give the cognitive inventory and WAI-SR questionnaires to the counselor, and send her to a separate room to complete them.]

[Enter the room] Thanks! [Give the cognitive inventory instructions and response form to the client]. Like before, just write down your thoughts, one per box. You have two minutes. [Time two minutes, then collect the response form]

Here are this week's questionnaires. [For weeks 3 and 5, give the WAI-SR questionnaire to the client. For weeks 2, 4, and 6 give the WAI-SR, Stage of Change, and Decisional Balance questionnaires to the client. While he/she is completing the questionnaire(s), set up up the equipment for the retrospective review.] Thanks! [Collect the questionnaire(s)]

Now it's time to watch the video of the conversation you just had. Like last time, you'll be videotaped again, and please try and point to the paper to show how much you trusted the counselor during the conversation, and try and talk about what you were feeling at the time. Any questions? [Answer questions]

[Turn on the large-screen display. Leave the room and close the door. Start video recording, and then playback. When playback is complete, shut off recording and reenter the room.]

[Thank and pay the participant]. Thanks! That's everything for today. We'll see you next week. [Give \$10 payment to the client, and, for the second through fifth weeks, schedule the next session]

Volunteers are needed for a research study on how people talk about exercise

\$60 and 6 free weeks of exercising counseling

Researchers at Northeastern University are studying how people talk about exercise and other health-related issues. We are looking for volunteers who are willing to be videotaped talking to an exercise counselor once a week for 6 weeks.

You must be 18 or older, and a native English speaker in order to participate. We are looking for volunteers who are **not** currently exercising regularly.

The study will be held at Northeastern University. The first session takes 1 hour, and the rest take 30 minutes.

If you are interested, please call (617) 373-4605, or email expt@neu.edu.

Recruitment Material

Volunteers Needed for a Research Study

\$60 and 6 weeks of free exercise counseling

Researchers at Northeastern University are studying how people talk about exercise and other health-related issues.

We are looking for volunteers who are willing to be videotaped talking to an exercise counselor once a week for 6 weeks.

You must be 18 or older, and a native English speaker in order to participate. We are looking for volunteers who are **not** currently exercising regularly.

The study will be held at Northeastern University. The first session takes 1 hour, and the rest take 30 minutes.

**If you are interested,
call (617) 373-4605, or email expt@neu.edu.**



Exercise counseling study (617)-373-4605 expt@neu.edu	Exercise counseling study (617)-373-4605 expt@neu.edu	Exercise counseling study (617)-373-4605 expt@neu.edu	Exercise counseling study (617)-373-4605 expt@neu.edu	Exercise counseling study (617)-373-4605 expt@neu.edu	Exercise counseling study (617)-373-4605 expt@neu.edu	Exercise counseling study (617)-373-4605 expt@neu.edu	Exercise counseling study (617)-373-4605 expt@neu.edu
---	---	---	---	---	---	---	---

Recruitment Material

APPENDIX A. EXPERIMENTAL PROTOCOL FOR THE COLLECTION OF
216 THE EXERCISE COUNSELING CORPUS

Please take a few minutes to answer the following questions to see if you are eligible to participate.

	<u>YES</u>	<u>NO</u>
1. Has your doctor ever said that you have a heart condition and that you should only do physical activity recommended by a doctor?	_____	_____
2. Do you feel pain in your chest when you do physical activity?	_____	_____
3. In the past month, have you had chest pain when you were not doing physical activity?	_____	_____
4. Do you lose your balance because of dizziness, or do you ever lose consciousness?	_____	_____
5. Do you have a bone or joint problem that could be made worse by a change in your physical activity?	_____	_____
6. Is your doctor currently prescribing drugs (for example, water pills) for your blood pressure or heart condition?	_____	_____
7. Do you know of <u>any other reason</u> why you should not do physical activity?	_____	_____

Eligibility Screening: PAR-Q

Northeastern University, College of Computer and Information Science
Timothy W. Bickmore, Ph.D.
Nonverbal Behavior during the Development of a Counseling Relationship

Informed Consent to Participate in a Research Study

We are inviting you to take part in a research study. This form will tell you about the study, but the researcher will explain it to you first. You may ask this person any questions that you have. When you are ready to make a decision, you may tell the researcher if you want to participate or not. You do not have to participate if you do not want to. If you decide to participate, the researcher will ask you to sign this statement and will give you a copy to keep.

Why am I being asked to take part in this research study?

You are being asked to participate in this study because you are an English speaking adult, and are not currently engaged in regular exercise.

Why are you doing this research study?

The purpose of this research is to study how people talk to each other when discussing exercise and other health-related issues.

What will I be asked to do?

If you decide to take part in this study, we will first ask you some questions about your background and ask you to fill out some questionnaires. Some of these questionnaires will tell us about your personality, while others will ask how you feel about exercise.

You will have a short conversation about exercise with a counselor. This conversation will be videotaped. After the conversation, we will ask you some more questions. You will then watch a videotape of the conversation, while answering questions about it.

We will give you a pedometer to take home with you, which you will carry for the next 5 weeks. This pedometer will record how many steps you walk each day.

Once a week, you will return here for another session. You will have another conversation with the counselor, answer questions, and watch a videotape of the conversation while answering questions about it.

Where will this take place and how much of my time will it take?

The study will take place in the Human-Computer Interaction Laboratory at Northeastern University. The first session, today, will take approximately 1 hour.

You will come back for 5 additional sessions about once a week. Each additional session will take approximately 30 minutes.

Consent Form

Will there be any risk or discomfort to me?

You will give up approximately 1 hour of your time today, plus five additional half-hour meetings once a week.

There is a possibility that some of the questions you will be asked might make you feel uncomfortable. In that case, you are free to refuse to answer any question and/or discontinue the study. Your refusal to answer will not have any impact on your participation in the study.

There are some risks associated with increasing your physical activity, such as physical injury.

Will I benefit by being in this research?

You may receive no direct benefit from taking part in the study, although you may increase the amount of physical activity you perform on a regular basis. Your participation may help the investigators learn to build computer systems that will help people get more exercise.

Who will see the information about me?

Information from this study may be used for research purposes and may be published. However, your name will not be used in any publications. All videotapes will be stored in a locked cabinet, accessible only to the researchers, and will be not be shown to anyone unless you give explicit permission for them to be used for teaching purposes.

In rare instances, authorized people may request to see research information about you and other people in this study. This is done only to be sure that the research is done properly. We would only permit people who are authorized by organizations such as Northeastern University or the federal government to see this information.

If I don't want to take part in the study, what choices do I have?

You have the option to not participate in the study.

What will happen if I suffer any harm from this research?

No special arrangements will be made for compensation or for payment for treatment solely because of your participation in this research.

Can I stop my participation in this study?

Your participation in this research is completely voluntary. You do not have to participate if you do not want to. Even if you begin the study, you may quit at any time. If you do not participate or if you decide to quit, you will not lose any rights, benefits, or services that you would otherwise have.

Who can I contact if I have questions or problems?

If you have questions or concerns at any time, or if you need to report an injury while participating in this research, contact TIMOTHY BICKMORE at (617) 373-5477.

Consent Form

Who can I contact about my rights as a participant?

If you have any questions about your rights in this research, you may contact Nan C. Regina, Director, Human Subject Research Protection, 960 Renaissance Park, Northeastern University, Boston, MA 02115. Tel: 617.373.4588, Email: irb@neu.edu. You may call anonymously if you wish.

Will I be paid for my participation?

You will be paid \$10 for each completed session, at the end of the session. You will receive a total of \$60 if you participate in the entire study.

Will it cost me anything to participate?

There are no costs to you for participating in this research study.

Is there anything else I need to know?

- You must be at least 18 years old to participate.
- You will be one of approximately 30 people to be asked to participate in this study.

I agree to take part in this research.

Signature of person agreeing to take part

Date

Printed name of person above

Signature of person who explained the study to the participant above and obtained consent

Date

Printed name of person above

Consent Form

Use of Videotape

We would also appreciate your consent to use very short clips of the videotapes to show in educational contexts. If we do use short clips of you, we will edit them to only show short examples of your interaction with the counselor.

This consent is entirely separate from your consent to participate in the experiment and may be withdrawn at any time in the future. To give your permission to have an excerpt from the videotape potentially shown for research and teaching purposes, please sign below. Please note that if you do not wish to give your permission for this, you may still participate in the study.

I have read and understand the above and agree to potentially have an excerpt from my videotaped interaction shown for educational purposes.

Name _____

—

Signature _____ Date _____

Consent Form

Counselor Instructions

You will be meeting with each participant up to six times, approximately once per week. During each session, you will discuss the participant's exercise behavior and any related issues which arise. Your overall goal for the intervention is for the participant to begin a program of moderate and regular exercise, which you believe that he/she will be able to maintain once the intervention is ended.

The specific target behavior is brisk walking; however please feel free to suggest additional or alternative activities whenever you feel it is appropriate. Each session is expected to consist of a 10-15 minute conversation, which you may conduct however you feel will be most effective.

You will be given basic demographic information (including age, height, and weight) on each participant. All participants have indicated that they are not currently maintaining a regular exercise program, and that they have no known health problems which would prevent them from engaging in regular exercise. Each participant will be given a pedometer for the duration of their participation in the study and instructed to wear it every day. You will have access to step counts gathered from these pedometers.

All sessions with participants will be videotaped for analysis. You will not have access to these videotapes during the duration of the study. Please do not discuss the videotapes themselves, or the analysis of them, with participants. If a participant asks about them during a session, please tell them only that you do not look at the videotapes, and that they are analyzed for research purposes only.

Participants will also answer a series of questionnaires before the first session, and after every session. You will also complete questionnaires after each session. You will not have access to these questionnaires during the duration of the study. Please do not discuss the questionnaires with participants during sessions. If a participant asks about them during a session, please tell them only that you do not look at the questionnaires, and that what they answer will not change what you do during the session in any way.

Counselor Instructions

Please take a moment and answer a few questions about yourself:

Date of Birth: _____

Sex: M / F

Height: _____

Weight: _____

Do you smoke: Y / N

Ethnic Background (check one):

American Indian or Alaskan Native _____
Asian or Pacific Islander _____
Black, Not of Hispanic Origin _____
White, Not of Hispanic Origin _____
Hispanic _____

Marital Status (check one):

Single _____
Married _____
Divorced/Widowed _____

Last grade of school completed (check one):

Less than high school (0-8) _____
Some high school _____
High school graduate or GED _____
Technical school education _____
Some college _____
College graduate _____
Advanced degree _____

Occupation: _____

Sociodemographics

Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who likes to spend time with others? Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement.

Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly
1	2	3	4	5

I See Myself as Someone Who...

- | | |
|---|--|
| ___1. Is talkative | ___23. Tends to be lazy |
| ___2. Tends to find fault with others | ___24. Is emotionally stable, not easily upset |
| ___3. Does a thorough job | ___25. Is inventive |
| ___4. Is depressed, blue | ___26. Has an assertive personality |
| ___5. Is original, comes up with new ideas | ___27. Can be cold and aloof |
| ___6. Is reserved | ___28. Perseveres until the task is finished |
| ___7. Is helpful and unselfish with others | ___29. Can be moody |
| ___8. Can be somewhat careless | ___30. Values artistic, aesthetic expression |
| ___9. Is relaxed, handles stress well | ___31. Is sometimes shy, inhibited |
| ___10. Is curious about many different things | ___32. Is considerate and kind to almost everyone |
| ___11. Is full of energy | ___33. Does things efficiently |
| ___12. Starts quarrels with others | ___34. Remains calm in tense situations |
| ___13. Is a reliable worker | ___35. Prefers work that is routine |
| ___14. Can be tense | ___36. Is outgoing, sociable |
| ___15. Is ingenious, a deep thinker | ___37. Is sometimes rude to others |
| ___16. Generates lots of enthusiasm | ___38. Makes plans and follows through with them |
| ___17. Has a forgiving nature | ___39. Gets nervous easily |
| ___18. Tends to be disorganized | ___40. Likes to reflect, play with ideas |
| ___19. Worries a lot | ___41. Has few artistic interests |
| ___20. Has an active imagination | ___42. Likes to cooperate with others |
| ___21. Tends to be quiet | ___43. Is easily distracted |
| ___22. Is generally trusting | ___44. Is sophisticated in art, music, or literature |

Please check: Did you write a number in front of each statement?

Personality: BFI

Page 1 of 2

The statements below concern how you feel in emotionally intimate relationships. We are interested in how you *generally* experience relationships, not just in what is happening in a current relationship. Respond to each statement by writing a number to indicate how much you agree or disagree with the statement.

Strongly disagree	Disagree	Disagree a little	Neither agree nor disagree	Agree a little	Agree	Strongly agree
1	2	3	4	5	6	7

___1. I usually discuss my problems and concerns with others.

___2. It helps to turn to people I am close to in times of need.

___3. I find that others don't want to get as close as I would like.

___4. Sometimes others change their feelings about me for no apparent reason.

___5. I talk things over with people I am close to.

___6. I find it easy to depend on others.

___7. I do not often worry about being abandoned.

___8. It makes me mad that I don't get the affection and support I need from others.

___9. I find it difficult to allow myself to depend on others.

___10. I worry a lot about my relationships.

___11. When a person I am close to is out of sight, I worry that he or she might become interested in someone else.

___12. I get uncomfortable when others want to be very close.

___13. I'm afraid that once someone gets to know me, he or she won't like who I really am.

___14. I feel comfortable sharing my private thoughts and feelings with others.

___15. My desire to be very close sometimes scares people away.

___16. Other people make me doubt myself.

___17. I often worry that others don't really love me.

___18. I prefer not to be too close to others.

Please check: Did you write a number in front of each statement?

Attachment Style: ECR-R

The statements below concern how you feel in emotionally intimate relationships. We are interested in how you *generally* experience relationships, not just in what is happening in a current relationship. Respond to each statement by writing a number to indicate how much you agree or disagree with the statement.

Strongly disagree	Disagree	Disagree a little	Neither agree nor disagree	Agree a little	Agree	Strongly agree
1	2	3	4	5	6	7

- ___ 19. I worry that I won't measure up to other people.
- ___ 20. I rarely worry about others leaving me.
- ___ 21. I worry that others won't care about me as much as I care about them.
- ___ 22. I am very comfortable being close to others.
- ___ 23. I often wish that others' feelings for me were as strong as my feelings for them.
- ___ 24. I don't feel comfortable opening up to others.
- ___ 25. I tell people I am close to just about everything.
- ___ 26. It's easy for me to be affectionate with others.
- ___ 27. Other people only seem to notice me when I'm angry.
- ___ 28. I find it relatively easy to get close to others.
- ___ 29. People that I am close to really understand me and my needs.
- ___ 30. It's not difficult for me to get close to others.
- ___ 31. I am nervous when others get too close to me.
- ___ 32. I prefer not to show others how I feel deep down.
- ___ 33. I'm afraid that I will lose others' love.
- ___ 34. When I show my feelings for others, I'm afraid they will not feel the same about me.
- ___ 35. I feel comfortable depending on others.
- ___ 36. I often worry that others will not want to stay with me.

Please check: Did you write a number in front of each statement?

Attachment Style: ECR-R

We are now interested in your thoughts about the conversation.

Please list any thoughts, regardless of whether they are about you, the counselor, the situation, or anything else; whether they are positive, negative or neutral.

Please be completely honest. The counselor will not see your responses.

Simply write the first thought you have in the first box, the second in the second box, and so on. Please put only one idea or thought in a box. Don't worry about spelling, handwriting, grammar, or punctuation.

You will have 2 minutes to write. You can write as many thoughts as you like; you do not need to fill every box. If you fill all the boxes, and wish to list more thoughts, please continue on to the next page.

Cognitive Inventory (Instruction)

We are now interested in your thoughts about the conversation.

Please list any thoughts, regardless of whether they are about you, the client, the situation, or anything else; whether they are positive, negative or neutral.

Please be completely honest. The client will not see your responses.

Simply write the first thought you have in the first box, the second in the second box, and so on. Please put only one idea or thought in a box. Don't worry about spelling, handwriting, grammar, or punctuation.

You will have 2 minutes to write. You can write as many thoughts as you like; you do not need to fill every box. If you fill all the boxes, and wish to list more thoughts, please continue on to the next page.

Cognitive Inventory (Counselor Instruction)

Cognitive Inventory (Response Form)

Below is a series of statements about experiences people might have with their counseling or counselor. Some items refer directly to your counselor with an underlined space -- as you read the sentences, mentally insert the name of your counselor in place of _____ in the text.

For each statement, please take your time to consider your own experience and then respond by writing a number.

Seldom	Sometimes	Fairly often	Very often	Always
1	2	3	4	5

- ___1. As a result of these sessions I am clearer as to how I might be able to change.
- ___2. What I am doing in counseling gives me a new way of looking at my problem.
- ___3. I believe _____ likes me.
- ___4. _____ and I collaborate on setting goals for my counseling.
- ___5. _____ and I respect each other.
- ___6. _____ and I are working toward mutually agreed-upon goals.
- ___7. I feel that _____ appreciates me.
- ___8. We agree on what is important for me to work on.
- ___9. I feel _____ cares about me even when I do things that he/she does not approve of.
- ___10. I feel that the things I do in counseling will help me to accomplish the changes that I want.
- ___11. We have established a good understanding of the kind of changes that would be good for me.
- ___12. I believe the way we are working with my problem is correct.

Please check: Did you write a number in front of each statement?

Therapeutic Alliance: WAI-SR

APPENDIX A. EXPERIMENTAL PROTOCOL FOR THE COLLECTION OF
230 THE EXERCISE COUNSELING CORPUS

Below is a series of statements about experiences people might have with their counseling or counselor. Some items refer directly to your client with an underlined space -- as you read the sentences, mentally insert the name of your client in place of _____ in the text.

For each statement, please take your time to consider your own experience and then respond by writing a number.

Seldom 1	Sometimes 2	Fairly often 3	Very often 4	Always 5
____1. As a result of these sessions _____ is clearer as to how he/she might be able to change.				
____2. What _____ is doing in counseling gives him/her a new way of looking at his/her problem.				
____3. _____ believes I like him/her.				
____4. _____ and I collaborate on setting goals for his/her counseling.				
____5. _____ and I respect each other.				
____6. _____ and I are working toward mutually agreed-upon goals.				
____7. _____ feels that I appreciate him/her.				
____8. We agree on what is important for _____ to work on.				
____9. _____ feels that I care about him/her even when he/she does things I do not approve of.				
____10. I feel that the things _____ does in counseling will help him/her to accomplish the changes that he/she wants.				
____11. We have established a good understanding of the kind of changes that would be good for _____.				
____12. _____ believes the way we are working with his/her problem is correct.				

Please check: Did you write a number in front of each statement?

Therapeutic Alliance: WAI-SR (Counselor)

Regular Exercise is any *planned* physical activity (e.g., brisk walking, aerobics, jogging, bicycling, swimming, rowing, etc.) performed to increase physical fitness. Such activity should be performed 3 to 5 times per week for 20-60 minutes per session. Exercise does not have to be painful to be effective but should be done at a level that increases your breathing rate and causes you to break a sweat.

Do you exercise regularly according to that definition? (Please check one):

_____ Yes, I have been for MORE than 6 months.

_____ Yes, I have been for LESS than 6 months.

_____ No, but I intend to in the next 30 days.

_____ No, but I intend to in the next 6 months.

_____ No, and I do NOT intend to in the next 6 months.

Stage of Change

The statements below look at positive and negative aspects of exercise. Read the following items and write a number next to each statement to indicate how important each statement is with respect to your decision to exercise or not to exercise in your leisure time.

If you disagree with a statement and are unsure how to answer, the statement is probably not important to you.

Not important	A little bit important	Somewhat important	Quite important	Extremely important
1	2	3	4	5

___1. I would have more energy for my family and friends if I exercised regularly.

___2. I would feel embarrassed if people saw me exercising.

___3. I would feel less stressed if I exercised regularly.

___4. Exercise prevents me from spending time with my friends.

___5. Exercising puts me in a better mood for the rest of the day.

___6. I feel uncomfortable or embarrassed in exercise clothes.

___7. I would feel more comfortable with my body if exercised regularly.

___8. There is too much I would have to learn to exercise.

___9. Regular exercise would help me have a more positive outlook on life.

___10. Exercise puts an extra burden on my significant other.

Please check: Did you write a number in front of each statement?

Decisional Balance

As you watch the conversation, try to think about how you were feeling at the time.

Place your finger on the line below to indicate how much you feel that you agreed with the statement at the point in the conversation that you are watching. You can place your finger anywhere on the line, not just on one of the marks.

You may feel that you agreed with this statement more or less at different times during the conversation. Please move your finger whenever you feel that your opinion changed.

I trust the counselor

Disagree strongly

Agree strongly

Retrospective Trust Assessment

APPENDIX B

Coding Manual for Discourse Structure

This coding manual was used for the annotation of topic boundaries in the analysis of posture shifts (Chapter 5).

B.1 Introduction

This manual gives coding directions for identifying and *discourse segment boundaries* in the exercise counseling corpus: points at which a conversation can most naturally be divided into smaller discourse segments.

You will be using transcripts that have already been separated into phrases based on the occurrence of silence. You will mark each phrase — a phrase being here a sequence of words spoken by a participant without intervening silence — to indicate whether there is a discourse segment boundary at that phrase.

B.2 Identifying Segment Boundaries

We assume that most of the time (not necessarily all of the time) participants in conversation believe that they are mutually discussing a particular topic. A discourse segment boundary occurs whenever either participant attempts to change this mutual topic. You should mark a boundary wherever you

believe an attempt occurred, even if it was not successful (i.e., even if the other participant did not join in discussing the intended topic).

The topic they are attempting to introduce need not be new to the conversation. It may be a previously discussed topic. It may also be a sub-topic of the current topic or a previous topic, or it may be a return to the main topic, having discussed a sub-topic.

To indicate the exact location and nature of a boundary, use one of the following tags:

new A segment begins at the beginning of a phrase. The phrase is devoted to either discussing a new topic or to attempting to introduce it, and adds no substantial information to the old topic.

overlap A segment begins somewhere in the middle of the phrase. The phrase has both content that adds something to the previous topic (typically at the beginning), as well as content that is devoted to the new topic (typically at the end).

drift A boundary occurs somewhere near a segment, but the exact location is unclear. Typically, a participant has begun by discussing one topic, and gradually changed to a second topic, but it is difficult to identify an exact point at which the topic changed.

Phrases often begin with acknowledgments (e.g. “okay”) or markers (e.g. “so...”) that, by themselves, add little information. A topic shift that begins with these but has no other content relating to the previous topic should be coded as “new” rather than “overlap”.

When “drift” is used, it should almost always be used on a several phrases in sequence. You should mark the shortest sequence of phrases that could possibly contain a segment boundary.

B.3 Examples

Here are two examples of phrases that should be coded as “new”. Note that although the beginning of the second example (“right but”) refers to a previous topic, it adds little to it, so this is still coded as “new” rather than “overlap”.

1 (counselor) let’s talk about exercise

1 (client) right but about the walking

In order to be coded as “overlap”, a phrase both introduce a new topic and add something to the old one. For example:

1 (client) I like the walking but about next week

B.4 General Suggestions

Discourse markers such as “so”, “well”, “now” often occur at topic shifts, and can be a hint that a phrase may be a segment boundary. Other hints are disfluencies or filled pauses (“um”, “uh”) and phrases that explicitly reference the topic of discussion (“let’s talk about...”). However, you should use these only as hints, and always decide whether a boundary is present based on the definitions above.

B.5 Difficult or Special Cases

B.5.1 Greetings

By convention, the first phrase of a conversation should always be marked as a segment boundary.

Most or all conversations will begin with some type of ritual greeting and social dialogue (e.g., “Hi, how are you?”); common exchanges that require little thought and little specific knowledge of the conversation partner. You should try and mark a discourse segment boundary at the point that less ritualized dialogue begins. Note that this may still be social dialogue, not focused on the specific counseling task.

B.5.2 Topic Shifts over Multiple Phrases

A topic may often be introduced over several phrases. You should always mark the segment boundary on the first phrase that contributes any substantial information about the new topic. In the following example, you should mark “new” on line 3:

- 1 (counselor) okay so
- 2 (counselor) um
- 3 (counselor) let’s talk about last week
- 4 (client) okay
- 5 (counselor) how did the walking go

B.5.3 Subtopics

In some cases, it may be difficult to decide whether a phrase introduces a sub-topic of the current topic, or merely continues the current topic. This rule may be useful: if there is a clear point later on at which the participant appears to go back to the larger topic — particularly if marked in any way — both points should be coded as segment boundaries. If you are unable to decide, you should prefer coding fewer segment boundaries rather than more.

B.5.4 Lengthy Narrative

Some conversations include lengthy narrative segments, during which one participant talks extensively about a topic while the other contributes no more than minimal acknowledgments (“okay”, “mmhmm”). These passages often include several sub-topics, and these may be among the most difficult to code reliably. The suggestions given above (under “General Suggestions” and “Subtopics”) may be useful. Lengthy narrative passages are also where the “drift” tag is most likely to be applicable.

B.5.5 Vague or Open-Ended Questions

One participant may introduce a topic only broadly, leaving it to the conversation partner to decide a more specific discussion topic. This may be done with an open-ended question, as in the following example:

- 1 (counselor) so
- 2 (counselor) um
- 3 (counselor) do you have any other questions
- 4 (client) um
- 5 (client) wearing the pedometer thing

A new topic has introduced, so at least one discourse segment boundary should be marked. However, there are two possible locations: on line 3, the counselor introduces an open-ended topic (unanswered questions), and on line 5, the client responds with the specific topic to be discussed (wearing the pedometer). Arguably, a boundary could be marked at line 3, at line 5, or both.

In most cases like this, try and mark only the phrase that first tries to begin a new topic — line 3 in this example — even if it does not completely

decide what that topic will be. This choice is somewhat arbitrary, but will limit the number of extra boundaries marked, and may increase reliability.

APPENDIX C

Coding Manual for Posture Shifts

This coding manual was used for the annotation of posture shifts in the analysis of that behavior (Chapter 5).

C.1 Introduction

This manual gives directions for identifying and coding posture shifts in the exercise counseling corpus. The corpus includes 32 conversations between a behavior change counselor and several different clients. Each conversation takes place in the same setting, with a similar arrangement: both client and counselor are seated in swiveling chairs, facing each other at a distance of about a couple feet. You will code posture shifts based on silent video of the conversations, recorded simultaneously from 3 different cameras: front and side views of the counselor, and a front view of the client, all focused mainly on the waist up.

The primary goal of the coding effort is to mark every movement, made by either counselor or client, that is identifiable as a posture shift according to the definitions given below. For each shift, you will also describe the movement by coding several features:

1. Timing: when does the shift begin and end?

2. Direction and movement pattern: what body parts move, and in which direction(s)?
3. Energy: How much effort is expended in the movement?
4. Co-occurrences: Are other movements performed simultaneously with the shift?

C.2 Identifying Posture Shifts

A posture shift is a movement during conversation (whether a participant is speaking or silent) that meets *all* of the following conditions:

1. It changes the overall (gross) position of the body, including at least the trunk, legs and lower body, or both. Movement of the head, shoulders, hands, or feet may be included, but by themselves they are not sufficient.
2. It is a distinct action, rather than being part of a continuous or repeated movement. It should be fairly easy to identify start and end points.
3. It is not caused by the performance of either:
 - a) A movement that appears to be consciously intended to communicate to the conversation partner, such as a head nod or an emphatic hand gesture, or
 - b) A purposeful physical action, such as leaning down to tie a shoe, or getting up out of the chair.

C.3 Coding Features of Shifts

C.3.1 Timing

Mark the times you believe a posture shift begins and ends with good accuracy. It is not necessary to go frame by frame to identify an exact position, but try

to be within a quarter second (7–8 frames) of where you believe the shift begins or ends.

C.3.2 Movement Pattern

Some shifts are simple movements, while others are more complex, and are made up of smaller sub-movements with changes of direction.

shift A simple movement, primarily in one direction, with no major changes of direction.

shift-and-return A movement in one direction, followed by a movement in the opposite direction. The two movements need not be the same energy or size; for example, a large movement forward and a smaller movement back should still be coded as a shift-and-return.

other Any other pattern of movement.

In some cases, a shift may contain some very small movements; You should consider smaller movements when coding movement pattern *only* if they would be large enough to count as posture shifts by themselves. For example, a large movement forward followed by a barely perceptible movement back should be coded as a shift, not shift-and-return.

C.3.3 Direction

forward/back Movement that brings the upper body closer to, or further away, respectively, from the conversation partner.

left/right Movement that brings the upper body to the participant's left or right (*note*: not the camera's left or right!), respectively.

lower-body Shifting of weight or other movement primarily involving the legs and lower body.

rotate Rotating in the chair.

other Any movement not described above.

In the case of a shift-and-return, code the direction of the shift, not the return. In the case of a movement pattern coded as “other”, if there is an obvious primary movement within it, give the direction of that movement, and if not, use “other”.

Some movements may appear to be in multiple directions simultaneously; for example, a participant may move forward and left, or back while rotating in the chair. If there is clearly a higher energy direction, code that. If not, use the order in the list above as a tie-breaker: prefer forward to left, and so on.

C.3.4 Energy

For each posture shift, estimate the total energy or effort expended by the participant to perform the movement. Fast movements are higher energy than slow movements, and large movements are higher energy than small movements.

Code energy with a linear scale from 1–10, where 1 is the least energetic movement you could identify, and 10 is the most energetic movement possible without the participant being forced to leave their chair.

If a posture shift includes several sub-movements (e.g. a shift-and-return rather than a simple shift), code the total energy; that is, the sum of all parts of the shift.

C.3.5 Co-occurrence

As above, movements similar to posture shifts may be caused by other actions, such as emphatic head nodding. In other cases, another noticeable action may occur at the same time as a posture shift, without causing it — that is, the

action and the posture shift could reasonably occur separately. If so, code the following common actions that occur simultaneously with posture shifts:

none No, or only small actions.

grooming Moving hair, straightening clothes, or similar actions.

laughter Visibly laughing or giggling.

other Any other large action.

C.4 General Suggestions

It will be easier (although more time consuming) to code the client and counselor separately, in two passes. When coding the counselor, be sure to keep an eye on both cameras, as some movements may be more apparent from the front view or the side view.

You are not required to code a conversation from start to finish with no breaks. It may be helpful to take regular short breaks in order to keep attention focused.

In some cases, it may be easier to notice changes in position than the movements themselves, particularly when other movements are occurring at the same time. If you notice that a participant's upper body position has changed noticeable, it is usually worth going back and looking to see if there was a posture shift.

C.5 Difficult or Special Cases

C.5.1 Small Movements

Very small movements — those that would be coded with an energy of 1, or sometimes 2 — can be very hard to code reliably, particularly when any other

action is occurring. Avoid placing too much effort on coding these movements, and focus on coding the larger ones reliably.

C.5.2 Repetitive Movement

In some cases, participants may perform repetitive motions during a conversation, which can last quite a long time — often a minute or more. The most common are repetitive back and forward motions, and repetitive swiveling in the chair. Repetitive motion should *not* be marked as a posture shift, since there is not distinct action (just continuous motion).

While long-running repetitive motion should be obvious, in some cases it may be difficult to decide whether movement is a posture shift or short-running repetitive motion. If a movement includes more than two repetitions (e.g. forward, back, forward, back, forward, back) or lasts more than a few seconds, it should normally be treated as repetitive motion, and not coded as a posture shift.

C.5.3 Co-occurring Actions

When a possible posture shift occurs at the same time as another action, it may be difficult to decide if the possible shift was caused by the co-occurring action, and thus should not be coded. The decision to make is: would it be possible for the other action to occur unmodified without the possible posture shift? While each possible shift should be examined carefully, here are some common examples:

- Possible shifts that co-occur with grooming behavior, or with laughing, usually *should* be coded.
- Possible shifts that co-occur with very emphatic head nodding, hand gestures, or shrugging, or with actions like picking up an object or tying a shoe, usually *should not* be coded.

C.5.4 Single versus Multiple Shifts

In some cases, it may appear possible to code a sequence of movements as either a single shift or as multiple shifts. For example, a forward movement followed by a backward movement could be coded as two shifts or a single shift-and-return.

You should prefer coding a single shift over two or more shifts. Only code multiple shifts if there is a clearly identifiable period of time between the end of one movement, and the beginning of the next, at least one second in duration.

C.5.5 Conversational Openings

In the exercise counseling corpus, every conversation begins with the counselor walking in and taking a seat. The counselor's movements should be coded as posture shifts (where appropriate) starting from the moment when she is first in a seated position in the chair. Note that sitting in a chair is often followed very closely by several movements. These are all potential posture shifts that can be coded; only the initial sitting motion is not.

The client will always begin the conversation seated. All client movements can potentially be coded as posture shifts from the beginning of the video. However, be sure to exclude movements such as reaching up to shake hands.

C.6 Changes from Earlier Drafts

There has been extensive rewriting and clarification from earlier drafts. In addition, the following substantive changes were made to the coding scheme:

- Lower-body movement (including chair movement) is now coded.
- The “other” movement pattern was added.
- The coding of movement direction was simplified, removing rotate left,

rotate right, raise left, raise right, raise, and drop, while adding lower-body and chair rotation codes.

- Coding of energy was simplified to use a 1–10 scale rather than 1–100.
- Coding of some co-occurring actions was added.

APPENDIX D

Coding Manual for Nonverbal Behavior in Openings

This coding manual was used for the annotation of nonverbal behavior in the analysis of conversation openings and reopenings (Chapter 6).

D.1 Introduction

This coding scheme is intended for an investigation of how nonverbal behavior varies over time in multi-conversation discourse, both across multiple conversations and within individual conversations. The specific focus during development of the scheme is a study of conversational openings, but the scheme may also be more generally useful.

This effort is intended to produce an accurate description of the form of (some of the) nonverbal behavior in these conversations. We code the timing, shape, extent, speed, and other physical features of behavior, but not the communicative function (e.g. semantic meaning, affect, turn-management features).

This manual is heavily based on the MUMIN coding scheme [2], which has been previously used for annotation of nonverbal behavior with good reliability on most behaviors. Relative to MUMIN, we omit all coding related to communicative function and simplify coding of facial expressions and hand

gestures. We add annotation of “resting” hand position, of “adapter” gestures that release bodily tension, and modify the annotation of posture.

D.2 Materials and Coding Task

The materials for this coding task are 32 one-minute video clips, each taken from the beginning of a conversation. Each video has two participants, the *counselor* and the *client*. One counselor appears in all videos; there are several clients. Each video has three (simultaneous) camera angles: a front and side view of the counselor, and a front view of the client.

Each conversation begins with the client seated, and the counselor entering the room. The clips have been chosen to begin as soon as the counselor is seated.

Coding will be performed twice for each behavior detailed below, separately for the counselor and the client. In the case of the counselor, all coding should be a single unified judgment based on both camera views.

All coding will be performed using the ANVIL tool [88], version 5.0.22 or higher¹. This tool allows for the coding of “events” on multiple “tracks”, where each event has a start time, an end time, and a set of features that varies for each track. For this task there will be two tracks for each behavior category below: one for the counselor, and one for the client.

D.3 Visibility

As participants do not assume precisely the same position in all videos, some features of the participants’ bodies may not be clearly visible, rendering some annotation difficult or impossible. The most frequently obstructed are the hands (off the bottom of the frame) and the eyes (by shadows and/or glasses).

¹The current version is available at <http://anvil-software.de>

Before coding any nonverbal behavior, view a video and note the visibility of the eyes and hands separately for each participant. If the hands or eyes are coded as “obstructed”, you may skip coding of hand gestures or gaze direction, respectively, for that participant.

visible The participant’s hands/eyes are entirely or mostly visible throughout the video, and easily coded.

partially-visible The participant’s hands/eyes are partially obstructed or otherwise difficult to code. The reliability of coding may be reduced.

obstructed The participant’s hands/eyes are entirely or mostly obstructed or otherwise difficult to code. Coding accurately will be very difficult or impossible.

Any major obstruction affecting parts of the body other than the hands or eyes should be noted in a comment.

D.4 Nonverbal Behavior

Most nonverbal behaviors are coded in terms of *movement* rather than *position*: you need only annotate when a participant either is moving, or shifts out of a “neutral” position (as defined for each type of behavior). The exception is arm openness, which should be annotated continuously (i.e. there should be a position coded at all times in the video).

In ANVIL, you should create events for all time periods at which a participant moves out of “neutral”, except for arm openness, where the events should cover the entire track. These tracks may have only a single event, in the case where a participant does not change their position during the video.

Many behaviors include an “other” category. These should be used only when none of the defined categories can be used, and a comment should be added to the annotation.

D.4.1 Gaze Direction

Annotate the video to mark segments where a participant is *not* looking toward the conversation partner. All directions are relative to the conversation partner; e.g., if a participant's eyes are away from the partner and up relative to the partner, this should be coded as "up" whether or not the participant's head is also tilted up.

up Away from the conversation partner, and primarily upwards: at more than a 45 degree angle from horizontal.

down Away from the conversation partner, and primarily downwards: at more than a 45 degree angle from horizontal.

side Away from the conversation partner, and primarily sideways (either left or right): at less than a 45 degree angle from horizontal.

other The participant is gazing away from the conversation partner, but the direction is unclear and/or does not match one of the above categories.

D.4.2 Eyebrows

Annotate the video to mark segments where a participants eyes are either raised or lowered relative to their neutral position.

raise The brows are moved up toward the forehead (AU 1 and/or 2).

frown The brows are contracted and moved down toward the nose (AU 4).

other The brows have moved from a neutral position, but in some arrangement other than "raise" or "frown". Use this category if only one eyebrow is raised/lowered.

Small and brief eyebrow raises can be easily missed. Often, these co-occur with a widening of the eyes and/or with abrupt head movements, often a tilt,

jerk, or turn (as defined below). You may want to pay extra attention to the eyebrows during such events.

D.4.3 Head Movement

Annotate the video to mark segments where a participant moves their head, either with a movement of the whole trunk or separately. All head movements should be marked with two features: the type of movement, and whether it is a single or repeated movement.

Very small movements will often occur as the participant's head balances on his or her shoulders. You should try and avoid coding these small movements. The goal is to mark movements that are noticeable to the conversation partner and may be taken to have meaning, and/or movements that substantially change the position of the head relative to the conversation partner.

As a general rule, you should code a movement if *either* (or both) of the following conditions holds, and not otherwise:

1. Any part of the head moves by two inches or more.
2. The movement is part of a clearly identifiable, repetitive gesture (e.g., nodding or shaking).

Note that participants may commonly nod using small movements, particularly as a “backchannel” when the conversation partner is talking. Following the second condition given here, you should try and code these movements whenever possible, even if the movements would be too small to code otherwise.

nod Movement of the head down-and-up.

jerk A single quick head movement up, sometimes with a movement back down. If this movement is repeated (see below), it should generally be coded as “nod” instead.

back Movement of the head backwards and/or up, either by itself or as part of a movement of the whole trunk.

forward Movement of the head forward and/or down, either by itself or as part of a movement of the whole trunk.

turn Movement that rotates the head to either side, or both.

tilt Leaning the head to either side. The head may also turn somewhat; use this category in that case rather than “turn”.

other Movement that does not fit any of the above categories. This includes cases where a participant appears to perform two movements at once; for example nodding overlaid on a backward or forward movement.

The same type of head movement is often repeated without pause; for example, a participant may continue nodding for several seconds, or may “turn” repeatedly (i.e. shaking the head to indicate disagreement). Code such repeated movement as a single event, starting at the beginning of the initial movement and continuing until there is a noticeable pause in movement (at least a half second) or the participant performs a different category of head movement.

single A single movement is performed, or a movement in one direction followed immediately by a return (e.g., moving the head down and back up should be coded as “nod” and “single”).

repeated The same movement is repeated two or more times without pause.

D.4.4 Mouth Shape

Annotate segments of video whenever the lips are not in a neutral shape. A neutral shape is the shape the lips take when the face is relaxed, not smiling or frowning, or displaying any other recognizable emotion. Note that mouth

shapes *should* still be annotated when a participant is speaking or otherwise has his or her mouth open.

corners-up Corners turned up, as in a smile (AU 12 and/or 13).

corners-down Corners turned down, as in a frown (AU 15, possibly with 17).

protruded Lips rounded and protruded.

retracted Lips sucked into the mouth.

other A mouth shape that does not fit into any of the above categories. This includes asymmetrical shapes (e.g., turning up one corner only).

A few cases may be more difficult:

1. Mouth shapes that a participant has taken at the beginning of the video are easily missed: for example, if a participant is already smiling when the video begins. These may be easier to notice when the expression ends and the mouth returns to a neutral position.
2. Some categories, particularly “retracted”, are unlikely to occur during speech.
3. Smaller and less emphatic smiling (“corners-up”) can be missed, particularly during speech. It may be easier to first notice movement of the cheeks, possibly with dimpling.

D.4.5 Hand Gesture

Annotate segments of video where the participant gestures with his or her hands. Gestures should be annotated whether they appear to be intentional or unintentional (e.g., adapters). Very small movements (e.g., twitching fingers) need not be annotated.

D.4.5.1 Semiotic Category

Semiotic category refers to the meaning or function of a hand gesture, although particular categories may have characteristic shapes. There are five categories given here, which can be placed into two broader groups. First, there are gestures that refer to or represent something the participants are talking about (“deictic”, “iconic”, and “emblematic”), as a pronoun like “it” might refer to something. These categories are distinguished based on *how* the gesture identifies what it is referring to. Second, there are gestures that do not refer to anything (“beat” and “adapter”).

Although each gesture should be considered individually, we note that “beat” and “adapter” are usually the most common categories. Of the remaining categories, “iconic” is the most general, and should usually be used unless a gesture is clearly a “deictic” or “emblematic”.

deictic Represents something the participants are talking about by indicating its physical location, often by pointing. This includes pointing or other gestures (e.g. tapping fingers on the chest) which a participant uses to emphasize that they are referring to themselves or the conversation partner.

iconic Represents something the participants are talking about by similarity in hand shape, trajectory, or speed. Examples include gestures that indicate the size of an object. Metaphoric gestures are included in this category (i.e., gestures that represent an abstract feature of something rather than a physical feature). Also included are gestures that indicate some ordering of things (first, second, next, previous, etc.).

emblematic Gestures that have meaning based on social convention (e.g., thumbs up).

beat Rhythmic, simple, baton-like movements that tend not to vary in shape along with the content of speech. Beat gestures do not appear to refer to

anything in the discourse; rather, they mark timing and emphasis. Often, a participant will leave his or her hand position wherever it happens to be, and just add a small movement. Multiple short beat gestures in quick succession are common.

adapter Gestures that appear to have no conscious communicative function, and act to relieve bodily tension. Examples include wringing hands, and grooming behaviors (e.g. running fingers through hair).

other A gesture that does not fit any of the above categories, or appears to fit into two or more categories.

D.4.5.2 Handedness

Gestures may be performed with the left hand, the right, or both. In the case of a single hand, we do not currently annotate which hand is used.

single-hand Either the left or right hand is used.

both-hands Both the left and the right hands are used. The hands need not perform identical or symmetrical movements.

D.4.5.3 Spatial Extent

Gestures may be “broad”, taking up a large amount of space, or narrow. The spatial extent of a gesture is defined in terms of the participant’s “center”, which is the region in front of the participant’s torso (at or above the stomach, and not over the head) easily reachable without extending the arms. Note that gestures at the face (common with “adapters”) are considered within the center.

minimal The gesture remains within the center, and involves only minimal movement, typically only moving the hands with very little movement of the arms.

center The gesture remains within the center, with more movement than “minimal”.

periphery The gesture extends outside the center to either side of the participant, or above his or her face.

forward The gesture extends outside the center in front of the participant toward the conversation partner.

D.4.6 Shoulder Raising

Mark all segments of the video in which a participant shrugs or otherwise temporarily raises either one shoulder or both shoulders. For the purposes of this annotation, “raising” is moving the shoulder(s) up above whatever position a participant assumes for the majority of the video. For example, in some videos a participant may lean heavily on the chair, with his or her shoulders fairly far down; when raised, the shoulders may still be lower than in other videos.

both Both the left and right shoulders are raised.

single Either the left shoulder or the right shoulder alone is raised.

D.4.7 Arm Openness

Arms can be either closed or opened. A closed position is one that generally moves the arms and/or hands to a participant’s front, between the participant and the partner, while open positions move the arms and/or hands out of this area.

Arm openness is coded here by looking at the lateral (left–right) position of the upper arms and elbows. The arms of the chairs are used as a reference point. The shoulder position may also change along with the upper arms, and can serve as an alternate cue.

Note that arm openness, unlike all other behaviors in this manual, should be annotated continuously: that is, the segments you mark should cover the entire video. In some cases, you may code only a single segment for a participant who never changes his or her arm position during a video, while in others, multiple segments should be coded.

shifting The participant is in the process of moving, and the open/closed position of the arms is unclear or changing.

closed Both arms are farther in than the arms of the chair. Arms are often touching the torso, and hands may be joined.

half-open One arm is positioned as in “closed”, and one as in “open”.

open Both arms are either resting on the chair arms, or lined up with them. Hands may or may not be joined.

half-back One hand positioned as in either “closed” or “open”, and the other as in “back”.

back Both arms are farther out than the arms of the chair. The arms may sometimes be hanging down to the side, or behind a participant’s head.

APPENDIX E

Experimental Protocol for the Evaluation Study

The evaluation study was almost entirely automated, with contact between researchers and participants (via email) occurring only in the case of technical problems or other unanticipated difficulties. Recruitment materials directed potential participants to a website where all experiment sessions, including intake, were performed. The system also contacted participants automatically via email. The following sections detail the behavior of the website and the automated email.

E.1 Experiment Sessions

E.1.1 Intake Session

When a potential participant navigated to the **Front Page** and clicked on “Get Started”, the system performed the following steps:

1. Present **Screening Questionnaire**.
 - a) If the participant is under 18 years of age (question 1), present **Ineligibility Response: Age** and exit.
 - b) If the participant is not a native English speaker (question 2), present **Ineligibility Response: Language** and exit.

- c) If the participant does not have computer access (question 3), present **Ineligibility Response: Computer Access**, and exit.
 - d) If the participant has a health issue which increases the risks of physical activity (questions 4–10), present **Ineligibility Response: Physical Activity Readiness**, and exit.
 - e) If the participant is in the Action or Maintenance stages of change (question 11, first or second response), present **Ineligibility Response: Stage of Change** and exit.
2. Present **Consent Form**. If the participant indicates consent (checking “I agree to participate” and selecting “Yes, continue”), then proceed, otherwise exit.
 3. Present **Sign-up Form**. If a participant has previously signed-up using the same email address, present **Error: Already Signed Up**, otherwise proceed.
 4. Record the participant as signed-up, randomize the participant into a study condition, and send the **Sign-up Email**.
 5. The participant navigates to the link provided in the **Sign-up Email**. Record the participant as having started the study; the current date is week 1 of the study.
 6. Present the **Demographics** questionnaire.
 7. Present the **Pre-Intervention Decisional Balance** questionnaire.
 8. Proceed to the first session, as below.

E.1.2 Weekly Sessions

A weekly session was held either immediately following intake steps (for the first week), or when a participant navigated to the **Front Page** and logged in (in the second through sixth weeks):

1. If a weekly conversation has not been held, present the virtual counselor.
2. If not yet collected for the current week, present the **Nonverbal Immediacy (SRIB)** questionnaire.
3. If not yet collected for the current week, present the **Therapeutic Alliance (WAI-SR)** questionnaire.
4. If not yet collected for the current week, present the **Behavioral Realism** questionnaire.
5. If the participant is in the first through fifth study weeks, present **Session Completion** and exit.
6. If the participant is in the sixth study week, then:
 - a) If not yet collected, present the **Post-Intervention Assessment** questionnaire.
 - b) Present **Final Session Completion**.

E.1.3 Post-Intervention Sessions

For a period of two weeks after the sixth weekly session, if a participant logged in, the system performed the following steps:

1. If not yet collected, present the **Post-Intervention Assessment** questionnaire.
2. If a conversation has not been held during the current day, present the virtual counselor.
3. Present **Final Session Completion**.

If a participant attempted to log in following this two week period, the system presented **End of Participation**.

E.2 Email Messaging

Participants were considered *withdrawn* and did not receive most email messages if either:

- they followed the link provided in all email reminders (*explicitly withdrawn*), or
- they did not log in for two consecutive weeks during the six-week intervention period (*implicitly withdrawn*).

At the beginning of each week (approximately midnight Monday), the system sent the following messages:

- From the set of participants who had a weekly conversation in the previous week (including the initial session but not including post-intervention sessions), the system selected one at random and delivered the **Drawing Notification Email**.
- For each participant in the second through sixth weeks of the study who was not withdrawn, the system delivered the **Weekly Reminder Email**.
- For each participant in the two week period following the final weekly session who was not explicitly withdrawn and had not completed the Post-Intervention Assessment questionnaire, the system delivered the **Final Assessment Email**.
- For each participant who had just completed the two week period following the final weekly session (including participants who had withdrawn, implicitly or explicitly), the system delivered the **Debriefing Email**.

Two days prior to the end of each week (approximately midnight Saturday), for each participant in the second through sixth weeks of the study who was not withdrawn and had not yet had a weekly conversation, the system delivered the **Additional Reminder Email**.

Volunteers needed for an online research study

Researchers at Northeastern University are building animated computer-controlled characters that act as counselors to help people improve their health.

We are looking for volunteers to help test a physical activity counselor that you can talk to in your web browser. Volunteers will briefly talk to the counselor once a week for six weeks. All sessions can be done online at any time that is convenient to you.

You will be entered in a drawing with a chance to win a \$50 gift card to Amazon.com each week you take part in the study.

You must be 18 or older, and a native English speaker in order to participate. You will need regular access to a computer with a high-speed (not dial-up) internet connection and the ability to play audio with speakers or headphones. We are looking for volunteers who are in good health and are **not** currently exercising regularly.

If you are interested, go to: <http://wonder.ccs.neu.edu/counselor-study>

Recruitment Material

Volunteers Needed for an Online Research Study

Researchers at Northeastern University are building animated computer-controlled characters that act as counselors to help people improve their health.

We are looking for volunteers to help test a physical activity counselor that you can talk to in your web browser. Volunteers will briefly talk to the counselor once a week for six weeks. All sessions can be done online at any time that is convenient to you.

You will be entered in a drawing with a chance to win a \$50 Amazon.com gift card each week you participate in the study.

You must be 18 or older, and a native English speaker in order to participate. You will need regular access to a computer with a high-speed (not dial-up) internet connection and the ability to play audio with speakers or headphones. We are looking for volunteers who are in good health and are **not** currently exercising regularly.

If you are interested, go to:

<http://wonder.ccs.neu.edu/counselor-study>



Online counselor study http:// wonder.ccs.neu.edu/	Online counselor study http:// wonder.ccs.neu.edu/	Online counselor study http:// wonder.ccs.neu.edu/	Online counselor study http:// wonder.ccs.neu.edu/	Online counselor study http:// wonder.ccs.neu.edu/	Online counselor study http:// wonder.ccs.neu.edu/	Online counselor study http:// wonder.ccs.neu.edu/	Online counselor study http:// wonder.ccs.neu.edu/	Online counselor study http:// wonder.ccs.neu.edu/	Online counselor study http:// wonder.ccs.neu.edu/
--	--	--	--	--	--	--	--	--	--

Recruitment Material

Longitudinal Evaluation of a Web-Based Physical Activity Counselor

If you are currently participating in this study, you can log in here:

Email:

Password:

[Log In](#)

If not, are you interested in participating?

Researchers at Northeastern University are building animated computer characters that act as counselors to help people improve their health.

We are looking for volunteers to help test a physical activity counselor that you can talk to in your web browser. Volunteers will briefly talk to the counselor once a week for six weeks. All sessions can be done online at any time that is convenient to you.

You will be entered in a drawing with a chance to win a \$50 gift card to Amazon.com each week you take part in the study.

You must be 18 or older, and a native English speaker in order to participate. You will need regular access to a computer with a high-speed (not dial-up) internet connection and the ability to play audio with speakers or headphones. We are looking for volunteers who are in good health and are **not** currently exercising regularly.

[Get Started](#)

Front Page (web form)

Thank you for your interest in this study.

Before we can begin, we have a few questions to check whether you are eligible to participate:

Are you at least 18 years old?

Yes	No
-----	----

Are you a native English speaker?

Yes	No
-----	----

Do you have regular access to a computer with a high-speed (not dial-up) internet connection and the ability to listen to audio through speakers or headphones?

Yes	No
-----	----

Has your doctor ever said that you have a heart condition and that you should only do physical activity recommended by a doctor?

Yes	No
-----	----

Do you feel pain in your chest when you do physical activity?

Yes	No
-----	----

In the past month, have you had chest pain when you were not doing physical activity?

Yes	No
-----	----

Do you lose your balance because of dizziness, or do you ever lose consciousness?

Yes	No
-----	----

Do you have a bone or joint problem that could be made worse by a change in your physical activity?

Yes	No
-----	----

Is your doctor currently prescribing drugs (for example, water pills) for your blood pressure or heart condition?

Yes	No
-----	----

Do you know of any other reason why you should not do physical activity?

Screening Questionnaire (web form)

Yes	No
-----	----

Regular exercise is any *planned* physical activity (e.g., brisk walking, aerobics, jogging, bicycling, swimming, rowing, etc.) performed to increase physical fitness. Such activity should be performed *3 to 5 times* per week for *20-60 minutes* per session. Exercise does not have to be painful to be effective but should be done at a level that increases your breathing rate and causes you to break a sweat.

Do you exercise regularly according to that definition? (Please select one):

Yes, I have been for MORE than 6 months
Yes, I have been for LESS than 6 months
No, but I intend to in the next 30 days
No, but I intend to in the next 6 months
No, and I do NOT intend to in the next 6 months

Continue

Screening Questionnaire (web form)

We are sorry, but you are not eligible to participate in this study. We can accept only participants who are least 18 years old.

Thank you for your time, and please feel free to mention this study to anyone you know who may be interested in participating.

If you have any further questions about this study, please feel free to contact Daniel Schulman, phone: 617-373-4605, email: schulman@ccs.neu.edu, the person mainly responsible for the research. You can also contact Timothy Bickmore, phone: 617-373-5477, email: bickmore@ccs.neu.edu, the Principal Investigator.

Ineligibility Response: Age (web form)

We are sorry, but you are not eligible to participate in this study. We can accept only participants who are native English speakers.

Thank you for your time, and please feel free to mention this study to anyone you know who may be interested in participating.

If you have any further questions about this study, please feel free to contact Daniel Schulman, phone: 617-373-4605, email: schulman@ccs.neu.edu, the person mainly responsible for the research. You can also contact Timothy Bickmore, phone: 617-373-5477, email: bickmore@ccs.neu.edu, the Principal Investigator.

Ineligibility Response: Language (web form)

We are sorry, but you are not eligible to participate in this study. We can accept only participants who have regular access to a computer with a suitable internet connection and the ability to play audio.

Thank you for your time, and please feel free to mention this study to anyone you know who may be interested in participating.

If you have any further questions about this study, please feel free to contact Daniel Schulman, phone: 617-373-4605, email: schulman@ccs.neu.edu, the person mainly responsible for the research. You can also contact Timothy Bickmore, phone: 617-373-5477, email: bickmore@ccs.neu.edu, the Principal Investigator.

Ineligibility Response: Computer Access (web form)

We are sorry, but you are not eligible to participate in this study. Your answers indicate that you may have a medical condition which would increase the risks of beginning an exercise or physical activity program.

Thank you for your time, and please feel free to mention this study to anyone you know who may be interested in participating.

If you have any further questions about this study, please feel free to contact Daniel Schulman, phone: 617-373-4605, email: schulman@ccs.neu.edu, the person mainly responsible for the research. You can also contact Timothy Bickmore, phone: 617-373-5477, email: bickmore@ccs.neu.edu, the Principal Investigator.

**Ineligibility Response: Physical Activity Readiness
(web form)**

We are sorry, but you are not eligible to participate in this study. We are currently looking for participants who do **not** exercise at least three times a week for at least twenty minutes a session.

Thank you for your time, and please feel free to mention this study to anyone you know who may be interested in participating.

If you have any further questions about this study, please feel free to contact Daniel Schulman, phone: 617-373-4605, email: schulman@ccs.neu.edu, the person mainly responsible for the research. You can also contact Timothy Bickmore, phone: 617-373-5477, email: bickmore@ccs.neu.edu, the Principal Investigator.

Ineligibility Response: Stage of Change (web form)

Northeastern University, College of Computer and Information Science
Name of Investigator(s): Timothy Bickmore, Daniel Schulman
Title of Project: Longitudinal Evaluation of a Web-Based Virtual Exercise Counselor

Request to Participate in Research

We would like to invite you to participate in a web-based online research study. The purpose of this study is to evaluate how people respond to an animated computer character designed to act as a counselor which discusses exercise and physical activity with users. The computer character is automatically controlled and is not monitored by the research staff or any person. If you choose to participate, you will have six weekly conversations with this animated counselor, and answer questions about your attitudes toward the conversation, the counselor, and physical activity. Each weekly session should take about 10-15 minutes, and can be done at any time you like during a week.

We are asking you to participate in this study because you are an English-speaking adult, in good health, and not currently engaging in regular physical activity. You must have regular access to a computer capable of using the animated counselor, including high-speed internet and the ability to play audio. **You must be at least 18 years old to participate in this study.**

The decision to participate in this research project is voluntary. You do not have to participate and you can refuse to answer any question. Even after you begin an online session, you can stop at any time.

The possible risks or discomforts of the study are minimal. You may feel a little uncomfortable answering some questions. If you choose to increase your physical activity, you may have a risk of injury. You will not be asked to perform any physical activity more strenuous than brisk walking, and are not required to do any physical activity in order to participate.

No special arrangements will be made for compensation or for payment for treatment solely because of your participation in this research.

You will receive no direct benefit from participating in this study. Your participation may help us learn about how to build computer systems that help people get more exercise.

As a token of our appreciation, you will be entered in a drawing for a \$50 gift card to Amazon.com each time you complete a weekly session. The odds of winning will vary depending on the number of participants who complete a session in a given week. If you win, you will receive your gift card via email. If you win in one week, you will remain eligible for drawings in later weeks.

Your part in this study is anonymous to the researchers. Your responses to survey questions are anonymous to the researchers. However, because of the nature of web based surveys, it is possible that respondents could be identified by the IP address or other electronic record associated with the response. Neither the researcher nor anyone involved with this survey will be capturing those data. Any reports or publications based on this research will use only group data and will not identify you or any individual as being affiliated with this project.

Consent Form (web form)

If you have any questions regarding electronic privacy, please feel free to contact Mark Nardone, IT Security Analyst via phone at 617-373-7901, or via email at privacy@neu.edu.

If you have any questions about this study, please feel free to contact Daniel Schulman, phone: 617-373-4605, email: schulman@ccs.neu.edu, the person mainly responsible for the research. You can also contact Timothy Bickmore, phone: 617-373-5477, email: bickmore@ccs.neu.edu, the Principal Investigator.

If you have any questions regarding your rights as a research participant, please contact Nan C. Regina, Director, Human Subject Research Protection, 960 Renaissance Park, Northeastern University, Boston, MA 02115. Tel: 617-373-7570, Email: irb@neu.edu. You may call anonymously if you wish.

By clicking on the button below you are indicating that you consent to participate in this study. Please print out a copy of this consent form for your records.

☐ I agree to participate in this study.

No, thanks.

Yes, continue!

Consent Form (web form)

Thank you!

To get started, please give your first name (or whatever you would like the counselor to call you), a valid email address, and a password which you will use to log in and talk to the counselor.

We'll send you a confirmation email at the address you provide. When you respond to that email, you'll get started with the first weekly session.

First Name:

Email:

Password:

Retype Password:

[Continue](#)

Sign-up Form (web form)

Longitudinal Evaluation of a Web-Based Physical Activity Counselor

According to our records, you have already signed up for this study using the email address you provided. You can log in here:

Email:
Password:

[Log In](#)

Error: Already Signed Up (web form)

A confirmation email has been sent. To continue, please verify that the email address is valid by responding to that message.

Sign-Up Completion (web form)

Page 1 of 2

Please take a moment and answer a few questions about yourself.

Year of Birth:

Sex

Male

Female

Ethnic background

American Indian or Alaskan Native

Asian or Pacific Islander

Black, Not of Hispanic Origin

White, Not of Hispanic Origin

Hispanic

Other

Marital Status

Single

Married

Divorced or Widowed

Other

Last grade of school completed

Less than high school (0-8)

Some high school

High school graduate or GED

Technical school education

Demographics (web form)

Some college
College graduate
Advanced degree

Continue

Demographics (web form)

Page 2 of 2

The statements below look at positive and negative aspects of exercise. Read the following items and write a number next to each statement to indicate how important each statement is with respect to your decision to exercise or not to exercise in your leisure time.

If you disagree with a statement and are unsure how to answer, that statement is probably not important to you.

I would have more energy for my family and friends if I exercised regularly.

Not important	A little bit important	Somewhat important	Quite important	Extremely important
---------------	------------------------	--------------------	-----------------	---------------------

I would feel embarrassed if people saw me exercising.

Not important	A little bit important	Somewhat important	Quite important	Extremely important
---------------	------------------------	--------------------	-----------------	---------------------

I would feel less stressed if I exercised regularly.

Not important	A little bit important	Somewhat important	Quite important	Extremely important
---------------	------------------------	--------------------	-----------------	---------------------

Exercise prevents me from spending time with my friends.

Not important	A little bit important	Somewhat important	Quite important	Extremely important
---------------	------------------------	--------------------	-----------------	---------------------

Exercising puts me in a better mood for the rest of the day.

Not important	A little bit important	Somewhat important	Quite important	Extremely important
---------------	------------------------	--------------------	-----------------	---------------------

I feel uncomfortable or embarrassed in exercise clothes.

Not important	A little bit important	Somewhat important	Quite important	Extremely important
---------------	------------------------	--------------------	-----------------	---------------------

I would feel more comfortable with my body if I exercised regularly.

Not important	A little bit important	Somewhat important	Quite important	Extremely important
---------------	------------------------	--------------------	-----------------	---------------------

There is too much I would have to learn to exercise.

Not important	A little bit important	Somewhat important	Quite important	Extremely important
---------------	------------------------	--------------------	-----------------	---------------------

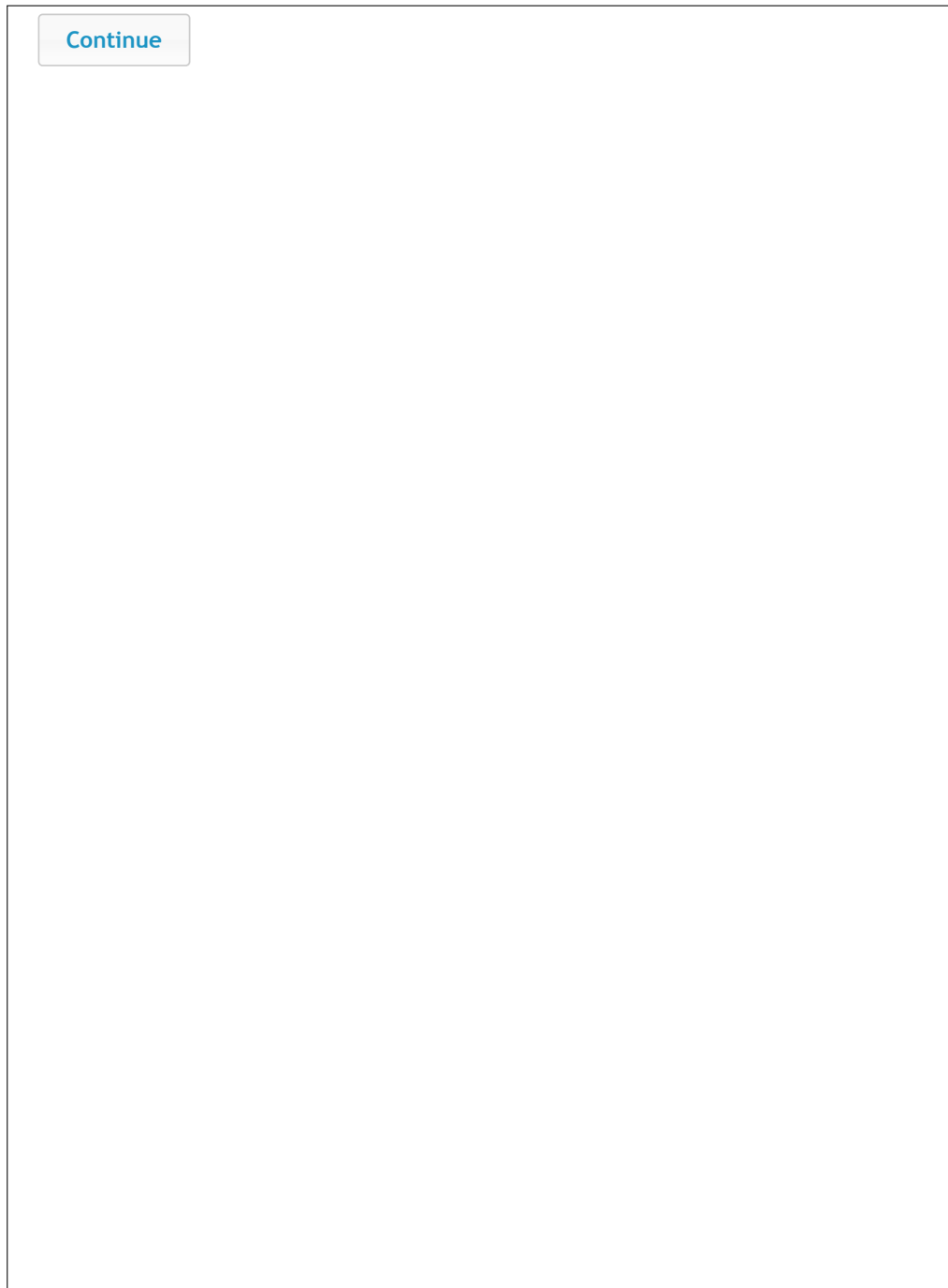
Regular exercise would help me have a more positive outlook on life.

Not important	A little bit important	Somewhat important	Quite important	Extremely important
---------------	------------------------	--------------------	-----------------	---------------------

Exercise puts an extra burden on my significant other.

Not important	A little bit important	Somewhat important	Quite important	Extremely important
---------------	------------------------	--------------------	-----------------	---------------------

Pre-Intervention Decisional Balance (web form)



Continue

Pre-Intervention Decisional Balance (web form)

Page 1 of 3

Below are a series of statements that describe the ways some people behave while talking with or to others. For each statement, choose the answer that most closely describes the agent's behavior when talking to you.

Karen uses a monotone or dull voice while talking to people.

Never	Rarely	Occasionally	Often	Very Often
-------	--------	--------------	-------	------------

Karen looks at people while talking to them.

Never	Rarely	Occasionally	Often	Very Often
-------	--------	--------------	-------	------------

Karen frowns while talking to people.

Never	Rarely	Occasionally	Often	Very Often
-------	--------	--------------	-------	------------

Karen has a very tense body posture while talking to people.

Never	Rarely	Occasionally	Often	Very Often
-------	--------	--------------	-------	------------

Karen moves away from people while talking to them.

Never	Rarely	Occasionally	Often	Very Often
-------	--------	--------------	-------	------------

Karen uses a variety of vocal expressions while talking to people.

Never	Rarely	Occasionally	Often	Very Often
-------	--------	--------------	-------	------------

Karen smiles while talking to people.

Never	Rarely	Occasionally	Often	Very Often
-------	--------	--------------	-------	------------

Karen looks away from people while talking to them.

Never	Rarely	Occasionally	Often	Very Often
-------	--------	--------------	-------	------------

Karen has a relaxed body posture while talking to people.

Never	Rarely	Occasionally	Often	Very Often
-------	--------	--------------	-------	------------

Karen is "stiff" while talking to people.

Nonverbal Immediacy (SRIB) (web form)

Never	Rarely	Occasionally	Often	Very Often
-------	--------	--------------	-------	------------

Karen moves close to people while talking to them.

Never	Rarely	Occasionally	Often	Very Often
-------	--------	--------------	-------	------------

Karen is expressive while talking to people.

Never	Rarely	Occasionally	Often	Very Often
-------	--------	--------------	-------	------------

Karen looks bland or neutral when talking to people.

Never	Rarely	Occasionally	Often	Very Often
-------	--------	--------------	-------	------------

Continue

Nonverbal Immediacy (SRIB) (web form)

Page 2 of 3

Below is a series of statements about experiences people might have with their counseling or counselor. For each statement, please take your time to consider your own experience and then respond by selecting one of the choices below it.

As a result of these sessions, I am clearer as to how I might be able to change.

Seldom

Sometimes

Fairly often

Very often

Always

What I am doing in counseling gives me a new way of looking at my problems.

Seldom

Sometimes

Fairly often

Very often

Always

I believe Karen likes me.

Seldom

Sometimes

Fairly often

Very often

Always

Karen and I collaborate on setting goals for my counseling.

Seldom

Sometimes

Fairly often

Very often

Always

Karen and I respect each other.

Seldom

Sometimes

Fairly often

Very often

Always

Karen and I are working toward mutually agreed-upon goals.

Seldom

Sometimes

Fairly often

Very often

Always

I feel that Karen appreciates me.

Seldom

Sometimes

Fairly often

Very often

Always

We agree on what is important for me to work on

Seldom

Sometimes

Fairly often

Very often

Always

I feel Karen cares about me even when I do things she does not approve of.

Seldom

Sometimes

Fairly often

Very often

Always

I feel that the things I do in counseling will help me to accomplish the changes that I want.

Therapeutic Alliance (WAI-SR) (web form)

Seldom	Sometimes	Fairly often	Very often	Always
--------	-----------	--------------	------------	--------

We have established a good understanding of the kind of changes that would be good for me.

Seldom	Sometimes	Fairly often	Very often	Always
--------	-----------	--------------	------------	--------

I believe the way we are working with my problems is correct.

Seldom	Sometimes	Fairly often	Very often	Always
--------	-----------	--------------	------------	--------

[Continue](#)

Therapeutic Alliance (WAI-SR) (web form)

Page 3 of 3

Please think about the counselor's behavior during your conversation, and answer whether you agree or disagree with each of the following statements.

The virtual counselor acted like a real person.

Strongly disagree

Disagree

Neither agree nor disagree

Agree

Strongly agree

The virtual counselor moved like a real person.

Strongly disagree

Disagree

Neither agree nor disagree

Agree

Strongly agree

I felt like I was interacting with a real person.

Strongly disagree

Disagree

Neither agree nor disagree

Agree

Strongly agree

For each of the following statements, please think about whether you feel that the counselor's behavior in your conversation felt compared to how a person might behave in a similar situation.

The counselor looks at me....

Much too little

Too little

About the right amount

Too much

Much too much

The counselor changes her body posture...

Much too little

Too little

About the right amount

Too much

Much too much

The counselor smiles at me...

Much too little

Too little

About the right amount

Too much

Much too much

The counselor nods her head...

Much too little

Too little

About the right amount

Too much

Much too much

The counselor's speech is...

Much too slow

Too slow

About right

Too fast

Much too fast

Continue

Behavioral Realism (web form)

Longitudinal Evaluation of a Web-Based Physical Activity Counselor

Thanks! You've completed everything we need for the study for now, and please come back next week! You'll be able to talk to the counselor again next _____.

Do you have any questions or comments? Please feel free to tell us what you think about the counselor or the study.

Send Comment

Email:

Password:

Log In

Session Completion (web form)

Page 4 of 4

The statements below look at positive and negative aspects of exercise. Read the following items and write a number next to each statement to indicate how important each statement is with respect to your decision to exercise or not to exercise in your leisure time.

If you disagree with a statement and are unsure how to answer, that statement is probably not important to you.

I would have more energy for my family and friends if I exercised regularly.

Not important	A little bit important	Somewhat important	Quite important	Extremely important
---------------	------------------------	--------------------	-----------------	---------------------

I would feel embarrassed if people saw me exercising.

Not important	A little bit important	Somewhat important	Quite important	Extremely important
---------------	------------------------	--------------------	-----------------	---------------------

I would feel less stressed if I exercised regularly.

Not important	A little bit important	Somewhat important	Quite important	Extremely important
---------------	------------------------	--------------------	-----------------	---------------------

Exercise prevents me from spending time with my friends.

Not important	A little bit important	Somewhat important	Quite important	Extremely important
---------------	------------------------	--------------------	-----------------	---------------------

Exercising puts me in a better mood for the rest of the day.

Not important	A little bit important	Somewhat important	Quite important	Extremely important
---------------	------------------------	--------------------	-----------------	---------------------

I feel uncomfortable or embarrassed in exercise clothes.

Not important	A little bit important	Somewhat important	Quite important	Extremely important
---------------	------------------------	--------------------	-----------------	---------------------

I would feel more comfortable with my body if I exercised regularly.

Not important	A little bit important	Somewhat important	Quite important	Extremely important
---------------	------------------------	--------------------	-----------------	---------------------

There is too much I would have to learn to exercise.

Not important	A little bit important	Somewhat important	Quite important	Extremely important
---------------	------------------------	--------------------	-----------------	---------------------

Regular exercise would help me have a more positive outlook on life.

Not important	A little bit important	Somewhat important	Quite important	Extremely important
---------------	------------------------	--------------------	-----------------	---------------------

Exercise puts an extra burden on my significant other.

Not important	A little bit important	Somewhat important	Quite important	Extremely important
---------------	------------------------	--------------------	-----------------	---------------------

Post-Intervention Assessment (web form)

Regular exercise is any *planned* physical activity (e.g., brisk walking, aerobics, jogging, bicycling, swimming, rowing, etc.) performed to increase physical fitness. Such activity should be performed *3 to 5 times* per week for *20-60 minutes* per session. Exercise does not have to be painful to be effective but should be done at a level that increases your breathing rate and causes you to break a sweat.

Do you exercise regularly according to that definition? (Please select one):

Yes, I have been for MORE than 6 months
Yes, I have been for LESS than 6 months
No, but I intend to in the next 30 days
No, but I intend to in the next 6 months
No, and I do NOT intend to in the next 6 months

[Continue](#)

Post-Intervention Assessment (web form)

Longitudinal Evaluation of a Web-Based Physical Activity Counselor

Thanks! You've completed the study.

The counselor is still available, and you can come back here for a quick check-in with her up to once each day. You'll be able to talk to the counselor again tomorrow, _____.

Do you have any questions or comments? Please feel free to tell us what you think about the counselor or the study.

[Send Comment](#)

Email:

Password:

[Log In](#)

Final Session Completion (web form)

Longitudinal Evaluation of a Web-Based Physical Activity Counselor

The study is now closed. Thank you for your participation.

Do you have any questions or comments? Please feel free to tell us what you think about the counselor or the study.

Send Comment

End of Participation (web form)

Hello *name*

Thank you for participating in the animated physical activity counselor study.

To continue, please click on the link below, or copy and paste it into the address bar of your web browser:

<http://wonder.ccs.neu.edu/counselor-study/confirm?token=token>

After you click on the link, you'll be asked to fill out a few short questionnaires, then have your first session with the animated counselor.

Sign-Up Email Template

Congratulations *name*!

You have won the weekly drawing for the animated physical activity counselor study. You should be receiving a \$50 Amazon.com gift card by email.

Thank you again for participating in this study.

Remember that you are still eligible to win another drawing: you will be entered in a new drawing each week that you have a conversation with the counselor.

Drawing Notification Email Template

Hello *name*,

This is week *week* of the animated physical activity counselor study. To log in, and have your conversation with the counselor for this week, please click on this link (or copy and paste it into the address bar of your web browser):

<http://wonder.ccs.neu.edu/counselor-study>

Each week that you have a conversation with the counselor, you are entered in a drawing to win a \$50 gift card to Amazon.com. Even if you have won a drawing before, you are still eligible to win every week.

You can talk to the counselor any time you want during the week, day or night.

If you do not want to receive these reminder emails, please click on this link:

<http://wonder.ccs.neu.edu/counselor-study/withdraw?token=token>

Weekly Reminder Email Template

Hello *name*

Thank you for participating in the animated physical activity counselor study!

It looks like you haven't yet had a conversation with the counselor for this week. To log in and have your conversation, you can click on this link (or copy and paste it into the address bar of your web browser):

<http://wonder.ccs.neu.edu/counselor-study>

Each week that you have a conversation with the counselor, you are entered in a drawing to win a \$50 gift card to Amazon.com. Even if you have won a drawing before, you are still eligible to win every week.

You can talk to the counselor any time you want during the week, day or night.

If you do not want to receive these reminder emails, please click on this link:

<http://wonder.ccs.neu.edu/counselor-study/withdraw?token=token>

Additional Reminder Email Template

Hello *name*,

Your participation in the animated physical activity counselor study is now finished. If possible, we ask you to please log in and complete a final set of questionnaires which tell us how you feel about physical activity. The questionnaires are short and should only take a couple minutes to complete.

To log in, click on this link, or copy and paste it into the address bar of your web browser:

<http://wonder.ccs.neu.edu/counselor-study>

Thank you again for your time. Your participation helps us to develop better systems to promote health and well-being.

If you do not want to receive these reminder emails, please click on this link:

<http://wonder.ccs.neu.edu/counselor-study/withdraw?token=token>

Final Assessment Email Template

Dear Participant,

Thank you for participating in the animated physical activity counselor study. During this study, you were asked to have several conversations with an animated computer-controlled counselor, and were told that the purpose of this study was to evaluate how people respond to this counselor.

However, we did not tell you that participants were randomly placed into one of three groups. In all groups, participants had similar conversations with the counselor, but the verbal and nonverbal behavior of the counselor was different. In the first group, the counselor's behavior was consistent throughout the study, while in the other two, there were changes in later conversations, including faster speech on some words, fewer posture shifts, fewer smiles and frowns, fewer head movements, and looking away from you more while speaking. In the second group, these changes matched those observed in a human counselor, while in the third they were exaggerated, and changed three times as much.

The purpose of this manipulation was to study the effect of these changes in behavior on participants' attitude toward the counselor, their willingness to talk to the counselor, and on the counselor's effectiveness in promoting physical activity. We did not disclose this at the start of the study in order to avoid influencing your response to the counselor.

If you have any questions about this study, please feel free to contact Daniel Schulman, phone: 617-373-4605, email: schulman@ccs.neu.edu, the person mainly responsible for the research. You can also contact Timothy Bickmore, phone: 617-373-5477, email: bickmore@ccs.neu.edu, the Principal Investigator.

If you have any questions regarding your rights as a research participant, please contact Nan C. Regina, Director, Human Subject Research Protection, 960 Renaissance Park, Northeastern University, Boston, MA 02115. Tel: 617-373-7570, Email: irb@neu.edu. You may call anonymously if you wish.

Debriefing Email Template

APPENDIX F

Sample Conversation Transcripts for the Evaluation Study

The following are a set of example conversation transcripts for the user-agent conversations in the evaluation study (Chapter 8). These conversations do not represent a single participant’s interactions with the system; rather they are intended to be illustrative of the range of possible interactions.

Users’ input to the system was largely limited to multiple choice from a short list of possible utterances. This is indicated in the transcripts below by giving the full set of possible user utterances at each turn, with the chosen one in italics.

F.1 Initial Conversations

At the first conversation, participants receive, following an introduction to the system, an in-conversation Stage of Change assessment (turns 23–24 in the first transcript below). Participants who give an answer indicative of an early stage of change (precontemplation or contemplation) receive content inspired by Motivational Interviewing [112]; in the first conversation this consists of a discussion of the pros and cons of physical activity, with reflective and reframing responses. Participants who give an answer indicative of a later stage of

change receive a goal setting intervention, beginning in the first conversation with a discussion of the long-term goal and negotiation of a first weekly goal.

F.1.1 Initial Motivational Interviewing

- (1) **Agent:**
Hello. To talk to me, just click the buttons to the right.
- (2) **User:**
OK, this is where I click.
- (3) **Agent:**
Hi, NAME. My name is Karen.
- (4) **User:**
Hi Karen.
Hello Karen.
- (5) **Agent:**
It's nice to meet you. I'm looking forward to talking with you.
- (6) **User:**
Nice to meet you too.
OK.
What will we talk about?
- (7) **Agent:**
Great! We are going to be working together to help you get more walking and improve your daily physical activity.
- (8) **User:**
OK.
What will we do?
What should I do?
- (9) **Agent:**
Okay. When I talk about physical activity, what kinds of things come to mind?
- (10) **User:**
Healthy.
Fit.
Fun.
Embarrassing.
Sweat!
Not fun.
- (11) **Agent:**
I understand. There are lots of things that can make it hard to get regular physical activity. That's why we're going to be talking about brisk walking, which is simple and easy to fit into your life.
- (12) **User:**

Okay.

Great!

(13) **Agent:**

First of all, let's talk about how you feel about physical activity, right now. That will help up have a better conversation about walking.

(14) **User:**

OK.

(15) **Agent:**

The long-term goal is to get up to doing at least thirty minutes of physical activity, at least five days every week.

(16) **User:**

OK.

No problem.

Sounds like a lot.

What is physical activity?

(17) **Agent:**

When I talk about exercise, I mean structured physical activity that makes you breathe hard or break a sweat.

(18) **User:**

How hard should I work?

Structured?

What kind of activity?

OK.

(19) **Agent:**

Things like housework, golfing using a cart, and walking around the office do not count. But things like brisk walking, bicycling, or playing sports are great!

(20) **User:**

OK.

I get it.

(21) **Agent:**

So, that's what I mean by physical activity. And like I said, the goal is thirty minutes, five days a week.

(22) **User:**

OK.

No problem.

Sounds like a lot.

(23) **Agent:**

How do you feel about this? Do you think you could work up to doing this much physical activity, every week?

(24) **User:**

Probably not.

I'm considering it.

Maybe.

Yes, I will.

- I seriously intend to.
I'm doing it now!
- (25) **Agent:**
Thanks.
- (26) **Agent:**
What's the most important reason to get more physical activity that you can think of right now?
- (27) **User:**
I'd have more energy.
I'd feel less stressed.
I'd be in a good mood.
I'd feel more fit.
I'd feel more positive about life.
Something else.
Nothing.
- (28) **Agent:**
A little walking or other physical activity is a great way to reduce stress!
There's nothing like getting moving to take your mind off things.
- (29) **User:**
OK.
Yes.
- (30) **Agent:**
And what makes it hardest to do more physical activity?
- (31) **User:**
It's embarrassing.
It takes too much time.
It's uncomfortable.
It's difficult.
It's hard on my friends and family.
Something else.
Nothing.
- (32) **Agent:**
I understand. Exercise can often be sweaty or just uncomfortable, and that can be a big deal, especially if you're not used to it. I like walking because it's easy to do at your own pace, and you can wear comfortable clothes if it helps.
- (33) **User:**
OK.
Yes.
Maybe.
- (34) **Agent:**
So, you're not sure about physical activity because it might be uncomfortable. On the other hand, it could help you feel less stressed. If we work together to deal with these obstacles, do you think you are willing

to try some easy ways to get you started with some activity?

(35) **User:**

OK.

Great!

Not right now.

(36) **Agent:**

That's okay. It is always your choice. I'd like to keep talking with you about your physical activity.

(37) **User:**

Let's talk now.

Not right now.

Actually, I'm ready to try walking.

(38) **Agent:**

Okay.

(39) **Agent:**

That's all for today. But before you go, let's talk about the rest of our sessions, and when we will talk again.

(40) **User:**

OK.

When should we talk again?

(41) **Agent:**

We should try and talk once per week for the next six weeks. Each week, we'll catch up on what you've been doing since the last time we talked.

(42) **User:**

OK.

What day should we talk?

What time should we talk?

I'm not sure I can.

(43) **Agent:**

We can talk whenever you want during the week, day or night, but it can help to schedule an appointment.

(44) **User:**

OK.

I guess.

I'm not sure I can.

(45) **Agent:**

So, when do you want to talk again?

(46) **User:**

Monday.

Tuesday.

Wednesday.

Thursday.

Friday.

Saturday.

Sunday.

- I'm not sure.
- (47) **Agent:**
Great! I'll see you then.
- (48) **User:**
OK.
- (49) **Agent:**
Okay, NAME. I look forward to talking with you again!
- (50) **User:**
OK. Me, too.
Bye, Karen.
- (51) **Agent:**
Bye, NAME. Have a good week!

F.1.2 Initial Goal Setting

- (1) **Agent:**
Hello. To talk to me, just click the buttons to the right.
- (2) **User:**
OK, this is where I click.
- (3) **Agent:**
Hi, NAME. My name is Karen.
- (4) **User:**
Hi Karen.
Hello Karen.
- (5) **Agent:**
It's nice to meet you. I'm looking forward to talking with you.
- (6) **User:**
Nice to meet you too.
OK.
What will we talk about?
- (7) **Agent:**
I'm glad you asked. We are going to work together to help you get more walking and improve your daily physical activity.
- (8) **User:**
OK.
What will we do?
What should I do?
- (9) **Agent:**
Okay. When I talk about physical activity, what kinds of things come to mind?
- (10) **User:**
Healthy.

Fit.

Fun.

Embarrassing.

Sweat!

Not fun.

(11) **Agent:**

Great! It sounds like you have a really positive attitude. All kinds of physical activity are great, but we're going to be talking about brisk walking, which is simple and easy to fit into your life.

(12) **User:**

OK.

Great!

(13) **Agent:**

First of all, let's talk about how you feel about physical activity, right now. That will help up have a better conversation about walking.

(14) **User:**

OK.

(15) **Agent:**

The long-term goal is to get up to doing at least thirty minutes of physical activity, at least five days every week.

(16) **User:**

OK.

No problem.

Sounds like a lot.

What is physical activity?

(17) **Agent:**

How do you feel about this? Do you think you could work up to doing this much physical activity, every week?

(18) **User:**

Probably not.

I'm considering it.

Maybe.

Yes, I will.

I seriously intend to.

I'm doing it now!

(19) **Agent:**

Thanks.

(20) **Agent:**

Okay, I think it's time to get moving! In the next week, I want you to try and do some brisk walking.

(21) **User:**

I can do that.

I'm not sure.

I don't think I can.

How much should I walk?

- I can walk more than once!
- (22) **Agent:**
I know it can be difficult to get started. Is there anything stopping you from walking that you would like to talk about?
- (23) **User:**
Actually, I think I will walk.
Yes, there are some problems.
Nothing I want to talk about.
I just don't want to.
- (24) **Agent:**
Okay. What is making it difficult to set a walking goal?
- (25) **User:**
Nothing I want to talk about.
I don't have time.
I just don't think I can do it.
I just don't feel like it.
The weather is terrible.
Something else.
- (26) **Agent:**
A lot of people think they do not have the time to walk, but I find that if you make yourself the priority and schedule exercise in your calendar first, you can fit other things around it.
- (27) **User:**
OK.
- (28) **Agent:**
I hope that helps a little. Do you think you can try and go for a walk?
- (29) **User:**
Yes, I will try.
No thanks.
- (30) **Agent:**
Great! I'm sure you'll do it.
- (31) **User:**
OK.
Thanks.
- (32) **Agent:**
That's all for today. But before you go, let's talk about the rest of our sessions, and when we will talk again.
- (33) **User:**
OK.
When should we talk again?
- (34) **Agent:**
We should try and talk once per week for the next six weeks. Each week, we'll catch up on what you've been doing since the last time we talked.
- (35) **User:**
OK.

- What day should we talk?
What time should we talk?
I'm not sure I can.
- (36) **Agent:**
We can talk whenever you want during the week, day or night, but it can help to schedule an appointment.
- (37) **User:**
OK.
I guess.
I'm not sure I can.
- (38) **Agent:**
So, when do you want to talk again?
- (39) **User:**
Monday.
Tuesday.
Wednesday.
Thursday.
Friday.
Saturday.
Sunday.
I'm not sure.
- (40) **Agent:**
Great! I'll see you then.
- (41) **User:**
OK.
- (42) **Agent:**
Okay, NAME. I look forward to talking with you again!
- (43) **User:**
OK.
Me, too.
Bye, Karen.
- (44) **Agent:**
Bye, NAME. Have a good week!

F.2 Routine Conversations

The second through fourth conversations (referred to here as “routine” conversations) all have the same content conditional on a participant’s prior behavior. Participants initially assessed at an early Stage of Change continue to receive Motivational Interviewing-inspired content — a discussion of reasons for mo-

tivation or confidence, with reflective and reframing responses — until they indicate strong motivation and confidence for change, at which point they begin receiving goal-setting content. Participants receiving goal-setting content, either from the initial conversation or later, negotiate a series of weekly goals, and begin each conversation by reviewing the prior goal and receiving either positive reinforcement or problem-solving.

The fifth conversations (not shown here) were similar to the routine conversations, with the exception of a brief addition just prior to the end of the conversation, noting to the participant that the next conversation would be the final conversation.

F.2.1 Motivational Interviewing

- (1) **Agent:**
Hello, NAME.
- (2) **User:**
Hello.
Hi, Karen.
Hi Karen, good to see you again.
Hi Karen, how are you?
- (3) **Agent:**
I'm glad to see you this week.
- (4) **User:**
OK.
Sure.
Good to see you too.
- (5) **Agent:**
How are you doing today?
- (6) **User:**
Great!
Good.
Okay.
Not bad.
Could be better.
Terrible.
- (7) **Agent:**
That's good to hear.
- (8) **User:**

OK.

Thanks.

And how are you?

It's been a fun week.

I feel good about walking this week.

(9) **Agent:**

Doing well as usual, thanks.

(10) **User:**

OK.

Great.

Sure.

(11) **Agent:**

Okay. How motivated are you to do more physical activity? Can you give it a number, on a scale from 0 to 10?

(12) **User:**

3

(13) **Agent:**

It sounds like you have some reasons to feel motivated to get more physical activity, since you answered 3 instead of zero. What would it take to move you up to 7? What stops you from feeling more motivated?

(14) **User:**

I just haven't thought about it.

Physical activity won't help me right now.

Other things are more important.

Physical activity won't help me.

I'll never manage to do it.

Actually, I DO feel very motivated.

I'd rather not talk about this now.

(15) **Agent:**

In that case, let's talk about some of the benefits of physical activity. One of the great things about physical activity is that it helps you both now and in the future. Which would you like to talk about?

(16) **User:**

Benefits now.

Benefits later.

Neither.

(17) **Agent:**

Sure. Besides helping you to get and stay healthy and fit, physical activity also can give you more energy and reduce stress. Many people find it helps to keep up a positive outlook on life.

(18) **User:**

OK.

(19) **Agent:**

Thanks for talking about this with me. How are you feeling about physical activity after our discussion?

- (20) **User:**
I'm not sure.
About the same.
I feel more motivated.
- (21) **Agent:**
Okay. We can talk about this again another time.
- (22) **Agent:**
That's about all for today. Do you want to pick a day to talk next week?
- (23) **User:**
Sure, let's pick a day.
Whatever day you want.
I'm not sure which day is good.
I'm not sure I can talk next week.
I can't talk next week.
- (24) **Agent:**
I understand that finding time can be hard, but it's important that we keep working on your physical activity. If it's hard to pick a day right now, that's fine too. I'm always here, so we can talk whenever you want.
- (25) **User:**
I will pick a day.
I'll talk, but I'm not sure when.
I'm still not sure I can talk.
- (26) **Agent:**
Okay. If you change your mind, remember that I am always available.
- (27) **User:**
OK.
- (28) **Agent:**
Okay, NAME. I look forward to talking with you again!
- (29) **User:**
OK.
Me too.
Bye, Karen.
- (30) **Agent:**
Bye, NAME. Have a good week!

F.2.2 Motivational Interviewing, and Goal Setting

- (1) **Agent:**
Hello, NAME.
- (2) **User:**
Hello.
Hi Karen.
Hi Karen, good to see you again.

Hi Karen, how are you?

(3) **Agent:**
I'm doing well, thanks for asking.

(4) **Agent:**
How are you?

(5) **User:**
Great!
Good Okay.
Not bad.
Could be better.
Terrible.

(6) **Agent:**
That's always good to hear.

(7) **User:**
OK.
Thanks.
It's been a fun week.
I feel good about walking this week.
I don't feel good about the walking.

(8) **Agent:**
Great! That's what we're here to talk about.

(9) **User:**
OK.

(10) **Agent:**
Okay. How confident are you that you could do more physical activity, on a scale from 0 to 10?

(11) **User:**
8

(12) **Agent:**
Great! It sounds like you are really confident that you can get more physical activity.

(13) **User:**
OK.
Yes!
Not really.
I'm not sure.

(14) **Agent:**
Tell me more. What is keeping you from being even more confident?

(15) **User:**
It sounds like hard work!
I'm not good at exercise.
I'm bad at starting things.
I might start, but I won't keep it up.
I can't do it by myself.
I'm just not ready.

- Actually, I DO feel very confident.
I don't want to talk about this now.
- (16) **Agent:**
I understand! It's easy to procrastinate when starting something new.
I think everyone does that.
- (17) **User:**
OK.
That's right.
- (18) **Agent:**
So it sounds like you could probably be more confident if we could just get you started with something quick, and get you past that first step.
- (19) **User:**
OK.
Maybe.
No thanks.
- (20) **Agent:**
Thanks for talking about this with me. How are you feeling about physical activity after our discussion?
- (21) **User:**
I'm not sure.
About the same.
I feel more confident.
- (22) **Agent:**
Since you feel that physical activity is important and you are feeling pretty confident you can do more, we should talk next about the first steps you can do. Do you want to talk about this now?
- (23) **User:**
Yes.
Sure.
Not right now.
- (24) **Agent:**
Okay, I think it's time to get moving! In the next week, I want you to try and do some brisk walking.
- (25) **User:**
I can do that.
I'm not sure.
I don't think I can.
How much should I walk?
I can walk more than once!
- (26) **Agent:**
For now, even a ten minute walk is great. You should be walking briskly, or fast enough to feel that you're working a little harder than usual.
- (27) **User:**
Okay, I can do that.
I'm not sure.

- I don't think I can.
I can walk more!
- (28) **Agent:**
Great! I'm sure you'll do it.
- (29) **User:**
OK.
Thanks.
- (30) **Agent:**
Okay, I think that's all we need to talk about today. Do you want to schedule a day for our next talk?
- (31) **User:**
Sure, let's pick a day.
Whatever day you want.
I'm not sure which day is good.
I'm not sure I can talk next week.
I can't talk next week.
- (32) **Agent:**
So, what day should we talk again?
- (33) **User:**
Monday.
Tuesday.
Wednesday.
Thursday.
Friday.
Saturday.
Sunday.
I'm not sure.
- (34) **Agent:**
Great! See you then.
- (35) **User:**
OK.
- (36) **Agent:**
Okay, NAME. I look forward to talking with you again!
- (37) **User:**
OK.
Me too.
Bye, Karen.
- (38) **Agent:**
Bye, NAME. Have a good week!

F.2.3 Baseline Goal Setting

- (1) **Agent:**
Hello NAME.
- (2) **User:**
Hello.
Hi Karen.
Hi Karen, good to see you again.
Hi Karen, how are you?
- (3) **Agent:**
Thanks! I've been looking forward to it.
- (4) **User:**
OK.
Great.
And how are you?
- (5) **Agent:**
How was your week?
- (6) **User:**
Great!
Good.
Okay.
Not bad.
Could be better.
Terrible.
- (7) **Agent:**
I'm glad to hear that!
- (8) **User:**
OK.
Thanks.
It's been a fun week.
I feel good about walking this week.
- (9) **Agent:**
Last week, you said you would try and do some walking.
- (10) **User:**
I did walk once.
I walked more than once!
I wasn't able to walk.
I tried, but I couldn't.
- (11) **Agent:**
Great! That's a wonderful step toward your goals.
- (12) **User:**
Yes.
OK.
Thanks.
- (13) **Agent:**

So, it is time to start working up to our long-term goal of doing physical activity five times a week. Can you make a goal to walk for at least 20 minutes, 2 times a week?

(14) **User:**

I can do that.

I'm not sure.

I don't think I can.

I might have some problems.

I can walk more!

(15) **Agent:**

Great! I'm sure you'll do it.

(16) **User:**

OK.

Thanks.

(17) **Agent:**

That's about all for today. Do you want to pick a day to talk next week?

(18) **User:**

Sure, let's pick a day.

Whatever day you want.

I'm not sure which day is good.

I'm not sure I can talk next week.

I can't talk next week.

(19) **Agent:**

Thanks! I'm always here, so we can talk any day you want. If you don't want to pick a day now, that's fine too.

(20) **User:**

I don't want to pick a day.

I will pick a day.

I'm not sure I can talk next week.

I can't talk next week.

(21) **Agent:**

Great! Just talk to me whenever you have time. I will be here.

(22) **User:**

OK.

(23) **Agent:**

Okay, NAME. I look forward to seeing you next week!

(24) **User:**

OK.

Me too.

Bye, Karen.

(25) **Agent:**

Bye, NAME. Have a good week!

F.2.4 Incremental Goal Setting

- (1) **Agent:**
Hello, NAME.
- (2) **User:**
Hello.
Hi, Karen.
Hi Karen, good to see you again.
Hi Karen, how are you?
- (3) **Agent:**
Thanks for talking with me.
- (4) **User:**
OK.
Sure.
Good to see you too.
- (5) **Agent:**
How are you?
- (6) **User:**
Great!
Good.
OK.
Not bad.
Could be better.
Terrible.
- (7) **Agent:**
Okay, great.
- (8) **User:**
Thanks.
And how are you?
I feel good about walking this week.
It's been a fun week.
I feel good about walking this week.
I don't feel good about the walking.
- (9) **Agent:**
So, your goal from last time was to walk for 20 minutes, 3 days a week.
How did it go?
- (10) **User:**
I didn't walk at all.
I only walked once.
I walked less than that.
I did it!
I walked more!
- (11) **Agent:**
I'm sorry to hear that. But it sounds like even when you had a rough week, you still tried to do some walking. Would you like some tips that

may help with weeks like that?

(12) **User:**

Yes.

No.

(13) **Agent:**

Here's a plot of the walking you've been doing.

(14) **User:**

OK.

Thanks.

(15) **Agent:**

So, it's time to set a goal for next week. Can you try again to do 20 minutes, 3 times a week?

(16) **User:**

I can do that.

I'm not sure.

I don't think I can.

I might have some problems.

I can walk more!

(17) **Agent:**

Great! I'm sure you'll do it.

(18) **User:**

OK.

Thanks.

(19) **Agent:**

We're just about done for today. Can we schedule a day for our conversation next week?

(20) **User:**

Sure, let's pick a day.

Whatever day out want.

I'm not sure which day is good.

I'm not sure I can talk next week.

I can't talk next week.

(21) **Agent:**

I'm always here, so we can talk whenever you want. We don't have to pick a day now.

(22) **User:**

I don't want to pick a day.

I will pick a day.

I'm not sure I can talk next week.

I can't talk next week.

(23) **Agent:**

Great! I'll see you next week.

(24) **User:**

OK.

(25) **Agent:**

- Okay, NAME. I look forward to seeing you next week!
- (26) **User:**
OK.
Me too.
Bye, Karen.
- (27) **Agent:**
Bye, NAME. Have a good week!

F.3 Final Sessions

The sixth and final sessions followed a different pattern from prior conversations. Participants receiving a Motivational Interviewing-inspired intervention received a discussion of pros and cons of physical activity, similar to the initial conversation. Participants receiving a goal-setting intervention had a final discussion of the most recent weekly goal, with an opportunity for problem solving; they did not negotiate a new weekly goal.

In both cases, the conversation included a brief meta-dialogue about the state of the user-agent relationship; this is shown in turns 37–41 of the first transcript below.

F.3.1 Final Motivational Interviewing

- (1) **Agent:**
Hello, NAME.
- (2) **User:**
Hello.
Hi Karen
Hi Karen, good to see you again.
Hi Karen, how are you?
- (3) **Agent:**
Good to see you again.
- (4) **User:**
OK.
Sure.
Good to see you too.
- (5) **Agent:**
How are things going?

-
- (6) **User:**
Great!
Good.
Okay.
Not bad.
Could be better.
Terrible.
- (7) **Agent:**
Okay, great.
- (8) **User:**
Thanks
And how are you?
I feel good about walking this week.
I don't feel good about the walking.
- (9) **Agent:**
So, NAME, this is our last weekly session. I've really enjoyed talking with you.
- (10) **User:**
OK. Thanks.
I've enjoyed it too.
What happens now?
- (11) **Agent:**
I'll still be around for a couple more weeks if you want to do a quick check-in at any time.
- (12) **User:**
OK.
I'll do that.
What will we talk about today?
- (13) **Agent:**
Today, we'll wrap up everything we've discussed about walking any physical activity and try and make sure to get to any concerns you have.
- (14) **User:**
OK.
- (15) **Agent:**
Let's talk once more about some of your reasons for thinking about physical activity.
- (16) **User:**
OK.
I guess.
- (17) **Agent:**
What's the most important reasons to get more physical activity that you can think of right now?
- (18) **User:**
I'd have more energy.
I'd feel less stressed.

- I'd be in a good mood.
I'd feel more fit.
I'd feel more positive about life.
Something else.
Nothing.
- (19) **Agent:**
That's true. Even if a little exercise ight tire you out, it can give you more energy throughout the day.
- (20) **User:**
OK.
Yes.
- (21) **Agent:**
And what gets in your way of doing more physical activity?
- (22) **User:**
It's embarrassing.
It takes too much time.
It's uncomfortable.
It's difficult.
It's hard on my friends and family.
Something else.
Nothing.
- (23) **Agent:**
It can be hard to find time to exercise. Remember to try and fit some short brisk walks into your day.
- (24) **User:**
OK.
Yes.
Maybe.
- (25) **Agent:**
So, you're not sure about physical activity because it might be hard to fit in your day, but on the other hand, it can give you more energy. It sounds like you've spent some time thinking about physical activity. That can be difficult, so good for you.
- (26) **User:**
OK.
Yes.
I guess.
- (27) **Agent:**
Since this may be our last conversation, I thought it would be a good idea to talk one last time about some tips for walking.
- (28) **User:**
OK.
I guess.
No thanks.
- (29) **Agent:**

Okay. What kinds of problems would you like to talk about.

(30) **User:**

Nothing else, thanks.

Trouble finding enough time.

Feeling like I can't do it.

When I don't want to exercise.

Bad weather.

A different problem.

(31) **Agent:**

Even with all your other responsibilities, it is important to get some exercise in every day. You may want to try exercising for ten minutes at a time, several times a day. That way you can fit your exercise in, without taking away from anything else.

(32) **User:**

OK.

(33) **Agent:**

Do you want to hear about any other problems?

(34) **User:**

Nothing else, thanks.

Trouble finding enough time.

Feeling like I can't do it.

When I don't want to exercise.

Bad weather.

A different problem.

(35) **Agent:**

I hope these tips can help you in the future.

(36) **User:**

OK.

Thanks.

(37) **Agent:**

So, now that it's been several weeks, how do you feel about working with me?

(38) **User:**

Great!

Okay.

I'm not sure.

It's been kind of strange.

It's not that great.

I'd rather not say.

(39) **Agent:**

I'm sorry to hear that. I know that talking with a computer about physical activity, must be weird, and maybe a little uncomfortable.

(40) **User:**

Yes.

I guess.

- It's not that bad.
It is pretty weird.
It is repetitive.
It's not helpful to me.
- (41) **Agent:**
Okay. I'm sorry this didn't work out better, and I hope you still got something useful out of our talks.
- (42) **User:**
OK.
- (43) **Agent:**
And how are you feeling about physical activity now?
- (44) **User:**
I can handle it.
I'm not sure, but I think I'm good.
I'm not sure.
I'm worried about doing it on my own.
I'd rather not say.
- (45) **Agent:**
Don't worry. You've made great progress in the time we've talked together, and I'm sure you'll continue to do great.
- (46) **User:**
OK.
Thanks.
- (47) **Agent:**
So, NAME, that's all for today. I know I've said this a few times already, but I'm glad we were able to talk.
- (48) **User:**
OK.
Thanks.
I'm glad too.
- (49) **Agent:**
I'll still be around for a couple more weeks if you want a quick chat about walking and physical activity.
- (50) **User:**
OK.
Thanks.
I may do that.
I will do that.
- (51) **Agent:**
Goodbye, NAME.
- (52) **User:**
Goodbye.
Bye, Karen.

F.3.2 Final Goal Setting

- (1) **Agent:**
Hello NAME.
- (2) **User:**
Hello.
Hi Karen.
Hi Karen, good to see you again.
Hi Karen, how are you?
- (3) **Agent:**
I'm glad to see you this week.
- (4) **User:**
OK.
Sure
Good to see you too.
- (5) **Agent:**
How are you?
- (6) **User:**
Great!
Good.
Okay.
Not bad.
Could be better.
Terrible.
- (7) **Agent:**
I'm glad to hear that.
- (8) **User:**
Okay.
Thanks.
It's been a fun week.
I feel good about walking this week.
- (9) **Agent:**
I'm very glad to hear that.
- (10) **User:**
OK.
- (11) **Agent:**
So, NAME, this is our last weekly session. I've really enjoyed talking with you.
- (12) **User:**
OK.
Thanks.
I've enjoyed it too.
What happens now?
- (13) **Agent:**
Thanks! I'm glad to hear that.

- (14) **Agent:**
I'll still be around for a couple more weeks if you want to do a quick check-in at any time.
- (15) **User:**
OK.
I'll do that.
What will we talk about today?
- (16) **Agent:**
Today, we'll wrap up everything we've discussed about walking and physical activity and try and make sure to get to any concerns you have.
- (17) **User:**
OK.
- (18) **Agent:**
Last week, you said you might not be able to do any walking by today.
- (19) **User:**
I did walk once.
I walked more than once!
I didn't walk.
OK.
- (20) **Agent:**
Great! It sounds like you are really doing well! How many days did you walk?
- (21) **User:**
4
- (22) **Agent:**
Here's a plot showing your progress so far!
- (23) **User:**
OK.
Thanks.
- (24) **Agent:**
Okay. Since this may be our last conversation, I thought it would be a good idea to talk one last time about some tips for walking.
- (25) **User:**
OK.
I guess.
No thanks.
- (26) **Agent:**
Are you sure? I know we've already talked a lot, but you never know when some tips might be useful.
- (27) **User:**
OK, let's talk.
No thanks.
- (28) **Agent:**
OK.
- (29) **Agent:**

So, now that it's been several weeks, how do you feel about working with me?

(30) **User:**

Great!

Okay.

I'm not sure.

It's been kind of strange.

It's not that great.

I'd rather not say.

(31) **Agent:**

That's okay. No problem.

(32) **User:**

OK.

Thanks.

Let's just talk about walking.

(33) **Agent:**

And how are you feeling about physical activity now?

(34) **User:**

I can handle it.

I'm not sure, but I think I'm good.

I'm not sure.

I'm worried about doing it on my own.

I'd rather not say.

(35) **Agent:**

Great! I'm glad to hear that. I'm sure you'll do well.

(36) **User:**

OK.

Thanks.

(37) **Agent:**

So, NAME, that's all for today. I know I've said this a few times already, but I'm glad we were able to talk.

(38) **User:**

OK.

Thanks.

I'm glad too.

(39) **Agent:**

I'll still be around for a couple more weeks if you want a quick chat about walking and physical activity.

(40) **User:**

OK.

Thanks.

I may do that.

I will do that.

(41) **Agent:**

Goodbye, NAME.

(42) **User:**
Goodbye. Bye Karen