

Adapting to Personality Over Time: Examining the Effectiveness of Dialogue Policy Progressions in Task-Oriented Interaction

Alexandria Katarina Vail and Kristy Elizabeth Boyer

Department of Computer Science

North Carolina State University

Raleigh, North Carolina, USA

{akvail, keboyer}@ncsu.edu

Abstract

This paper explores dialogue adaptation over repeated interactions within a task-oriented human tutorial dialogue corpus. We hypothesize that over the course of four tutorial dialogue sessions, tutors adapt their strategies based on the personality of the student, and in particular to student introversion or extraversion. We model changes in strategy over time and use them to predict how effectively the tutorial interactions support student learning. The results suggest that students leaning toward introversion learn more effectively with a minimal amount of interruption during task activity, but occasionally require a tutor prompt before voicing uncertainty; on the other hand, students tending toward extraversion benefit significantly from increased interaction, particularly through tutor prompts for reflection on task activity. This line of investigation will inform the development of future user-adaptive dialogue systems.

1 Introduction

Throughout dialogue interactions, humans adapt to each other in a variety of ways (Cohen et al., 1981; Power, 1974; Wahlster and Kobsa, 1989). Some recent studies suggest that dialogue systems that mirror these adaptations to the user, e.g., by adopting the user’s vocabulary (Niederhoffer and Pennebaker, 2002) or linguistically aligning to the user’s context (Pickering and Garrod, 2004), may be more effective than those that do not. For supporting human dialogue, it has been demonstrated that tutorial dialogue systems improve in effectiveness when they adapt to user uncertainty (Forbes-Riley and Litman, 2007) or perform ‘small talk’ to increase the user’s trust in the system (Cassell

and Bickmore, 2003). Some studies have provided evidence that adapting to the user at the personality level also increases effectiveness; for example, users may become more agreeable when systems mirror their personality (Reeves and Nass, 1997), and varying levels of encouragement may help users of extraverted or introverted personalities accomplish a task more effectively (Tapus and Mataric, 2008).

With this substantial evidence that adapting to user personality may improve the effectiveness of a dialogue system, there is little investigation of how personality affects repeated interactions. For supporting human learning in particular, we hypothesize that taking personality into account may enhance outcomes by providing a more tailored experience. To explore this hypothesis, this paper presents an analysis that uses the change in human tutorial dialogue policies over repeated interaction with introverted and extraverted students to predict the effectiveness of the tutoring. We utilize a widely-used and validated questionnaire, the Big Five Inventory, to determine a personality profile for each student. We hypothesize that introverted and extraverted students learn more effectively under different dialogue policies. The results suggest dialogue policy progressions that could aid in the future development of personality-based user-adaptive tutorial dialogue systems.

2 Related Work

Humans adapt to their dialogue partner in a variety of ways: for example, using knowledge acquired through the dialogue to inform subsequent utterances (Carberry, 1989), maintaining a set of subdialogues (Litman and Allen, 1987), and structuring dialogue to achieve a common goal (Power, 1974), including asking particular sorts of questions (Cohen et al., 1981), reaching dialogue convergence (Mitchell et al., 2012), and understanding context-specific vocabulary (Grosz, 1983). It

has been strongly suggested by a number of studies that dialogue systems would benefit greatly from mirroring this sort of adaptation, e.g., by adopting the user’s syntax (Niederhoffer and Pennebaker, 2002), goal-oriented language (Brennan, 1996), and dialogue structure (Levelt and Kelter, 1982).

Some of these factors have been successfully applied to task-oriented dialogue systems. For example, ‘entrainment’ (the alignment between partners at various linguistic levels) has been shown to be predictive of task success in telephone conversation (Nenkova et al., 2008) and of less misunderstanding in personality-matching systems (Mairesse and Walker, 2010).

In order to gauge user personality, we utilize the Big Five Factor model, which was developed to objectively measure five particular aspects of a person’s personality (Goldberg, 1993). This personality model has been widely implemented in a number of studies of personality in dialogue systems, including recommender systems (Dunn et al., 2009) and conversational systems (Mairesse and Walker, 2010). The investigation of personality as it pertains to tutorial dialogue systems is a natural step for user-adaptive dialogue systems.

3 Tutorial Dialogue Corpus

The corpus under examination in this study consists of computer-mediated human-human textual dialogue (Mitchell et al., 2013; Ha et al., 2013). For each dialogue session, participants included one tutor and one student who cooperated with the goal of creating a working software artifact, a text-based adventure game, by the end of the repeated interactions. Students were first-year university students from an introductory engineering course who volunteered in exchange for course credit. No previous computer science knowledge was assumed or required. The tutors were primarily graduate students with previous experience in tutoring or teaching Java programming.

The tutorial sessions were conducted within a web-based textual dialogue interface for introductory programming in Java. The tutorial dialogue interface, displayed in Figure 1, consists of four panes in which the student interacts: the task description, the compilation and execution output, the student’s Java source code, and the textual dialogue messages between the tutor and the student. The student could modify, compile, and ex-

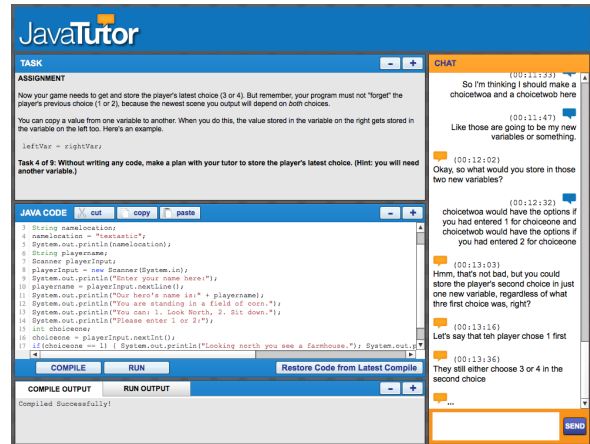


Figure 1: The task-oriented tutorial dialogue interface.

ecute Java code from within the interface, in addition to conversing with the tutor via the textual dialogue pane. The content of the interface was synchronized in real time between the student and the tutor; however, the tutor’s interactions with the environment were constrained to the textual dialogue with the student and the progression between tasks.

The corpus was collected during two university semesters in Fall 2011 and Spring 2012. A total of $N = 67$ students interacted with one of seven tutors to complete the series of interactions during this time frame. The tutoring curriculum was composed of six task-based lessons completed over four weeks, each constrained to forty minutes in duration. Each lesson consisted of multiple sub-tasks, with each lesson concluding at a milestone. This paper considers only the first four of the six lessons, because the fifth lesson suffered from significant data loss due to a database connectivity error, and the sixth lesson consisted of an unstructured review of the previous five lessons, and is therefore a different type of dialogue than the prior lessons. The structure of the corpus is illustrated in Table 1.

The sessions under consideration contained 67 students, with a total of 45,904 utterances: 13,732 student utterances and 32,172 tutor utterances. There were an average of 117 utterances per session: 82 tutor utterances (652 words) and 35 student utterances (184 words). Introverted students averaged 36 utterances and 172 words per session, while extraverted students averaged 34 utterances and 187 words per session. There was no statistically significant difference between in-

Tutor	Student	Lessons					
1	1	L1	L2	L3	L4	L5	L6
1	2	L1	L2	L3	L4	L5	L6
⋮							
2	15	L1	L2	L3	L4	L5	L6
⋮							
3	18	L1	L2	L3	L4	L5	L6
3	19	L1	L2	L3	L4	L5	L6
⋮							

Table 1: A diagram of the structure of the corpus. Gray cells indicate dialogue sessions that were not considered in the present analysis.

troverts and extraverts on these counts. The possible extraversion score on the questionnaire ranges from -10 (highly introverted) to 25 (highly extraverted), and the mean extraversion score of the students in our corpus was 6.40 (standard deviation 6.42). The distribution of scores across the sample was comparable to a normal distribution, as demonstrated by the histogram in Figure 2.

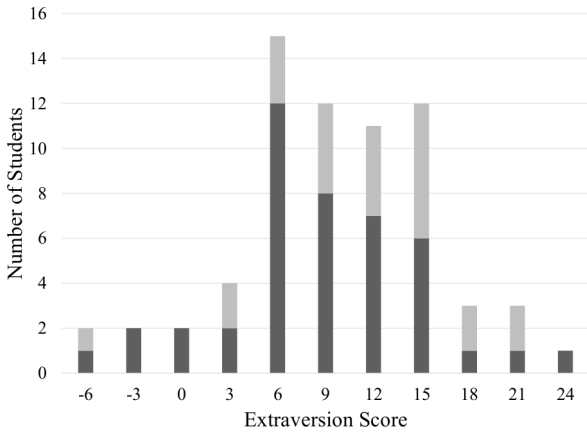


Figure 2: Histogram of extraversion scores across students in the corpus. Lighter bars indicate female students, while darker bars indicate male students.

3.1 Learning Gain

Students completed an identical pretest and posttest for each lesson. The average pretest and posttest scores for students scoring above and below the median extraversion score in the four lessons are detailed in Table 3 (determination of extraversion is detailed in Section 3.2). There was no statistically significant difference between the scores of extraverted and introverted students.

The tutoring was statistically significantly effective overall and within each student group ($p \ll 0.0001$, on all accounts).

Lesson	Pretest		Posttest	
	Introvert	Extravert	Introvert	Extravert
L1	50.69%	47.42%	71.63%	68.18%
L2	43.70%	38.96%	71.01%	73.59%
L3	55.88%	54.55%	67.65%	64.85%
L4	68.79%	65.66%	80.56%	79.97%

Table 3: Average pretest and posttest scores for each lesson.

This equation adjusts for negative learning gain in the rare cases that posttest score is less than pretest score (Marx and Cummings, 2007).

$$norm_gain = \begin{cases} \frac{post-pre}{1-pre} & post > pre \\ \frac{post-pre}{pre} & post \leq pre \end{cases} \quad (1)$$

Since pretest and posttest scores for introverts and extraverts were not identical, normalized learning gain was standardized within each group before developing models to predict learning (Section 4).

3.2 Extraversion vs. Introversion

One of the standard frameworks for identifying personality traits is the Big Five Factor model of personality (Goldberg, 1993). The standard method of testing for the Big Five personality traits is by questionnaire (John and Srivastava, 1999; Gosling et al., 2003). The students under consideration in this study were administered a Big Five Inventory survey, a type of self-assessment of personality, prior to any interaction with the tutorial dialogue system. The Big Five Inventory consists of 44 items to measure an individual on the Big Five Factors of personality: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (Goldberg, 1993). This study focuses on a student’s responses to the items reflective of extraversion and introversion. These items are identified in Table 4. Extraversion is defined as the part of the Big Five Factors that identifies gregariousness, assertiveness, activity, excitement-seeking, positive emotions, and warmth (John and Srivastava, 1999).

3.3 Dialogue Act Annotation

As described in the previous section, the corpus being considered here consists of 268 dialogues, four

Extraverted Student Dialogue Excerpt
STUDENT: So do we need an <code>else</code> statement for each one? [QI]
TUTOR: That wouldn't actually work. [AWH]
STUDENT: Really? [FNU]
TUTOR: See, because it's testing them each independently. [E]
TUTOR: So when it gets to 2 and 4, any other combination goes to its <code>else</code> . [E]
<i>Pause for 29 seconds.</i>
TUTOR: If we added an <code>else</code> clause for each statement, we'd end up with 3 of them printing out for every valid input. [E]
STUDENT: Oh. [ACK]
<i>Pause for 44 seconds.</i>
TUTOR: What else do you think we could try? [QP]
<i>Pause for 49 seconds.</i>
STUDENT: Well the first one worked last time because it was checking only <code>playerChoice</code> ...maybe <code>currentChoice</code> has something to do with this case. [AWH]

Introverted Student Dialogue Excerpt
STUDENT: The <code>else</code> applies no matter what because it doesn't have an <code>else if</code> to combine with? [QI]
TUTOR: Well, it's a little different than that. [AWH]
TUTOR: Each <code>if</code> statement applies no matter what. [I]
TUTOR: So, instead of checking the values as mutually exclusive conditions, each <code>if</code> is checked in sequence. [I]
<i>Pause for 22 seconds.</i>
TUTOR: Your <code>else</code> occurs only with the final <code>if</code> , regardless of what happened with the previous <code>if</code> statements! [E]
<i>Pause for 31 seconds.</i>
TUTOR: Let's fix it by doing the change that you started much earlier. [D]
<i>Pause for 50 seconds.</i>
TUTOR: Much better. :) [FP]
STUDENT: Thanks! [ACK]
<i>Pause for 22 seconds.</i>
TUTOR: Do you have any issues with the input checking as it is now? [QP]
<i>Pause for 46 seconds.</i>
STUDENT: I do not! [AYN]

Table 2: Excerpts of similar dialogue between an extraverted and an introverted student.

I see myself as someone who ...

- ... is talkative.
- ... is reserved.*
- ... is full of energy.
- ... generates a lot of enthusiasm.
- ... tends to be quiet.*
- ... has an assertive personality.
- ... is sometimes shy, inhibited.*
- ... is outgoing, sociable.

Table 4: Items of the Big Five Inventory reflective of a student's extraversion traits. Asterisks represent items negatively associated with extraversion.

for each of 67 students, with 45,904 utterances total. As described in this section, a portion of these dialogues were manually annotated, and then a supervised dialogue act classifier was trained on them and was used to tag the remaining dialogues.

The annotation scheme applied to the corpus consisted of 31 dialogue act tags grouped into four high-level categories (Statement, Question, Answer, Feedback) (Vail and Boyer, In press). This tagset represents a refinement of previous dialogue act tagsets developed for task-oriented tutoring (Ha et al., 2013). During this refinement,

emphasis was placed on decomposing frequent tags that tended to be broad, such as STATEMENT and QUESTION, in order to capture more fine-grained pedagogical and social phenomena in the dialogues. The annotation scheme is detailed in Table 5.

A total of 30 sessions (4,035 utterances) were manually annotated by a single annotator. Of those 30 sessions, 37% were annotated by a second independent annotator. Inter-annotator agreement on this subset reached a Cohen's kappa of $\kappa=0.87$ (agreement of 89.6%). These manually annotated sessions form the basis for developing an automated classifier.

The automated classifier was trained using the WEKA machine learning software (Hall et al., 2009). We used a J48 decision tree classifier, which has a low running time (Verbree et al., 2006) and as we will see, performed very well for this task. The classifier was provided the features listed in Table 6.

Before the construction of the classifier, the 30 sessions of the manually annotated corpus were systematically split into a training and a test set, consisting of 24 and 6 sessions, respectively; the test set contained the first three sessions with students identified as introverts and the first three sessions with students identified as extraverts. Utterances were defined as single textual messages.

Tag	Example	Session Type		κ
		Introvert	Extravert	
ACKNOWLEDGE (ACK)	<i>Okay.</i>	10.46%	10.36%	0.872
EXTRA-DOMAIN ANSWER (AEX)	<i>I'm doing great.</i>	1.33%	1.42%	0.813
READY ANSWER (AR)	<i>I'm ready.</i>	2.75%	3.08%	0.963
WH-QUESTION ANSWER (AWH)	<i>Line 9.</i>	8.14%	8.10%	0.819
YES/NO ANSWER (AYN)	<i>No, sir.</i>	2.99%	3.73%	0.839
CORRECTION (CO)	<i>*exclamation</i>	0.43%	0.41%	0.700
DIRECTIVE (D)	<i>Test what you have.</i>	6.01%	5.97%	0.888
EXPLANATION (E)	<i>Your code stops on line 2.</i>	31.48%	26.70%	0.822
NEGATIVE FEEDBACK (FN)	<i>No, that's wrong.</i>	0.02%	0.02%	0.615
ELABORATED NEGATIVE FEEDBACK (FNE)	<i>You're using the wrong function.</i>	0.21%	0.14%	0.689
NOT UNDERSTANDING FEEDBACK (FNU)	<i>I'm not sure.</i>	0.05%	0.04%	0.749
OTHER FEEDBACK (FO)	<i>That's okay.</i>	0.17%	0.16%	0.614
ELABORATED OTHER FEEDBACK (FOE)	<i>What you had was fine.</i>	0.29%	0.27%	0.665
POSITIVE FEEDBACK (FP)	<i>Very good!</i>	6.78%	5.45%	0.927
ELABORATED POSITIVE FEEDBACK (FPE)	<i>That's a very good approach.</i>	0.05%	0.12%	0.705
UNDERSTANDING FEEDBACK (FU)	<i>Ohh, I see!</i>	0.76%	0.92%	0.804
GREETING (GRE)	<i>Hello!</i>	2.59%	3.03%	0.941
INFORMATION (I)	<i>Variable names must be one word.</i>	4.55%	5.33%	0.859
OBSERVATION (O)	<i>As you see, we have a bug.</i>	0.25%	0.31%	0.760
EXTRA-DOMAIN OTHER (OEX)	<i>Calculus is difficult.</i>	1.49%	2.22%	0.789
CONFIRMATION QUESTION (QC)	<i>Does that work?</i>	0.16%	0.16%	0.857
DIRECTION QUESTION (QD)	<i>What do I do now?</i>	0.68%	0.58%	0.758
EVALUATIVE QUESTION (QE)	<i>Does that make sense?</i>	0.87%	0.83%	0.763
EXTRA-DOMAIN QUESTION (QEX)	<i>How are you today?</i>	0.42%	0.45%	0.781
FACTUAL QUESTION (QF)	<i>What line is it waiting on?</i>	4.10%	5.12%	0.832
INFORMATION QUESTION (QI)	<i>How do you add spaces?</i>	4.06%	4.91%	0.820
OPEN QUESTION (QO)	<i>How can you fix it?</i>	0.15%	0.14%	0.725
PROBING QUESTION (QP)	<i>Do you think that looks correct?</i>	4.99%	4.76%	0.731
QUESTION PROMPT (QQ)	<i>Any questions?</i>	2.49%	2.24%	0.978
READY QUESTION (QR)	<i>Are you ready to move on?</i>	2.47%	2.75%	0.989
REASSURANCE (R)	<i>We have plenty of time left.</i>	0.12%	0.15%	0.763

Table 5: Dialogue act tags comprising the annotation scheme, the average composition of a Lesson 4 session with introverted and extraverted students, and the Cohen's kappa achieved by the automated classifier.

Feature Description	Number of Features	
	Initial	Selected
TUTOR or STUDENT	1	1
Two-step tag history	2	2
Two-step category history	2	2
Number of tokens in the utterance	1	1
Existence of a question mark	1	1
Existence of word unigrams	1459	160
Existence of word bigrams	8959	150
Existence of POS unigrams	50	31
Existence of POS bigrams	928	152

Table 6: Features provided to the J48 automatic dialogue act classifier.

Feature selection was performed on the features occurring more than three times in the training set using the WEKA machine learning software: various top- N cut-offs were examined for performance on tenfold cross-validation after ranking the features by information gain. A peak in performance during cross-validation on the training set was observed at $N=500$ features.

The final dialogue act classifier includes the following features: speaker role, two-step dialogue act history (category and tag), utterance length, existence of the '?' token, existence of 160 unigrams and 150 bigrams, and existence of 31 part-of-speech unigrams and 152 part-of-speech bigrams.

The part-of-speech tagger used in this analysis was an n -gram tagger within the Natural Language Tool Kit for Python, trained on the NPS chat corpus (Bird et al., 2009; Forsyth and Martell, 2007). The classifier performance on the held-out test set consisting of 714 utterances was 80.11% accuracy, Cohen’s kappa of 0.786. This classifier was then used to tag dialogue acts in the remaining 41, 869 utterances.

4 Extraversion and Dialogue Policy

With the annotated corpus in hand, the goal is to examine how *dialogue policy progression*, as represented by tutors’ contextualized dialogue acts, occurs over time with students tending toward extraversion or introversion. We hypothesize that tutors adapt differently to introverted and extraverted students, and that students of different extraverted or introverted tendencies learn more effectively from different dialogue policies.

Students were binned into two groups, the ‘introverts’, consisting of the students scoring below or equal to the median extraversion score of 7, and the ‘extraverts’, consisting of the students scoring above the median score¹. These groups included 34 and 33 students, respectively.

We describe tutor dialogue policy by identifying the conditional probabilities of a tutor move following a student move (i.e., the probabilities $Pr(T_n|S_{n-1})$) during each session. In other words, we compute bigram probabilities over dialogue acts, where the second dialogue act of the bigram is a tutor move. Because the task-oriented nature of the dialogue allows for extended periods of dialogue silence while the student is working on the task, a WAIT tag was added to the corpus when there was a pause in the dialogue for more than twenty seconds. This threshold was chosen based upon qualitative inspection of the corpus. To identify the changes in this policy over time, we calculated the difference in the probability of each dialogue act bigram between the first and fourth lessons of each student-tutor pair. Finally, in order to allow for directly comparing parameter values across models, each column of predictors was standardized by subtracting the mean and dividing

¹We split on the median introversion/extraversion score as observed in our student sample rather than splitting on a larger population median because the range of personality traits differs significantly based on the sample. To date, no large study has examined university students in order to establish personality norms.

by the standard deviation.

After all of the bigram probabilities were standardized, we split the students into two groups based on median extraversion score: those tending toward extraversion and those tending toward introversion. A feature selection algorithm was then applied to each of these sets in order to identify the most relevant dialogue act bigram features for predicting learning. Any feature that provided non-positive information gain was eliminated from consideration. A stepwise linear regression model was then applied using the SAS statistical modeling software, resulting in the models displayed in Tables 7 and 8. Subscripts indicate the speaker of the dialogue act, student or tutor. Note that in each of these tables, the predictors are not just bigram probabilities, but *change* in that particular bigram probability from the *first to the fourth* dialogue within repeated-interactions tutoring.

Students Tending Toward Extraversion

Normalized Learning Gain =	Partial R^2	p
1.244 * $OEX_S \rightarrow FP_T$	0.228	< 0.001
-0.445 * $AYN_S \rightarrow R_T$	0.169	< 0.001
0.440 * $E_S \rightarrow QE_T$	0.139	0.001
0.359 * $QIS \rightarrow QF_T$	0.092	0.002
-0.298 * $AWH_S \rightarrow QO_T$	0.081	0.013
0.207 * $WAIT \rightarrow QP_T$	0.050	0.037
-0.226 * $QIS \rightarrow IT$	0.038	0.041
0.000 (intercept)		1.000
RSME = 50.97% of range in Normalized Learning Gain		

Table 7: Stepwise linear regression model for standardized Normalized Learning Gain in students scoring above the median in extraversion.

Students Tending Toward Introversion

Normalized Learning Gain =	Partial R^2	p
-0.447 * $QIS \rightarrow R_T$	0.262	0.003
0.371 * $QIS \rightarrow QP_T$	0.125	0.007
-0.331 * $QIS \rightarrow QQ_T$	0.092	0.015
-0.278 * $WAIT \rightarrow FPE_T$	0.083	0.018
0.384 * $AYN_S \rightarrow QQ_T$	0.067	0.010
0.288 * $ACK_S \rightarrow ET$	0.067	0.022
0.000 (intercept)		1.000
RSME = 60.89% of range in Normalized Learning Gain		

Table 8: Stepwise linear regression model for standardized Normalized Learning Gain in students scoring below the median in extraversion.

Several tutorial dialogue policy progressions were identified as statistically significantly associated with learning gain in both extraverted and introverted students. An increase in factual questions following extra-domain statements was associated with increased learning in students scoring above the median in extraversion, as was an increase in evaluative questions after explanations, an increase in the number of factual questions following information questions, and an increase in probing questions initiated after the conclusion of a sub-dialogue. On the other hand, extraverted students achieved a lower learning gain when tutors offered increasing reassurance after yes/no answers, asked more open questions after answers to WH-questions, or gave increasing instruction after an information question.

A similar number of tutorial dialogue policy progressions were identified as statistically significantly correlated with learning gain in introverted students. For these students, a higher learning gain was achieved when tutors followed more information questions with a probing question, more yes/no answers with a prompt for questions, or offered increasing explanation after acknowledgements. Students scoring below the median in extraversion achieved a lower learning gain when tutors offered more reassurance after information questions, more prompts for questions after information questions, or increasing elaborated positive feedback after pauses in the dialogue.

5 Discussion

This section examines the tutorial dialogue policy progressions that were identified as statistically significant to learning gain in these groups of students; recall that each feature represents a change over time in the probability that the second dialogue act follows the first. First we examine the extraverted student model, and then we examine the introverted student model. Dialogue excerpts illustrating these dialogue interactions are displayed in Appendix 1.

5.1 Extraverted Students

Students scoring higher in extraversion tend to be assertive, outgoing, and energetic (Goldberg, 1993). As the models show, these characteristics likely influence the extent to which particular dialogue policies are effective for supporting learning for extraverted students. For example,

the high energy nature of the extraversion personality trait may influence how dialogues transition. The model shows that students learned more when tutors progressed over time toward more positive feedback following extra-domain statements (*Extra-Domain Statement_S → Positive Feedback_T*) and toward more probing questions following pauses (*Wait → Probing Question_T*). Both of these bigrams indicate important transition points within dialogue. For the former, extra-domain statements represent off-topic utterances, whereas tutor positive feedback can only be task-related (if it were a positive response to an extra-domain statement, the response would also have been tagged extra-domain). For tutor probing questions following pauses, it is likely that extraverted students benefited from this adaptation over time because in being asked to reflect and explain their current understanding or goals, they may have been re-engaged. It should be noted that in general, asking students to self-explain can support learning (VanLehn et al., 1992).

Another example of a dialogue policy progression that emerged in the model and illustrates a widely known fact about tutoring is reflected in the *Information Question_S → Information_T* bigram, which when tutors progressed more toward this approach, is associated with decreased learning. Our prior work has shown that directing students what to do, even if they have just asked for such direction, is strongly associated with decreased learning (Mitchell et al., 2013).

Extraverted students tend to be assertive, and this characteristic influences how they make and interpret particular dialogue moves. An example of this can be seen within the model: when tutors progressed toward providing more reassurance after student yes/no answers, students learned less. This *Yes/No Answer_S → Reassurance_T* policy is likely a form of indirect feedback or politeness, both of which have been shown to be unhelpful, and sometimes harmful, to learning (Johnson and Rizzo, 2004), and this seems to be a particularly marked effect for extraverted students who may benefit more from direct evaluations of their answers. Another example of this indirect approach may be within the *WH-Question Answer_S → Open Question_T* tutor policy, whose increasing use over time was associated with lower student learning. Like reassurance, a follow-up question may be interpreted by extraverted students as an indirect in-

dication that the previous answer was incorrect, and a more direct approach may have been more helpful.

Finally, extraverted students tend to be talkative. This tendency is consistent with two of the model's findings regarding the helpfulness of particular types of tutor questions. Students tended to learn more when tutors progressed toward following student explanations with evaluative questions (*Explanation_S → Evaluative Question_T*). Although students' responses to evaluative questions (e.g., 'Do you understand?') are frequently considered to be inherently inaccurate, especially when students are first introduced to material, it may be the case that as students work on a task for an extended period of time, evaluative questions may become increasingly helpful. Another tutor questioning policy was also positively associated with learning gain for extraverted students: *Information Questions_S → Factual Question_T* involves the tutor answering a question with a question, potentially a very helpful strategy for talkative or highly social students.

5.2 Introverted Students

Students scoring lower in extraversion tend to be less talkative, more reserved, and more shy (Goldberg, 1993). This may result in introverted students being less outspoken about their understanding, and less likely to ask questions about misunderstandings. These characteristics affect the way that tutor choices impact student learning during tutoring. For example, when less talkative students ask information questions and tutors tend to provide more reassurance as time goes on, this *Information Questions_S → Reassurance_T* pair is associated with decreased student learning. It is possible that since introverts are less likely to speak up with a question, the "stakes" or importance of providing a direct answer may be higher for these students. Another dialogue policy progression that is not helpful for student learning is to provide elaborated positive feedback after a pause in dialogue (*Wait → Elaborated Positive Feedback_T*). Because pauses in the dialogue typically correspond to student task actions, it is possible that introverted students who are on the right track would benefit more from the tutor allowing them to continue working.

Introverted students also tend to describe themselves as shy or inhibited, which may be influential

in the apparent helpfulness of tutors' increasing their question prompts following student answers (*Answer Yes/No_S → Question Prompt_T*). This could be due to the fact that introverted students are prone to giving terse responses, and may need extra encouragement to ask questions if they are uncertain. Increasing the number of these prompts could increase the likelihood that more of the student's questions are voiced. Another helpful type of question for introverted students seems to be probing questions, even when they follow a student question (*Question Information_S → Probing Question_T*). A probing question is an indirect request for reflection, prompting the student to reconsider her approach; this has previously been shown to have a positive effect on learning gain (VanLehn et al., 1992).

6 Conclusion and Future Work

Adapting to personality during dialogue may substantially improve the effectiveness of both human-human interactions as well as interactions with dialogue systems. We have investigated the ways in which human tutorial dialogue policy progressions are associated with learning within a repeated-interactions dialogue study. The models indicate that depending on a student's tendencies toward introversion or extraversion, different dialogue policy progressions support higher learning. In particular, introverts may benefit from additional prompting and encouragement to speak their mind, while extraverts may benefit from being given opportunities to discuss their thoughts with a tutor.

While this study has focused on the extraversion facet of personality, future work may benefit from examining the other facets of the Big Five Factors: Neuroticism, Openness, Conscientiousness, and Agreeableness. How we may best design a tutorial dialogue policy around a more fully-featured model of the student's personality is an important research area. It will also be important to examine task actions closely in future analyses, as this may have significant effects on task-oriented dialogue system design in particular. Additionally, analyzing the intermediate sessions in order to capture a fuller picture of the interaction over time is a promising direction. Finally, examining tutor personality may also reveal important insight for the design of tutorial systems. It is hoped that these lines of investigation will lead to a next generation

of user-adaptive dialogue systems with increased effectiveness facilitated by their adaptation to personality traits.

Acknowledgements

The authors wish to thank the members of the LearnDialogue group at North Carolina State University for their helpful input. This work is supported in part by the Department of Computer Science at North Carolina State University and the National Science Foundation through Grant DRL-1007962 and the STARS Alliance, CNS-1042468. Any opinions, findings, conclusions, or recommendations expressed in this report are those of the participants, and do not necessarily represent the official views, opinions, or policy of the National Science Foundation.

References

- Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python*. O'Reilly Media, Inc.
- Susan E Brennan. 1996. Lexical entrainment in spontaneous dialog. In *Proceedings of ISSD*, pages 41–44.
- Sandra Carberry. 1989. Plan recognition and its use in understanding dialog. In *User Models in Dialog Systems*, pages 133–162. Springer.
- Justine Cassell and Timothy Bickmore. 2003. Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User Modeling and User-Adapted Interaction*, 13(1-2):89–132.
- Philip R Cohen, C Raymond Perrault, and James F Allen. 1981. Beyond Question-Answering. Technical report, DTIC Document.
- Greg Dunn, Jurgen Wiersema, Jaap Ham, and Lora Aroyo. 2009. Evaluating interface variants on personality acquisition for recommender systems. In *User Modeling, Adaptation, and Personalization*, pages 259–270. Springer.
- Kate Forbes-Riley and Diane Litman. 2007. Investigating human tutor responses to student uncertainty for adaptive system development. In *Affective Computing and Intelligent Interaction*, pages 678–689. Springer.
- Eric N Forsyth and Craig H Martell. 2007. Lexical and discourse analysis of online chat dialog. In *Semantic Computing, 2007. ICSC 2007. International Conference on*, pages 19–26. IEEE.
- Lewis R. Goldberg. 1993. The structure of phenotypic personality traits. *American Psychologist*, 48(1):26–34.
- Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in personality*, 37(6):504–528.
- Barbara J. Grosz. 1983. TEAM: A Transportable Natural-language Interface System. In *Proceedings of the First Conference on Applied Natural Language Processing*, pages 39–45, Santa Monica, California. Association for Computational Linguistics.
- Eun Young Ha, Christopher M Mitchell, Kristy Elizabeth Boyer, and James C Lester. 2013. Learning Dialogue Management Models for Task-Oriented Dialogue with Multiple Communicative Channels. In *Proceedings of the 14th Annual SIGDIAL Meeting on Discourse and Dialogue*, pages 204–213, Metz, France.
- Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Reutemann, and Ian H. Witten. 2009. The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, 11(1).
- Oliver P. John and Sanjay Srivastava. 1999. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, 2:102–138.
- W Lewis Johnson and Paola Rizzo. 2004. Politeness in tutoring dialogs: "Run the factory, thats what Id do". In *Intelligent Tutoring Systems*, pages 67–76. Springer.
- Willem JM Levelt and Stephanie Kelter. 1982. Surface form and memory in question answering. *Cognitive psychology*, 14(1):78–106.
- Diane J Litman and James F Allen. 1987. A plan recognition model for subdialogues in conversations. *Cognitive Science*, 11(2):163–200.
- Francois Mairesse and Marilyn A Walker. 2010. Towards personality-based user adaptation: psychologically informed stylistic language generation. *User Modeling and User-Adapted Interaction*, 20(3):227–278.
- Jeffrey D. Marx and Karen Cummings. 2007. Normalized change. *American Journal of Physics*, 75(1):87.
- Christopher M Mitchell, Kristy Elizabeth Boyer, and James C Lester. 2012. From strangers to partners: examining convergence within a longitudinal study of task-oriented dialogue. In *Special Interest Group on Discourse and Dialogue*, pages 94–98.
- Christopher M Mitchell, Eun Young Ha, Kristy Elizabeth Boyer, and James C Lester. 2013. Learner characteristics and dialogue: recognising effective and student-adaptive tutorial strategies. *International Journal of Learning Technology (IJLT)*, 8(4):382–403.
- Ani Nenkova, Agustin Gravano, and Julia Hirschberg. 2008. High frequency word entrainment in spoken dialogue. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies*, pages 169–172. Association for Computational Linguistics.
- Kate G Niederhoffer and James W Pennebaker. 2002. Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, 21(4):337–360.
- Martin J Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, 27(2):169–190.
- Richard Power. 1974. A computer model of conversation.
- Byron Reeves and C Nass. 1997. *The Media equation: how people treat computers, television, and new media*. Cambridge University Press.
- Adriana Tapus and Maja J Mataric. 2008. Socially Assistive Robots: The Link between Personality, Empathy, Physiological Signals, and Task Performance. In *AAAI Spring Symposium: Emotion, Personality, and Social Behavior*, pages 133–140.
- Alexandria Katarina Vail and Kristy Elizabeth Boyer. In press. Identifying Effective Moves in Tutorial Dialogue: On the Refinement of Speech Act Annotation Schemes. In *Proceedings of the 12th International Conference on Intelligent Tutoring Systems*, Honolulu, Hawaii, USA.
- Kurt VanLehn, Randolph M Jones, and Michelene TH Chi. 1992. A model of the self-explanation effect. *The Journal of the Learning Sciences*, 2(1):1–59.
- Daan Verbree, Rutger Rienks, and Dirk Heylen. 2006. Dialogue-act tagging using smart feature selection; results on multiple corpora. In *Spoken Language Technology Workshop*, pages 70–73. IEEE.
- Wolfgang Wahlster and Alfred Kobsa. 1989. *User models in dialog systems*. Springer.

Extraverted Student Dialogue Excerpt	
<i>Extra-Domain Other → Positive Feedback</i>	STUDENT: I haven't really done this before. TUTOR: You're doing well!
<i>Yes/No Answer → Reassurance</i>	STUDENT: Yes. TUTOR: Actually, I was wording the question incorrectly...
<i>Explanation → Evaluative Question</i>	STUDENT: But it still prompts for 3 or 4... TUTOR: Yes; does that make sense from what you learned about sequential program flow?
<i>Information Question → Factual Question</i>	STUDENT: What did I do wrong? TUTOR: What is your Scanner's name?
<i>WH-Question Answer → Open Question</i>	STUDENT: Previous. TUTOR: Why did <code>previousChoice</code> get assigned a value?
<i>Wait → Probing Question</i>	TUTOR: What do you think about your program's behavior?
<i>Information Question → Instruction</i>	STUDENT: There wouldn't have been any output? TUTOR: Yeah, but more than that, the program would report an error.
Introverted Student Dialogue Excerpt	
<i>Information Question → Reassurance</i>	STUDENT: So the previous answer needs to be stored as a part of <code>PlayerInput2</code> ? TUTOR: That would work fine.
<i>Information Question → Probing Question</i>	STUDENT: That's not what I want? TUTOR: Do you really want 'or'?
<i>Information Question → Question Prompt</i>	STUDENT: So I need an <code>else if</code> for every <code>if</code> statement? TUTOR: Do you have any questions?
<i>Wait → Elaborated Positive Feedback</i>	TUTOR: Nice, you compiled the code.
<i>Yes/No Answer → Question Prompt</i>	STUDENT: No, I got it. TUTOR: Any questions so far?
<i>Acknowledgement → Explanation</i>	STUDENT: Okay. TUTOR: When Java gets to the <code>nextLine()</code> , it will stop.

Appendix 1: Dialogue excerpts illustrating the dialogue interactions emergent as significant in the analysis. All excerpts originate from Lesson 4, at the end of the series of dialogue sessions.