5-2020

Question-Asking in Conversational Tasks : A Gender Comparison

Courtney Bell

Abstract

This corpus analysis used unexplored gender disparities in a previously collected speech corpus, the Montclair Map Task Corpus (MMTC) to examine stylistic patterns in utterance goals and question-asking. The MMTC utilized a collaborative map task to explore effective communication, gender differences in communication, and patterns in conversational style. The current study built upon those goals by engaging in a more in-depth analysis of the MMTC conversations. To accomplish this, the previously transcribed conversations were coded to determine turn goals. Turns that were observed as questions were further coded to determine question objectives. Based on prior research on affiliative and assertive speech, it was hypothesized that males would utilize challenge questions more than females and that females would utilize more information seeking and clarification questions. Overall, proportions of turns and questions did not vary based on pair sex or talker sex. Relationships between performance question type suggested no meaningful patterns based on pair or talker sex, however some trends were found in clarification questions, where same-sex male pairs and male talkers overall showed negative relationships with performance, while females showed positive relationships with performance.

MONTCLAIR STATE UNIVERSITY

Question-asking in conversational tasks: A gender comparison

By

Courtney Bell

A Master's Thesis Submitted to the faculty of

Montclair State University

In Partial Fulfillment of the Requirements

For the Degree Of

Master of Arts

May 2020

College/School CHSS

Department Psychology

Thesis Committee:

[redacted]
Thesis Sponsor: Dr. Jennifer Pardo

[redacted]
Committee Member: Dr. Alan Pehrson

[redacted]
Committee Member: Dr. John Paul Wilson

QUESTION-ASKING IN CONVERSATIONAL TASKS: A GENDER COMPARISON

A THESIS

Submitted in partial fulfillment of the requirements

For the degree of Master of Arts

By

COURTNEY BELL

Montclair State University

Montclair, NJ

2020

Acknowledgements

I would first like to thank Montclair State University for the opportunity to complete my thesis and graduate studies. I would also like to thank all the research assistants at the Speech Communications Lab at Montclair State University for their tireless work coding the transcriptions. Alexander Bachert, Brittany Nace, Daniel Rynerson, Gina Lisa, Jacqueline Long, Jessica Jean, Nina Flores, and Ruchael McNair. Without their help, I would have never completed this thesis on time. I would like to express my gratitude to Dr. Alan Pehrson and Dr. John Paul Wilson for their support and advice as my thesis committee members. I would finally like to thank my thesis sponsor, Dr. Jennifer Pardo. Her research is what inspired this study and I would have been lost without her guidance.

Table of Contents

Question-Asking in Conversational Tasks: A Gender Comparison

As a social species, humans rely on effective communication to relay important information to others. Because these interactions are such an integral aspect of our everyday lives, it is essential to understand the dynamics of communication, where disparities may exist, and ways in which communication can be improved. As a means of studying communication in the past, researchers used various types of corpora. Some of these include transcriptions from news reports, recorded phone calls, book excerpts and recorded work meetings. Alternately, researchers have utilized corpora that involve collaborative tasks as a method of studying communication, not only because they create a setting for participants to naturally interact, but also because collaborative tasks are a large portion of the interactions we partake in, whether it be a work collaboration, a group project for school or with family at home. While many of these collaborative tasks are utilized in information retrieval and workplace projects (Fidel, Bruce, Peitersen, Dumais, Grudin & Poltrock, 2000), others have been used by researchers to explore potential differences in communication preferences based on gender.

Gender Differences in Communication

Prior studies exploring communication dynamics in collaborative tasks have found gender differences in communication styles. While the verbal abilities of men and women do not differ (Hyde & Linn, 1988), there are reports of communication contrasts between genders. Communication dominance studies have found that men challenge their partners more often than women when working on a collaborative task (Adams, 1980) and that men use more assertive speech, while women use more affiliative speech (Leaper & Ayres, 2007; Mulac, Bradec & Gibbons, 2001). Research comparing talkativeness in men and women found that men speak more words and use more speaking turns than women (Leaper & Ayres, 2007; Wood, 1966).

Studies investigating interruptions in spontaneous speech found that based on the type of interaction, men use more intrusive interruptions than women (Anderson & Leaper, 1998). Workplaces are encouraged to diversify their teams, which has shown to improve group performance. However, when the team dynamic changes to a competitive nature, performance in mixed gender groups decreases, while performance in same sex groups remains the same as individual performance. Ivanova- Stenzel and Kübler (2010) created a study to test this by comparing gender composition and performance using team-shared wages as an incentive for performance when competing against another team. An analysis of performance found that individual performance was on par with same sex teams whether they were competing against same-sex men or same-sex women teams. In mixed sex groups, individual performances were greater than team performance showing a disconnect in mixed group competitive settings (Ivanova- Stenzel & Kübler, 2010).

According to a limited literature review by Smith-Lovin & Robinson (1992), these gender differences are due to a social status disparity, where men are granted higher social status and women are granted lower social status. Several gender dynamics perspectives compiled in this review attempt to explain this disparity. The socialization perspective reasons that historically, women bore and raised the children. Women were then predisposed to speak using social-emotional skills and had more experience with family coordination. These interactions were observed by their children who passed on the tradition. The cultural perspective points out the natural gender segregation children use when playing games and explains that though this is learned through social cues, it further complicates the dynamic of mixed gender conversations. Finally, the affect-control theory states that people enter a social situation with their own unconscious assessment of their power and status. As the interaction progresses, each person

makes judgments about others' power and status, which can alter the dynamic of their interaction. Other researchers have proposed similar gender dynamic theories. The socialization theory explains gender dynamics in adult communication as a continuation of what was learned in gender-segregated activities as a child. (Maltz & Borker, 1982). The social constructionist theory explains that situational needs are the prominent factor in the dynamics of communication but underlying elevated power and position of men and the subordinate status of women play a role as well (Hall, 2006a). The biological explanation suggests that evolutionary pressures for men to be more aggressive/assertive and women to be more nurturing/ affiliative are the cause of gender discrepancies in communication. Brain imaging studies also found that women's brains may have more developed language processing regions and a higher capacity for language processing that begins earlier in life (Anderson, 2006). Many experts believe that these theories are to be used in conjunction.

Smith-Lovin and Robinson, interested in testing these theories, randomly assigned participants to complete a task in either same sex or mixed sex groups. Number of turns, average length of turn, normal transitions, overlapped transitions, and interrupted turns were measured and compared based on the sex of the talker and the sex of the other members of the group. Interruptions were used as a measure a dominance. Analyses of these conversations found effects of pair gender on all these categories. Significant differences were observed between mixed sex pairs and same sex pairs only. The researchers concluded that none of these theories could explain these gender dynamics on their own, so a mixed-perspective approach that includes using these theories in conjunction with one another is best when evaluating gender dynamics in conversations.

One of the issues in deciding which theories may be at play, and even if these gender differences are generalizable, is the inconsistencies in the methodologies across studies. Leaper and Ayres (2007) conducted a meta-analysis to try to understand where issues in methodologies might affect results of studies. They also attempted to match or rule out conversational dynamic theories that may explain the results. This exhaustive meta-analysis including 149 studies reported from 1968 to 2004 exploring talkativeness in men and women as well as preferences towards assertive and affiliative speech. Their inclusion criteria were as follows: studies investigating gender effects on adult language, studies using quantitative observational methods, and studies that were published in journals or books.

Leaper and Ayres identified several potential modifiers in their assessment of talkativeness, assertive speech, and affiliative speech across multiple studies. They also compared operational definitions for talkativeness, assertive speech, and affiliative speech. An analysis of talkativeness found that men were more talkative than women. Significant modifiers of this were the operational definition of talkativeness (number words or utterances, rate of time speaking, mean length of utterance (MLU), duration of time talking, total turns, and total statements/speech acts), relationship between speakers (all groups were more talkative with strangers, but men had the largest variability between strangers and those they are familiar with), group size (men were more talkative in dyads), and group gender (men were more talkative in mixed gender groups than same sex groups). An analysis of affiliative speech found that women used more affiliative speech than men. Significant modifiers for this were operational definition (supportive speech, active understanding, agreement, acknowledging, general social-emotional speech), age (undergraduates had the most gender disparities), relationship (greater disparities in strangers), gender composition (greatest differences among same gender, not significant in

mixed), activity type, year of publication, and length of conversations. The analyses of assertive speech found that men used more assertive speech than women, but the effect size was weak. Significant modifiers for this were operational definitions (directive, giving information, suggestions, criticism, disagreement, or task-oriented speech), participant age(undergraduates had the greatest disparities), participant relationship (strangers produced the most significant findings), gender composition, and if the researcher was present (Leaper & Ayres, 2007). One the greatest limitations of this metanalysis is that many of the studies they used either did not report socioeconomic status and cultural background information or were not diversified enough to generalize the results.

It is important to keep the significant modifiers they found in mind when designing studies and interpreting results. Based on the Leaper & Ayres' findings, it would be predicted that a study using participants in dyads with no familiarity to one another would result in more talkativeness and assertive speech in men and more affiliative speech in women. These findings should be consistent across all different types of speaking turns, including questions, answers, and in direction-giving. To test this, specific affiliative and assertive types speech should be recorded and measured.

Anderson & Leaper (1998) also conducted a meta-analysis analysis including 43 studies reported from 1965 to 1996 exploring existing gender comparative conversational tasks. This meta-analysis found potential limitations in the way conversational studies, especially those regarding interruptions, are conducted. Interruptions are studied because it is thought that they show a dynamic of apparent power. One limitation is variability in the operational definitions of interruptions. The prior studies were categorized into one of three groups based on their operational definitions of interruptions; interruptions were undefined or only broadly defined,

backchannels (brief vocal responses by listener that does not result in the taking of conversational floor) and minimal responses were specifically excluded, or interruptions were defined as an intrusive act (distinctly different from overlaps).

The meta-analysis revealed that men interrupted significantly more than women, however, they found a weak effect size. Studies defining interruptions as an intrusive act had the most significant differences between genders with the most meaningful effect sizes. Multiple factors were also identified as potential moderators. In exploring the effect of author gender, studies with women authors reported most significant gender differences and effect sizes. Analyses also found that studies using groups of three or more resulted in more significant results over pairs, that men interrupt more with strangers than those they are familiar with, and that these results have a much larger magnitude in naturalistic settings (Anderson & Leaper, 1998).

Bilous and Krauss (1998) conducted an important study that investigated conversational differences based on partner sex. Their results captured some of the nuances of conversational dynamics that can be overlooked in meta-analyses. Bilous and Krauss assessed the accuracy of the male dominance hypothesis, which states that the gender disparities in communication stem from societal perspectives that men are the dominant gender, which is mirrored in the way individuals communicate with members of the opposite sex (Thorne & Henley, 1975). They were also interested in speech accommodation theory, which states that the nature of the relationship between the speaker, along with other factors, determines the type and frequency of accommodations, whether they be convergent or divergent. Convergence refers to an increase in communicative similarity between a pair of talkers and divergence refers to a decrease in communicative similarities between a pair of talkers. In their study, 60 participants were

randomly assigned to groups of 4, and further broken into pairs. Each participant had two conversational tasks one in a same sex pair and another in a randomly assigned mixed sex pair with a remaining group member. Each conversation was over 10 minutes and the last 8 minutes were transcribed. The tasks required an agreed-upon judgement/deliberation from the pair on one of two scenarios to either allocate scarce resources to people who want it or to rate articles on importance for a purpose. The dependent variables included were speech productivity (the number of words uttered excluding back-channels), interruptions (the frequency of overlaps in speaking turns, excluding those due to back-channels), short pauses (intra-utterance silence 1 second or less), long pauses (intra-utterance silence longer than 1 second), and listener back-channels.

Bilous and Krauss analyzed the relationships between multiple dependent variables and the independent variables, which were gender, gender composition (mixed or same sex), and discussion topic. They found an overall significant difference between same and mixed sex pairings across multiple measures. Women had more convergence in length of utterance than men. Women produced more words than men in same sex pairs but reduced their words in mixed sex pairs. Men had longer utterance length than women in same sex and mixed sex pairings, however, the differences in mixed sex pairs were smaller. Women interrupted more than men in same sex pairs but were equal in mixed sex pairs. Women overall converged more in same and mixed sex pairs. Participant gender and dyad gender composition were significant modifiers. Overall, the male dominance hypothesis fails to account for the observable behaviors in this study because both men and women showed accommodation on different attributes. Instead, Bilous and Krauss concluded that this dynamic is nuanced, where men maintained their power status in some instances by not making changes between same and mixed sex pairs, but women

had less speech and interruptions in mixed sex pairs than in same sex pairs and were the only ones to show divergence. This could be that women change their communication habits around men based on nuanced power dynamic.

Overall, results throughout these studies suggest a trend towards gender differences of some kind. Exhaustive meta-analyses such as Leaper and Ayres (2007) and Anderson and Leaper (1998) found potential modifiers for effects in research regarding gender communicative differences, which are important to consider when critically evaluating the literature. Limited literature reviews, such as Smith-Lovin & Robinson (1992) create a strong theoretical framework, however, only consider studies that focus on the specific theories they are interested in. Excluding studies that disagree with theories can create a false narrative, which is why these types of literature reviews should be viewed with caution. As society is consistently evolving, so should our theories about the dynamics within this society. Many of the studies used to understand gender dynamics in communication are at least 20 years old. In that time the social status disparity between men and women has changed and it is likely that these theories need to evolve with that. It is possible that what was observed in Bilious and Kruss (1998), that gender dynamic does not follow rigid conversational rules, but is highly complex and nuanced.

It is also important to note that Anderson's (2006) brain imaging studies should also be viewed with caution. These studies often come with a variety of issues from lack of power to inflated alpha levels. These studies also make inferential leaps about the implications of their findings. More experimental brain imaging studies suggested by Wang, Kessels, and Hu (2014) could be utilized in understanding more about social dynamics and dominance rather than making large leaps based on the size of brain regions is. They suggest manipulating circuitry within the prefrontal cortex. In their exhaustive literature review they determined that this region

is related to dominance and social status in rodents, which translates to humans due to the similar nature of brain structures and functioning. While these varied studies all have found disparities in the way men and women communicate, several situational factors across this research made it difficult to determine more specific patterns until now.

Montclair Map Task Corpus

The Montclair Map Task (Pardo, et al., 2019) was designed to examine effective communication, gender differences in communication, phonetic convergence in spontaneous speech, and patterns in conversational style. In this study, 96 participants were placed in pairs based on gender: male only (16 pairs), female only (16 pairs) or mixed pairs (16 pairs). Each pair was given a collaborative task that involved reconciliating variations in the composition of landmarks across 6 pairs of maps. Half of the landmarks were shared on both maps, while the other half differed. The goal was to find the locations and names of the missing landmarks without seeing their partner's map through verbal-only communication, and to draw markers for the missing landmarks on their map. These conversations were recorded, transcribed, and analyzed.

Initial analyses of the corpus indicated that males spoke more words than females, overlapped their speech more than females in same-sex pairs only, and had more turn exchanges. All gender pairs also increased in performance across map pairs. An evaluation of participants' map task performance found that only six out of the 576 maps had missing landmarks and only one map had an extra landmark. Further analyses of performance examined accuracy of placement of each landmark's location on the map. The average score on each map was 90% with no significant difference in scores across male and female participants.

An evaluation of task completion time found that it took participants an average of 32 minutes to complete the entire session (all 6 maps) with times ranging from 16-62 minutes. Time to complete the task was weakly correlated with performance, $r(46) = 0.28$, $p = 0.05$. An analysis of variance on time to complete the map task found no significant differences based on pair sex. Speaking frequencies were evaluated based on talker sex. Males spoke more than females—12% more in same sex pairs and 26% more in mixed pairs. Males also exchanged more turns than females in both same and mixed sex pairs and overlapped more in same sex pairs only.

In evaluating communication efficacy, it was determined that men spoke more and faster than women, but an analysis of partner map performance found that words spoken by male pairs were negatively correlated with partner map performance, $r(30) = -0.33$, $p = 0.06$. In mixed sex pairs, there was a positive correlation between words spoken and partner map performance, $r(30) = 0.48$, $p = 0.008$. In female pairs, there was a strong positive correlation between words spoken and partner map performance, $r(30) = 0.60$, $p = 0.0003$. Number of words spoken by same sex male pairings also dramatically reduced from map 1 to map 2. This means that when a female in a same sex pair spoke more it was strongly related with their partner performing better on the map task, but when a male in a same sex pair spoke more it was related to their partner performing worse on the map task.

In order to further understand these patterns, it is necessary to conduct more detailed analyses of the content of these corpus conversations. To investigate the observed gender disparity, the current study attempted to pinpoint an aspect of communication that may contribute to this gender difference using the Montclair Map Task's preexisting data set and an expanded coding scheme. The Montclair Map Task Corpus (MMTC) is appropriate to study this

because the utterances of the talkers were more balanced than those of similar corpora, such as the HCRC Map Task. To expand on the MMTC coding scheme, the current study used comprehensive coding of each turn of the first two maps, which is where the most variability was observed. This in-depth coding scheme required that the coders first established what the talker was trying to accomplish in each turn. Turns that aimed at asking a question were then further coded into question type.

Questions in Discourse

While many aspects of gender differences in communication have been investigated, conversational goals, specifically regarding the types of questions asked, have received limited attention. This circumstance hinders an understanding of effective communication because the question-answer form of communication is one of the most prevalent constructs of a conversation (Graesser, Baggett & Williams, 1996). Important factors in understanding question-asking are the goal and format of the question. Multiple studies previously mentioned found differences in affiliative and assertive speech between males and females (Leaper & Ayres; Mulac, Bradec & Gibbons, 2001). This should imply that there are different types of affiliative and assertive questions asked. As explained by Stivers and Enfield (2010) when describing how they came up with their question-response sequences, it is important that a researcher looks at what response a question is trying to elicit or what the talker's goal is when asking a question to properly categorize it.

A psycholinguistics program in the Netherlands was able to compile a question-asking coding scheme applicable to 10 different languages based on the consistency of their questions (Stivers & Enfield, 2010). While some of their categories are not applicable to a collaborative task, their classifications of initiation of repair, clarification, and information seeking would be

relevant to a collaborative task where gender disparities in question-asking are studied. These considerations served as the basis of the current study's coding scheme, which will code questions as *challenges*, *clarifications*, *information-seeking*, and *repetitions.*

In a study exploring gender and race power dynamics in conversations, Adams (1980) measured dominance using verbal *challenges*. Adams defined challenges as short statements of disagreement with another speaker. Because assertive speech is characterized by dominance and challenges have previously been used to measure assertive speech (Leaper & Ayres, 2007), challenge questions that assert dominance would be considered a type of assertive speech. While challenges have been measured within full corpora in the past, they have not been measured within questions.

Clark and Krych (2004) examined whether speakers used *clarification* utterances while speaking to an addressee, how they do so, and if and how they make corrections or alterations in their speech. These clarifications were defined as understanding checks and were characterized by the talker's attempt to make a message clearer. Since affiliative speech is characterized by active understanding (Leaper & Ayres, 2007), clarification questions would be a type of affiliative speech. Clarifications elicit understanding between both the speaker and listener and do not seek unequal power dynamics or dominance.

According to Reichel, Porner, Nowak and Cole (2015) when describing their measurements for cooperation in participants working on a collaborative task, "Cooperative behavior includes providing relevant and sufficient information". *Information seeking* questions are an affiliative speech because they are designed to give the partner the speaking floor and encourages cooperative behaviors, which is a characteristic of affiliative speech (Leaper and Ayres, 2007).

Levelt and Kelter (1982) explored question-asking and found consistencies in question formats and the types of answers they illicit. They also found that questions in the form of a *repetition* are used when a speaker is presented with novel information. One explanation for this is that repeating information is a way to remember the information presented to the listener. Levelt and Kelter tested this theory and found that those who repeated information in the form of a question had improved performance on a memory test.

The Current Study

The current study explores gender differences found in the MMTC by focusing on question styles in conversational interactions. The first two trials (out of six) in each task contained the starkest contrasts between male same-sex pairs, female same-sex pairs, and mixed pairs as well as the lowest performance scores. This could imply that most communication errors took place while the pairs were working on the first and second trials of the task. Because this study aims to understand where these differences occur and what communication errors can lead to diminished performance for either partner, only the first two trails were evaluated. Before question-asking could be evaluated, the speaking turns in the MMTC needed to be categorized into different types of speech to properly identify which utterances were questions. Some speaking turns included multiple types of speech, so this coding scheme categorized both the type of speech at the beginning of a speaking turn and at the end of that speaking turn.

Speaking turns in the MMTC were categorized into one of six different types of speech. Questions were one of the most common types of speech and defined as “the talker asked a question”. Answers were also quite common because they were in response to questions. They were defined as “the talker answered a question”. Directions were also very common in this type of task because participants were required to give physical directions regarding landmark

locations on their maps. They were defined as “the talker was providing directions for their partner”. Comparisons were also necessary in this task for both partners to understand which landmarks were missing from each of their maps. They were defined as “the talker made a statement comparing their map to their partner’s map”. Repetitions, while not very common in this task, are often used as a tool to show attentiveness to their partners’ speaking turn. Repetitions have various subtypes (semantic, syntactic, mirror, etc.) that could be explored in the future, however in this study, they were defined as “the talker repeated something their partner uttered in the preceding turn, verbatim”. Speaking turns that did not fit into any of these types of speech were categorized as “other”.

The question categories were further categorized by the speakers’ goal in each question, informed by previously discussed research and customized for the MMTC corpus. *Challenge* questions were defined as “the talker is asking a question that conveys uncertainty about the accuracy of the information their partner previously conveyed”. *Clarification* questions were defined as “the talker is asking a question to better comprehend previously discussed information”. *Information* questions were defined as “the talker is asking a question to seek new information”. *Repetition* questions were defined as “the talker mirrored a question they previously asked”. *Direction* question were defined as “the talker asked a question about the location of a previously discussed landmark”. Comparison questions were defined as “the talker is asking a question to compare their map information to their partner’s map information”. All the questions that did not fit into any other category were marked as “other”.

Based on prior research regarding affiliative and assertive speech in men and women, it is hypothesized that males, specifically in same sex male pairs, will utilize challenge questions more than female pairs, while female pairs will utilize more information seeking and clarification

questions and that these conversational patterns will be consistent across talker sex. Based on the prior gender disparities found between words spoken and map task performance it is also hypothesized that proportions of question types will be related to map performance based on pair and talker sex. To test this, conversations from the first two maps of the MMTC were coded at each speaker turn to assess variations in conversational style between same sex male pairs, same sex female pairs and mixed sex pairs.

Methods

Montclair Map Task Corpus

The current study is based on an original coding of a previously recorded speech corpus. The sections below explain the procedures used in collecting the corpus and describe the system used to complete the novel coding. For more information on The Montclair Map Task Corpus, see report by Pardo, et al. (2019)

Participants

Participants (96 total, 48 female) were recruited from Montclair State University between March 2013 and May 2014. Before starting the study, participants completed an IRB-approved (1213) informed consent form and a demographic questionnaire. Participants were excluded if they indicated that they were not native English speakers or if they were hard of hearing. They were randomly selected into gender pairs -male only pairs [MM], female only pairs [FM], and mixed sex pairs [MX] and given a conversational task. There were 96 total participants, 48 of which were male and 48 were female. Participants were compensated $20 for their time.

A demographics questionnaire collected data about the participants' geographical history, ethnicity, race, income level, language background, age, and sex. Participants varied in age from 18 to 38 years old, with an average age of 21 years old (SD = 2.8 years). Nearly 90% of all the

participants were born in and lived most of their lives in the New York/ New Jersey/ Pennsylvania area. Those that were not from the NY/NJ/PA area, all have lived in New Jersey for five or more years. When asked about ethnicity, 20 participants identified as Hispanic/Latino when asked to indicate whether they identify as Hispanic/Latino or not (2 participants left this blank). When asked about their race, 61 participants identified as White, 15 identified as Black, four identified as Asian, two as American Indian or Alaskan Native, one as Black/White, one as Native Hawaiian/White, one as Other and one who wrote in Latina (eight participants did not answer this question).

Language backgrounds were assessed by asking participants to list any languages they knew in addition to English, the year they began learning that language and their proficiency in that language (fluent, basic conversational, or school setting only). 26 of the participants indicated they are fluent in a second language and 19 did not answer this question. Many of those that indicated fluency in a second language listed the year of their birth as the year they began to learn the language, so it is likely that these individuals are bilingual.

Procedures

Participants were randomly assigned pairs: 16 female-only (FM), 16 male-only (MM), and 16 mixed sex (MX). The participants were all unfamiliar with each other and were unable to see one another throughout the task due to a divider placed in the room. Each participant sat in a sound-proof booth at a small table and used a pencil to draw on the maps as they completed the task. Participants wore microphones connected to computers situated outside the booth. Conversations were recorded using SoundStudio software.

Conversational Task

The HCRC Map task (Anderson et al., 1991) was used as a model for the Montclair Map Task. The Montclair Map Task (MMT) used six pairs of maps that consisted of routes around specifically labeled landmarks. Each participant received six maps that corresponded with their partner's maps. Each of these map pairs contained five landmarks that were also present on the partner's map (shared) and five landmarks that only appeared on their map. Additionally, the routes around the landmarks had a starting and finishing point that were the same on both maps. Each map pair had a different set of routes and landmarks. In total, there were 79 landmarks, with some appearing on multiple maps.

The aim of this task is for the pair to end up with identical maps by finding the landmarks they are each missing and draw markers for their locations on their maps. The instructions inform the participants that their goal is to find the names and locations of the five missing landmarks that are on their partner's map and to draw and label them as accurately as possible. Because the participants are unable to see their partner's map, this task will create an environment where the participants are required to converse in a naturalistic setting. The routes between landmarks in each map pair are identical with a starting and ending point to assist the pairs in having a shared frame of reference. See Appendix A for a sample set of maps used in the MMT

Task Accuracy

Map task accuracy was scored using two independent raters. Each rater examined the location of each landmark on the map and placed it into one of five categories: Correct location, adjacent location on the same side of the path, nearby location on the wrong side of the path, distant location (more than two inches away in any direction), and missing. A landmark correctly placed is worth 20 points, adjacent is worth 15 points, nearby is worth 10 points, distant

is worth five points, and missing is worth no points. With five landmark locations on each map, the maximum score a participant can receive on each map is 100. Rater's scores were averaged and used as the participants' final scores.

Transcription Protocol

Using Praat software, trained research assistants segmented and transcribed all the recorded conversations between each of the 48 pairs. Each recording indicated the time at which each utterance (individual talker completing a communicative objective; Sacks, Schegloff, & Jefferson, 1974) began and ended, which map the pair was working on (1-6), which talker was speaking, inter-turn intervals (gap of time between the ending of one talker's turn and the beginning of the other talker's turn), backchannels (utterances spoken by the other talker during the main talker's turn that do not start a new speaking turn, e.g., "uh-huh," "okay," etc.), overlapping speech (both talkers speak simultaneously), and landmark label phrases. Praat textgrid files stored the data from segmentation, transcription, and labeling, and conversational analyses used data extracted from these files. See Appendix B for a transcription example.

MMTC Coding Protocol

The MMT results indicated gender differences in multiple aspects of communication. In order to obtain a more in-depth understanding of the dynamics behind these differences, the MMTC was further coded. First, trained research assistants further categorized the interactions immediately after and right before each inter-turn interval (ITI) in the first two map attempts, because these maps had the most gender differences in previous analyses. These interactions were coded as either a question (the talker asked a question), an answer (the talker answered a question), a repetition (the talker repeated something their partner uttered in the preceding turn), direction (the talker was providing directions for their partner), comparison (the

talker made a statement comparing their map to their partner's map), or other (anything that doesn't fit into the other categories).

To explore potential differences in the types of questions being asked, further coding was performed on the interactions labelled questions. These questions were categorized based on the goal of the questions and loosely follows the coding scheme suggested in Stivers & Enfield's (2010) coding scheme for question-response sequences. These questions were categorized as either a clarification (when the talker is asking a question to better comprehend previously discussed information), information (when the talker is asking a question to seek new information), comparison (when the talker is asking a question to compare their map information to their partner's map information), challenge (when the talker is asking a question that conveys uncertainty about the accuracy of the information their partner previously conveyed), repetition (the talker mirrored a question they previously asked), direction (the talker asked a question about the location of a previously discussed landmark), or other (any type of question that does not fit into another category).

Data Treatment and Analyses

All data was analyzed using R statistical platform (R version 3.6.2 R Development Core Team, 2016). Analysis alpha levels were all set at .05 and were exploratory in nature. To investigate patterns of conversational style in the MMTC, frequencies of each turn type were compared based on pair sex, map, and talker sex. Male pairs overall used significantly more speaking turns than females in the entire corpus [$p < 0.03$], 12% more in same-sex pairs and 26% more in mixed sex pairs. There was also a variation across individual pairs and between maps 1 and 2 in number of turns. Female participants used 1,525 speaker turns in Map 1 and 1,492 speaking turns in Map 2, while males used 1,792 speaking turns in map 1 and 1,438 speaking

turns in map 2. This circumstance could show skewed results for male pairs or pairs with more turns since they would have had more opportunities than the other pairs to use each turn type. Frequency data also limits the full understanding of turn styles by missing out on information about the overall number of turns. Proportions of turns, however, are easily translated to percentages out of 100 and give a better picture of the overall usage of turns without skewing data. To normalize for differences in overall numbers of turns, the frequency of each turn type was converted into proportions for each map in all further analyses.

Four total statistical tests were used to analyze patterns in turn and question types. To assess effects of pair sex on turn type usage across maps, two three-way split plot ANOVAs included two repeated measure factors, map (1 versus 2) and proportions of coding type (Turn code categories or Question code categories), and one between-subject factor for pair sex (MM, FM, or MX). To assess effects of talker sex regardless of pairing, the two three-way split plot ANOVAs included two repeated measure factors, map (1 versus 2) and proportions of coding type, and one between-subjects factor for talker sex (Male versus Female).

To investigate patterns between the use of question types and performance, correlations were assessed using Pearson's R. Correlations between the proportions of each question type (overall, by pair sex, and by talker sex) and map performance (own and partner) for each map were determined using the cor.test function in R. In total, 144 correlation tests were performed however, 39 had too few observations to produce a result.

An observed pattern in the clarification questions was explored using two linear mixed regression analyses. The goal of the first analysis was to test whether map performance could be predicted by the use of clarification questions and pair sex. A second linear mixed regression analysis was used to determine the reliability of talker sex as a predictor. The goal of this was to

test whether map performance could be predicted by the use of clarification questions and talker sex. Both analyses were obtained through the lme function in R.

Results

To test the hypotheses that males will utilize more assertive questions such as challenges and females will utilize more affiliative questions such as clarifications and information-seeking, the turn coding first needed to be analyzed to determine whether there were stylistic differences in turn type usage based on pair or talker sex. When it was determined that the use of questions did not vary based on sex, stylistic differences in question type based on pair and talker sex were examined. To look closer at patterns of question styles, relationships between the use of each question type and map performance were determined and fit into a regression model.

Turn-by-turn Coding of Conversational Style

Frequencies of each turn type were converted into proportions of turns using the frequency of that turn type as the numerator and the total number of turns as the denominator in an equation (prop= F/n). Figure 1 below shows the overall proportion of turns in each category based on map (1 or 2). *Answers* made up about 25% of all turns and were the most prevalent turn type, followed by *other* turn types which were a little over 20%. *Question* turns made up about 20% of all turns, while *comparison* and *direction* turns each made up about 15% of all turns. *Repetitions* were infrequent and made up less than 5% of all turns. This pattern was consistent across both maps and across all gender pairs. A three-way split plot ANOVA examined the influence of pair sex (MM, FM, MX), map (1 and 2) and turn code (answer, question, comparison, direction, repetition, and other) on proportion of turns. While results indicated a main effect of turn type [$F(5,465) = 49.85$, $p < 0.0001$, $\eta^2 = 0.3$], and an interaction of map and turn type [$F(5,465) = 2.34$, $p < 0.04$, $\eta^2 = 0.005$], no significant effect of pair sex was found

([$F(10,465) = 0.4$, $p = 0.9$, $\eta^2 = 0.007$]. This implies that turn types varied in proportion overall, which was influenced by map, however, turn types did not vary based on pair sex. Figure 2 shows the proportion of turns in each category based on pair sex with a side by side comparison between maps.

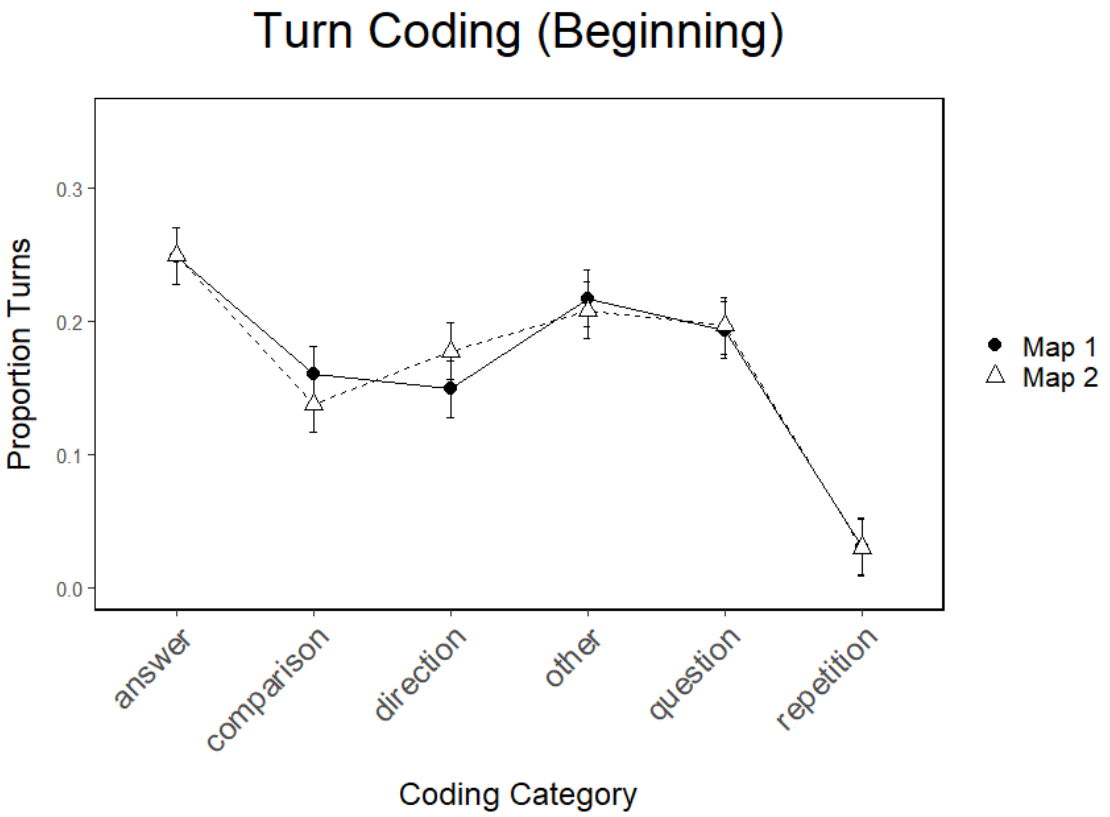

Figure 1: This figure shows the proportion of each type of turn by map. Error bars reflect 95% confidence intervals using Fisher's Least Significant Difference (FLSD).

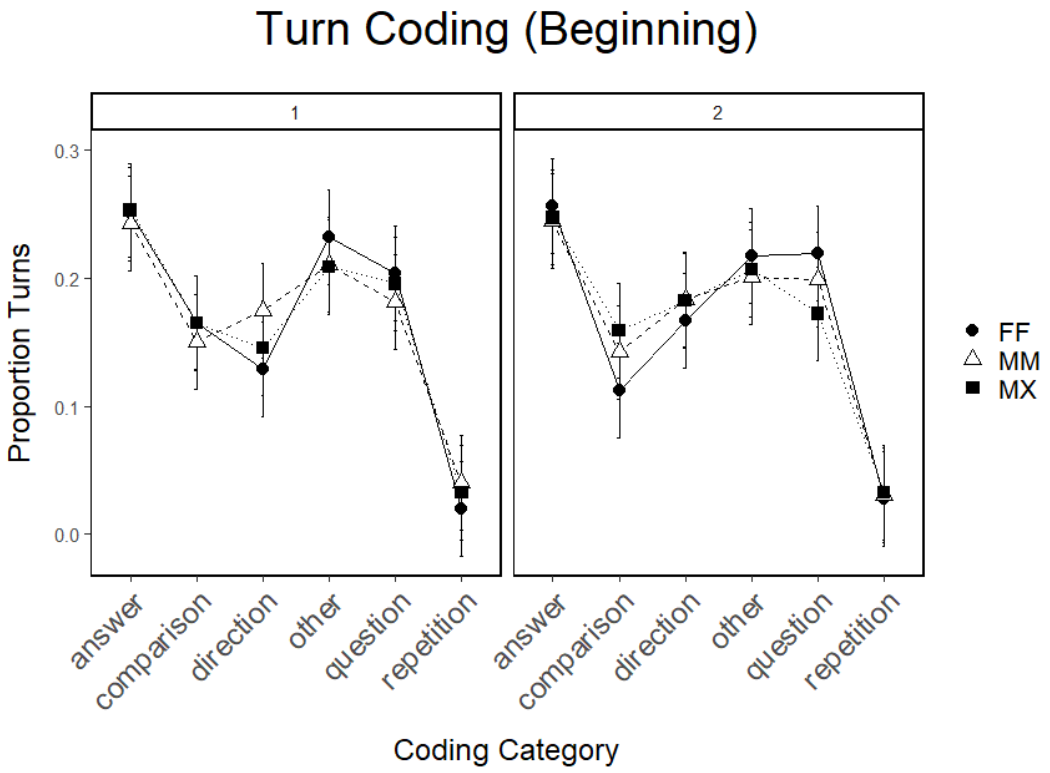

Figure 2: This figure shows the proportion of each turn type based on pair sex with a side by side comparison between maps. Error bars reflect 95% confidence intervals using FLSD.

To determine conversational style based on talker sex rather than pair sex, proportions of turn categories between talker sex and map were analyzed. Figure 3 below shows the proportion of each turn type based on talker sex overall, while Figure 4 compares the proportion of each turn type based on talker sex, with a side-by-side comparison of map. These talker sex patterns are consistent with the pair sex patterns found previously. A three-way split plot ANOVA examined the influence of talker sex (male and female), map, and turn code on proportion of turns. While results indicated a significant main effect of turn type [$F(5,470) = 50.56$, $p < 0.0001$, $\eta^2 = 0.3$] and an interaction of map and turn type[$F(5,470) = 2.35$, $p < 0.04$, $\eta^2 = 0.005$], no significant effect of talker sex and map were found [$F(5,470) = 1.2$, $p = 0.3$, $\eta^2 = 0.01$]. This

implies that turn types varied in proportion overall, which was influenced by map, however, turn types did not vary based on talker sex.

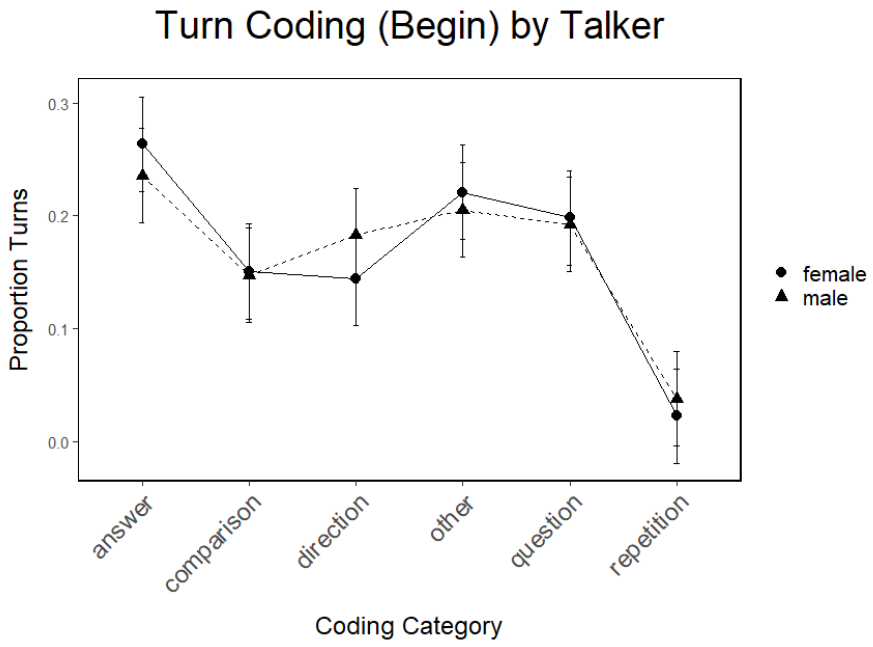

Figure 3: This figure shows the proportion of turns for each turn type based on talker sex. Data is consistent with the results of pair sex turn coding analyses. Error bars reflect 95% confidence intervals using FLSD.

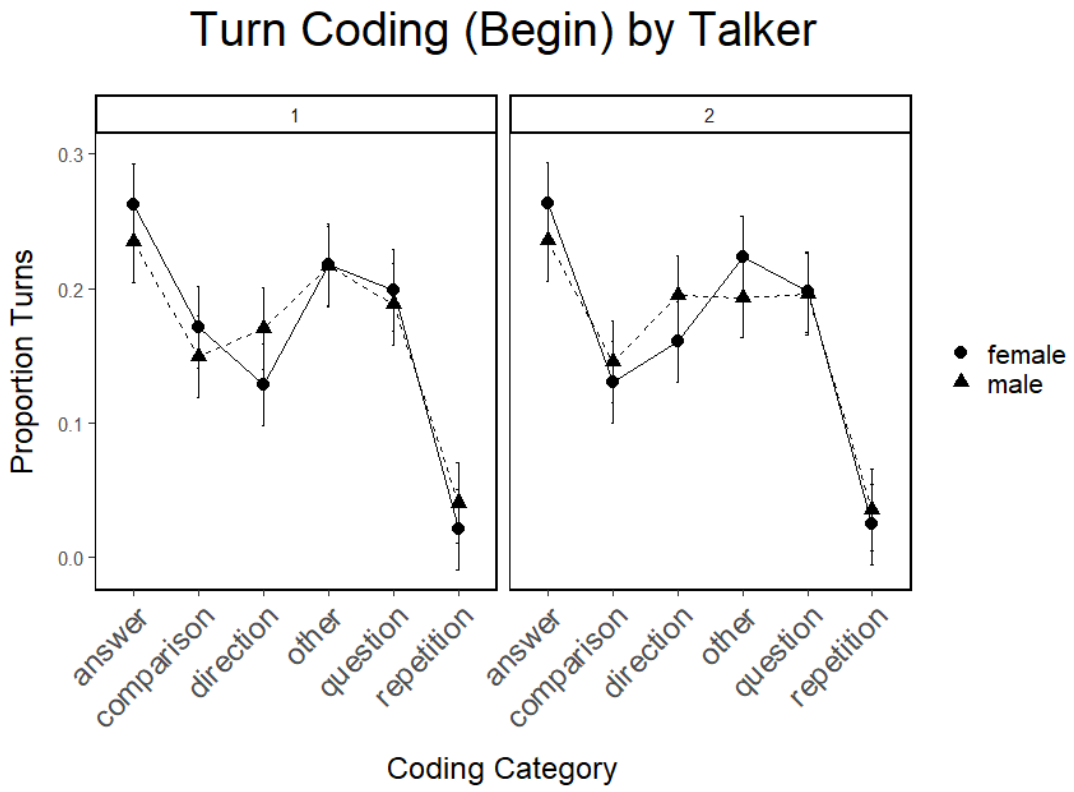

Figure 4: This figure mirrors the patterns in the talker sex and turn type comparison above; however, it is separated by map. Error bars reflect 95% confidence intervals using FLSD.

Question Style Analyses

To determine patterns of conversational styles in question-asking, proportions of each question type were compared based on pair sex, talker sex, and map. Figure 5 below shows the overall proportion of turns for each question category based on map. Clarification questions were used most frequently, followed closely by information questions. Comparison questions were asked at a moderate frequency, while challenge, direction, repetition, and other questions were asked very infrequently. This pattern is almost identical in both maps. A three-way split plot ANOVA examined the influence of pair sex, map, and question code (clarification, comparison, information, challenge, repetition, direction, and other) on proportion of turns. While results

implied a significant main effect of question type [$F(6,558) = 217.71$, $p < 0.0001$, $\eta^2 = 0.62$], no significant effects of pair sex were found [$F(12,558) = 1.42$, $p = 0.15$, $\eta^2 = 0.02$]. This implies that questions varied in proportion overall, however, turn types did not vary based on pair sex. Figure 6 shows the proportion of questions category based on pair sex with a side by side comparison between maps. This pattern is consistent with the findings from figure 5.

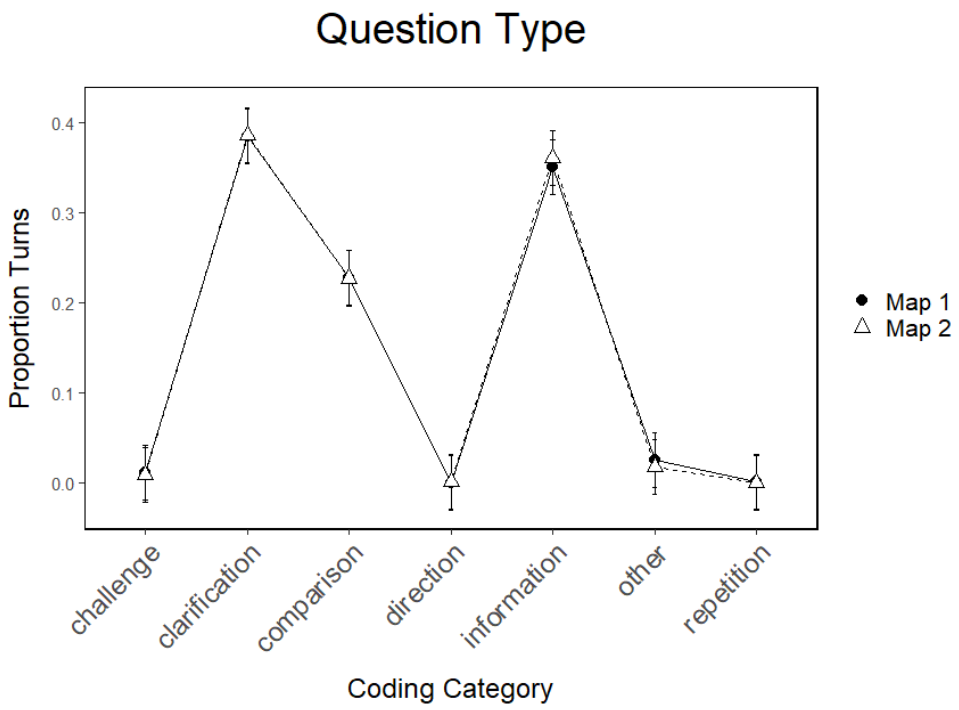

Figure 5: This shows the overall proportion of turns for each question category based on map. This pattern is consistent between both maps. Error bars reflect 95% confidence intervals using FLSD.

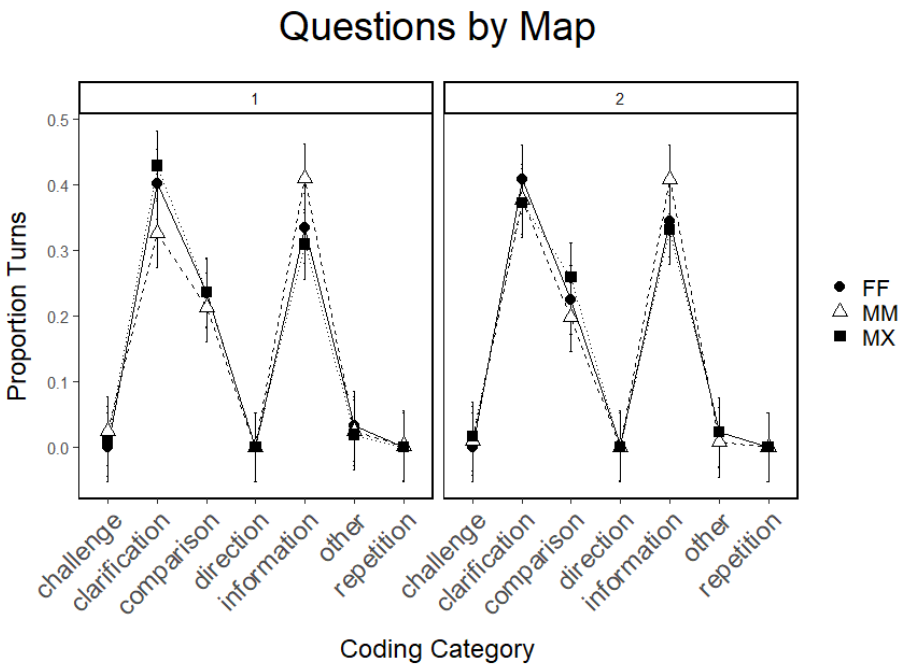

Figure 6: This shows the proportion of questions type based on pair sex with a side by side comparison between maps. Error bars reflect 95% confidence intervals using FLSD.

To understand patterns of question-asking styles based on talker sex, proportions of question types by different speakers were compared by map. Figure 7 below shows the overall proportion of turns for each question category based on talker. Patterns of question asking in talker sex are consistent with the patterns seen in figures 5 and 6, when evaluating pair sex. A three-way split plot ANOVA examined the influence of talker sex, map, and question code on proportion of turns. While a main effect of question type was found [$F(6,564) = 215$, $p < 0.0001$, $\eta^2 = 0.62$], no significant differences in talker sex were found [$F(6,564) = 0.65$, $p = 0.7$, $\eta^2 = 0.005$]. This implies that questions varied in proportion overall, however, turn types did not vary based on talker sex. Figure 8 shows the proportion of questions category based

on talker sex with a side by side comparison between maps. The patterns seen in figure 8 are consistent when separated by map.

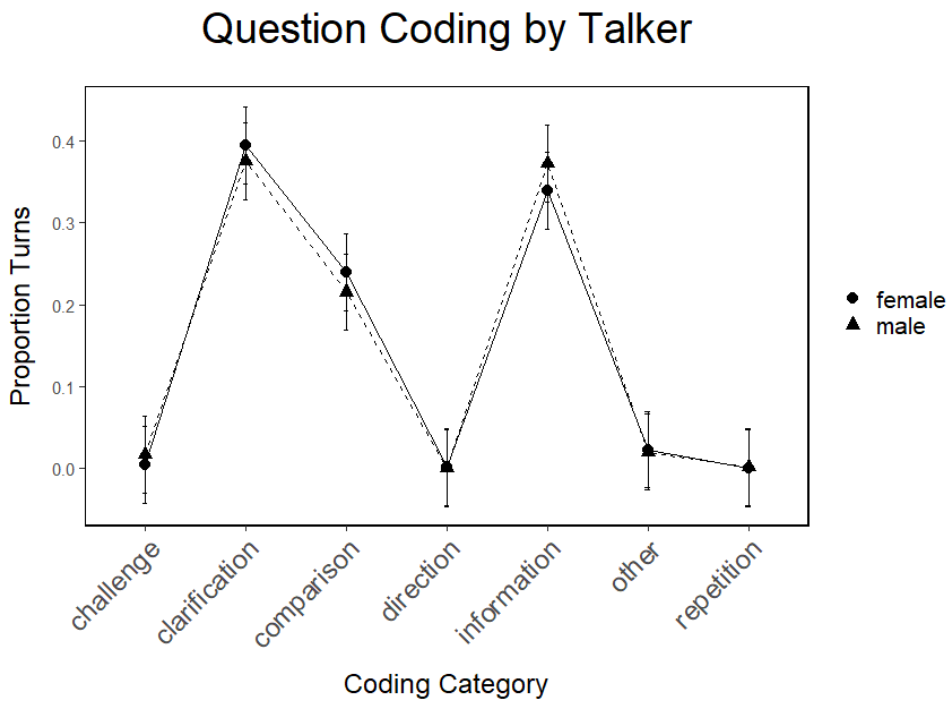

Figure 7: This figure shows the overall proportion of each question type based on talker sex. Error bars reflect 95% confidence intervals using FLSD.

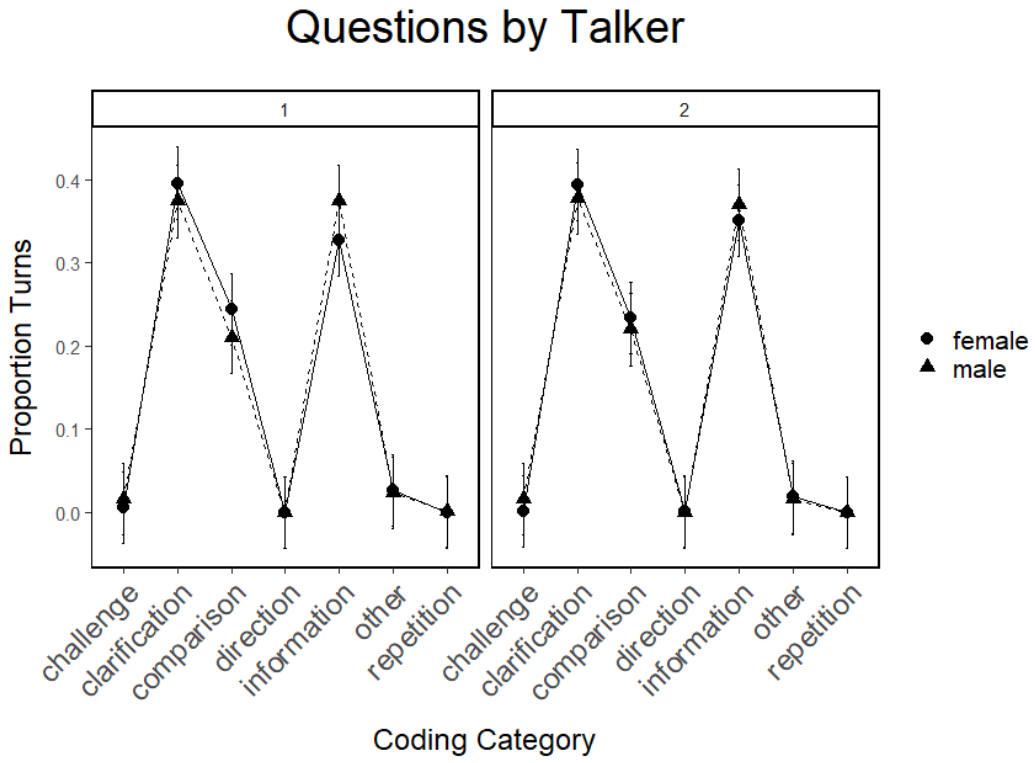

Figure 8: This graph mirrors the figure above but is separated by map. Error bars reflect 95% confidence intervals using FLSD.

Relations with Task Performance

To determine whether the proportion of question type was related to task performance, correlations for the talker's performance and their partner's performance were compiled for each coding category by talker sex and pair sex and were separated by map. Overall correlations between each coding category and map task performance were also determined. Most correlations did not reach significance or show any meaningful patterns related to pair or talker sex, with the exception of clarification questions. See Appendix C for complete correlation findings.

The only significant relationships between question type and performance were found in the clarification questions—40% of all same-sex female and mixed pairs' questions were clarification questions (SD= 0.22), while 36% of all same-sex males' questions were clarification questions (SD= 0.18). A marginally significant correlation was found between use of clarification questions overall and participants' own performance in Map 1 [$r(94) = -0.18$, $t = -1.80$, $p < 0.07$]. This indicates that use of clarification questions was associated with poorer performance on the talkers' own map task. Breaking this down by pair sex, there were significant correlations between proportion of clarification questions and participants' own task performance in same-sex male pairs for Map 1 [$r(30) = -0.43$, $t = -2.58$, $p < 0.01$]. Mixed sex pairs' clarification proportions also trended negatively with participants' own map performance, however these findings were not significant [$r(30) = -0.26$, $t = -1.46$, $p < 0.16$]. Conversely, same-sex female pairs' clarification questions were positively trended with map performance, but these were also not significant findings [$r(30) = 0.1$, $t = 0.53$, $p < 0.6$].

To assess whether these patterns of correlations found between performance and the use of clarification questions were influenced by pair sex, a linear mixed regression analysis was used to predict participant performance on the task in map 1 as a function of proportion of clarification questions and pair sex. For pair sex, same-sex female was the base level, and the model included the interaction between z-scale normalized proportion clarification questions and pair sex. The results of the analysis found that the proportion of clarification questions was not a reliable predictor of a participants' own performance in map 1 overall [$ß = 0.83$ (1.74), $t(90) = 0.48$, $p = 0.6$]. The analysis also found that same-sex female pairs' use of clarifications differed from same-sex male pairs in predicting own performance in map 1 [$ß = -6.10$, $t(90) = -2.64$, $p < 0.03$], but use of clarifications by mixed pairs did not differ from female pairs [$ß = -3.50$, $t(90) =$

-1.43, $p = 0.16$]. The implications from this are that same sex male pairs exhibited a greater negative relationship between the proportion of clarification questions used and their own task performance on map 1. The full regression model had only a marginal fit to the dataset, and accounted for only 5% of variance, likely due to lack of power in the dataset [$F(5,90) = 2.09$, $p = 0.07$, $R^2 = 0.10$, $R^2_{adjusted} = 0.05$].

Talker sex proportions of clarification questions were similar to pair sex findings—38% of all questions that males asked were clarification questions (SD = 0.19) and 40% of all questions females asked were clarification questions (SD = 0.22). A significant correlation was found between the proportion of clarification questions and partner task performance in male talkers overall for map 1 [$r(46) = -0.39$, $t = -2.88$, $p < 0.006$]. This implies that male talkers' partner performance in map 1 was significantly related to their use of clarification questions. In contrast, female talkers' partner performance was not correlated with clarification questions in map [$r(46) = 0.05$, $t = 0.37$, $p < 0.71$]. To assess whether these patterns of correlations found between performance and the use of clarification questions were influenced by talker sex, a linear mixed regression analysis was used to predict participant performance on the task in map 1 as a function of proportion of clarification questions and talker sex. The analysis confirmed that on basis of talker sex regardless of pairing, proportion of clarifications significantly predicted male talkers' partner performance in map 1 [$ß = -4.74$, $t(92) = -2.02$, $p < 0.03$]. The full regression model had only a marginal fit to the dataset, and accounted for only 7% of variance, likely due to lack of power in the dataset [$F(3,92) = 3.55$, $p = 0.02$, $R^2 = 0.10$, $R^2_{adjusted} = 0.07$].

Challenge questions occurred very rarely. Overall, only 22 out of a total of 2,165 questions asked were challenge questions. Fourteen of which were spoken by males and eight of which were spoken by females. Challenge questions were only utilized by male pairs and mixed

sex pairs in both maps. Direction, repetition, and other questions also had too few observations to conclude any meaningful patterns or stylistic consistencies. Even though information questions had a substantial amount of observations, they were not reliably related to performance and presented no meaningful patterns across pair or talker sex.

Overall, proportions of turns and questions did not vary based on pair sex or talker sex. These measures were largely consistent across map as well. Relationships between performance question type suggested no meaningful patterns based on pair or talker sex, however some trends were found in clarification questions, where same-sex male pairs and male talkers overall showed negative relationships with performance, while females showed positive relationships with performance.

Discussion

The goal of this study was to understand more about the nuances of gender dynamics in communication. Using unexplored gender disparities in a previously collected speech corpus, the Montclair Map Task Corpus, this study aimed to examine stylistic patterns in utterance goals and question-asking. The MMTC utilized a collaborative map task to understand effective communication, gender differences in communication, phonetic convergence in spontaneous speech, and patterns in conversational style. The current study built upon those goals by engaging in a more in-depth analysis of the MMTC dialogues. To accomplish this, the previously transcribed conversations were coded to determine turn goals (question, answer, repetition, direction, comparison, or other). Turns that were observed as questions were further coded to determine question objectives (clarification, challenge, information, comparison, direction, repetition, or other). Based on the Leaper and Ayres (2007) meta-analysis of affiliative and

assertive speech, it was hypothesized that males would utilize challenge questions more than females and that females would utilize more information seeking and clarification questions.

Questions accounted for approximately 20% of all the beginning turn categories. Answers accounted for about 25% of all turn types. There were no significant differences in the frequency of each turn type based on pair sex, talker sex, or even map. This means that all speakers consistently use the same broad types of speech turns when working collaboratively on a task. Turn coding analyses indicated that turn types remained consistent across map, pair sex and talker sex. Answers were the most common turn type across map and pair sex. Question and other turns were the next most frequent turn types. Comparisons and directions had an intermediate frequency. Repetitions were consistently infrequent across map and pair sex, which implies that the dynamic of communication between all the participants was very similar. This uniformity makes sense because the task was cooperative and not gender-related so they would not be influenced by gender dynamics. These findings are consistent with Leaper and Ayres' meta-analysis findings (2007). Operational definitions inconsistent with affiliative or assertive speech characteristics found no effects of pair or talker sex in their analyses. From a logistical standpoint it also would not make sense for proportions of turn types to vary based on sex because questions usually lead to an answer response, and directions and comparisons are necessary to complete the task at hand. Any deviation from those proportions would lead to poor performance.

Because questions and responses to questions make up about half of all speaker turns in this task, questions were an important facet of communication. Pair sex variations in question types were expected based on previously mentioned gender dynamics theories and studies. However, this corpus analysis did not find consistent effects of gender. One potential reason for

this could be societal roles have evolved since these gender dynamic hypotheses were made. Some of the studies used as a reference are over 30 years old. It is entirely possibly that over time, these power dynamics subside when society begins to reach equality or that the younger generation (this study used undergraduates) does not follow the same power dynamics as the previous generations used in studies because of changes to the family dynamic, gender roles, and other societal influences. It is also possible that the question coding scheme was not specific enough to capture the dynamics of communication properly. A question coding design similar to the dialogue act coding scheme used by Reichel, Pörner, Nowack & Cole's (2015) may have depicted a more robust gender dynamic, consistent with social gender dynamic theories. Like the current study, their design recorded the overall goals of each speaking turn but also indicated more detailed observations about their speaking turns. Examples of these are whether a statement agreed or disagreed with their speaking partner and whether the speaking turn was positive or negative. These measures were truncated to save time; however, agreeability is a characteristic of affiliative speech. Perhaps measuring this would have contributed to understanding affiliative speech patterns that were not seen in the current study.

The proportion of each question type did not significantly vary based on pair sex or talker sex or by map. Though this finding was not expected based on an analysis of previous literature it implies that speaker sex and partner sex has no effect on the overall proportion of turns for each type of utterance in the present corpus. Previous studies, including the MMTC, have found differences in communication based on talker and pair sex. Since the proportion of question turns did not significantly differ based on sex or partner sex, any differences had to then exist in the way these questions were asked. Those who had a weaker partner or personal performance, likely had the most miscommunications in their collaboration.

While challenge questions were hypothesized to have the most significant differences overall based on pair and talker sex, results were inconclusive. This is likely because challenge questions occurred less frequently than anticipated. Overall, there were only 22 challenge questions asked in maps one and two. 14 of which were asked by males and eight of which were asked by females. There were no same-sex female challenge questions asked across both maps. The only times females used challenge questions was when they had a male partner. It is quite possible that if all 6 maps were measured in this analysis, that these differences and correlations would have been significant. Results were expected to follow Leaper and Ayres' (2007) findings that men use more assertive speech than women. While the data trend towards this, there is no way to verify this until more challenge questions are observed.

Clarification questions showed some promise with respect to correlations with task performance based on pair and talker sex. Clarification questions are usually asked when a speaker's message is unclear to the listener, so it would make sense that clarification questions are correlated with performance on a collaborative task. Significant correlations were found between the proportion of clarification questions and map performance, with same-sex male pairs showing moderately negative relationships with own performance, and male talkers in general showing moderately negative relationships with partner performance. In contrast, female pairs and talkers showed no correlations with performance. This pattern is similar to that found in previously reported findings with respect to talkativeness. In both cases, more clarification questions (and more words) among same-sex male pairs were associated with lower performance levels. In the current study, regression analyses found these results to be marginally reliable, probably due to lack of power in the dataset.

Repetitions in the form of a question were recorded but only occurred two times, which were during collaborations on the second map. Inferences were not made due to the lack of observations recorded. Repetitions use could be a sign that the speaker is repeating themselves or their partner's previous statement because the listener was not clear on the speakers' message. This should be analyzed in future studies. Direction questions also occurred too infrequently to make any inferences. It is likely that most direction-seeking questions were regarding novel landmarks or were coded as clarification questions. Information proportion of turns varied in correlation with map task performance based on talker and partner sex. These correlations were weak, and regression analyses found information questions to be a poor predictor of task performance.

Comparison questions were not significantly correlated with map performance based on speaker or pair sex. Overall, the correlations changed from positive to negative between maps one and map two. Because none of these correlations were significant, comparison questions may just be a necessary type of question that has no relationship to gender dynamics or map performance.

Limitations and Future Research

As with any research, there were some limitations, which can help inform future studies. While the categories for turn type seemed to be a comprehensive coverage of all turn types, there was a somewhat large proportion of "other" categorized turn types. It is possible that a category of type of speaking turn was missed or should not have been included in the coding scheme. These "others" were concentrated at the beginning and ending sequences of speaking turns of the entire map and were often incomplete utterances but were counted as a full speaking turn. To

avoid coding speaking turns that have no goal, future research should only code speaker turns with complete utterances.

Hand gestures, nods and other non-verbal communications were not recorded. While the partners may not see each other, these could indicate understanding, confusion and disagreement giving better context to the transcriptions and possibly aid in the coding. Future studies could also code a video recording for non-verbal communication. If the full set of maps within the MMTC set were transcribed rather than just the first two maps, observations in coding categories that were too scarce to fully analyze, such as repetitions and challenges, may have contained enough data to make inferences regarding conversational styles. Future research on this corpus should utilize all six maps' conversational data.

It is possible that this coding system too simple to capture the nuances of the conversations or that the question categories need to be further broken down to understand more about what kinds of questions are being utilized. A coding category that was purposely omitted was off-topic or unrelated speech. These occurrences, while rare, were likely coded in the "other" category. It may have been beneficial to see how many utterances went off-topic and if that had any type of relationship with task performance.

As explained in Leaper and Ayres' meta-analysis (2007) investigating gender differences in speech, there are several ways in which the methods could have influenced the outcome of the study. Influences Leaper and Ayres explored that may have affected outcomes of the current study are the age of the participants (significant differences were found between studies using undergraduates versus older participants; undergraduates had a higher likelihood of showing sex differences), degree of intimacy (higher likelihood of showing sex differences in strangers rather than acquaintances), and size of group (dyads have more significant findings than groups of

three or more). Predictions regarding the outcomes of the current study, which used undergraduates who were strangers and placed in dyads were based on Leaper and Ayres' findings. This would include a greater probability of finding sex differences, however, the results of the current study were not consistent with theses outcomes, as no significant sex differences were found.

Other potential confounds Leaper and Ayres mention that could be at play in this study are within-pair dominance, individual personality traits, sexuality, self-esteem, and self-identified gender roles. Future research should focus on controlling for those potential confounds.

Overall, the current study did not find meaningful gender differences in communication, which is inconsistent with results in prior research. The current study's results indicated distinct patterns of similarity in turn types and question types across all gender pairs. Most of the question categories we either underpowered or had too few observations, so no conclusive patterns or inferences could be drawn. Further analyses of the MMTC should be conducted using the full dataset of six maps rather than the two maps used in the current study.

References

Adams, K. A. (1980). Who has the final word? Sex, race, and dominance behavior. *Journal of Personality & Social Psychology*, *38*, 1-8.

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., & Weiniert, R. (1991). The HCRC Map Task Corpus. *Language & Speech*, *34*(4), 351–366.

Anderson, K. J., & Leaper, C. (1998). Meta-analyses of gender effects on conversational interruption: Who, what, when, where, and how. *Sex Roles*, (3–4), 225-252

Anderson, L. R., & Blanchard, P. N. (1982). Sex differences in task and social-emotional behavior. *Basic & Applied Social Psychology*, *3*(2), 109-139

Andersen, P. A. (2006). The evolution of biological sex differences in communication. In D. J. Canary & K. Dindia (2nd Eds.), *Sex Differences and Similarities in Communication* (2nd ed., pp. 117-135). Mahwah, NJ: Lawrence Erlbaum.

Bilous, F.R. & Krauss, R.M. (1988). Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads. *Language & Communication*, *8*(3/4), 183-194.

Clark, H.H. & Krych, M.A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*, 62-81.

Davis, B. M., & Gilbert, L. A. (1989). Effect of dispositional and situational influences on women's dominance expression in mixed-sex dyads. *Journal of Personality & Social Psychology*, *57*, 294-300.

Fidel, R., Bruce, H., Peitersen, A., Dumais, S., Grudin, J., & Poltrock, S. (2000). Collaborative information retrieval. *Review of Information Behavior Research, 1* (1)

Graesser, A. C., Baggett, W., & Williams, K. (1996). Question-driven explanatory reasoning. *Applied Cognitive Psychology, 10*(7), 17–31.

Gravano, A. & Hirschberg, J. (2009). Turn-yielding cues in task-oriented dialogue. *Proceedings of SIGDIAL 2009: the 10th Annual Meeting of the Special Interest Group in Discourse and Dialogue*. 11-12 September 2009, London, UK.

Hall, J. A. (2006a). Nonverbal behavior, status, and gender: How do we understand their relations? *Psychology of Women Quarterly*, 30, 384-391.

Hyde, J.S. & Linn, M.C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*(1), 53-69.

Ivanova-Stenzel, R. & Kübler, D. (2011). Gender differences in team work and team competition. *Journal of Economic Psychology*,*32*, 797-808

Leaper, C. & Aryes, M. M. (2007). A meta-analytic review of gender variations in adults' language use: Talkativeness, affiliative speech, and assertive speech. *Personality & Social Psychology Review, 11*, 328-363.

Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology* 14, 78–106.

Maltz, D. N., & Borker, R. A. (1982). A cultural approach to male-female miscommunication. In J. J. Gumperz (Ed.), *Language and Social Identity* (pp. 196-216). Cambridge, UK: Cambridge University

Mulac, A., Bradac, J.J. & Gibbons, P. (2001). Empirical support for the gender-as-culture hypothesis: An intercultural analysis of male/female language differences. *Human Communication Research, 27*(1), 121-152.

Pardo, J.S., Urmanche, A., Gash, H., Wiener, J., Mason, N., Wilman, S., Francis, K & Decker, A. (2018). The Montclair map task: Balance, efficacy and efficiency in conversation interaction. *Language and Speech,* June, 1-21

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/

Reichel, U. D., Pörner, N., Nowack, D., & Cole, J. S. (2015). Analysis and classification of cooperative and competitive dialogs. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2015-January,* 3056-3060.

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn taking in conversation. *Language,* 50, 696-735. http://dx.doi.org/10.2307/412243

Smith-Lovin L., Robinson D.T. (1992) Gender and Conversational Dynamics. In: Gender, Interaction, and Inequality. Springer, New York, NY

Stivers, T., & Enfield, N. J. (2010). A coding scheme for question-response sequences in conversation: Question-response sequences in conversation across ten languages. *Journal of Pragmatics, 42*(10), 2620-2626

Thorne, B., & Henley, N. (1975). *Language and Sex: Difference and Dominance*. New York: Newbury House Publishers.

Wang, F., Kessels, H.W. & Hu. (2014). The mouse that roared: neural mechanisms of social hierarchy**.** *Trends in Neurosciences*, 37(11), DOI: 10.1016/j.tins.2014.07.005

Wood, M. M. (1966). The influence of sex and knowledge of communication effectiveness on spontaneous speech. *Word 22*(1), 112-137.

Appendix A: Sample MMT Pair of Maps

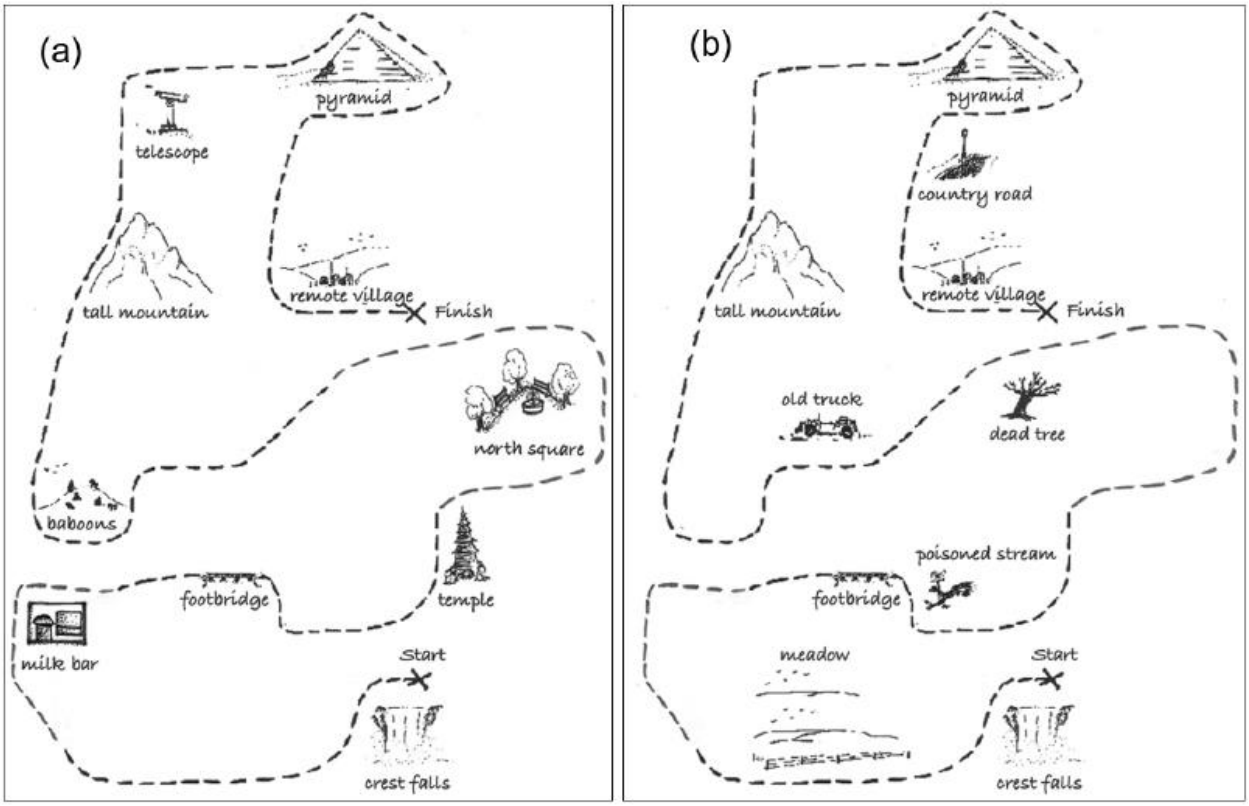

Figure 1: A sample pair of maps used in the MMT

Appendix B: Excerpt from Task Transcriptions

Talker A: from the start um there would be the camera shop

Talker B: yeah I have that one too

Talker A: and then under that is the parked van

Talker B: that one I'm missing

Talker A: there's a museum okay and then where that line go --slopes down is the yacht club

Talker B: okay I have that one too

Appendix C: Question Coding and Performance Correlations

OWN PERF	Total Cor	SSMA	SSFM	MX	MA	FEM
clarification	*-0.18	**-0.43**	*0.1*	*-0.26*	*-0.25	*-0.11*
challenge	*-0.09*	*-0.06*	NA	*-0.09*	*-0.08*	*-0.09*
direction	NA	NA	NA	NA	NA	NA
information	*0.14*	*0.32	*0.05*	*0.1*	*0.13*	*0.17*
repetition	*0.04*	*0.08*	NA	NA	*0.05578*	NA
comparison	*0.08*	*0.05*	*-0.09*	*0.22*	*0.15*	*-0.01*

PARTN PERF	Total Cor	SSMA	SSFM	MX	MA	FEM
clarification	*-0.19	*-0.33	*0.07*	**-0.34**	**-0.39**	*0.06*
challenge	*-0.04*	*0.08*	NA	*-0.19*	*0.06*	*-0.2*
direction	NA	NA	NA	NA	NA	NA
information	*0.12*	*0.15*	*0.02*	*0.26*	*0.22*	*0.02*
repetition	*0.04*	*0.09*	NA	NA	*0.07*	NA
comparison	*0.07*	*0.03*	*-0.06*	*0.21*	*0.13*	*-0.02*

This table shows correlations between Map 1 performance (own-OWN PERF or partner-PARTN PERF) and the proportion of each question type (clarification, challenge, information, direction, repetition, other) based on pair sex (same sex male-SSMA, same sex female-SSFM, and mixed sex-MX), talker sex (male-MA or female-FEM), or overall- Total Cor. Italic indicates not significant, * indicates marginally significant, **bold** indicates $p < .05$, NA indicates there were not enough observations to run a successful correlation.

OWN PERF	Total Cor	SSMA	SSFM	MX	MA	FEM
clarification	*-0.05*	-0.27	*-0.03*	*0.11*	*-0.23*	*0.08*
challenge	*-0.1*	*-0.25*	NA	*-0.01*	*-0.16*	*0.11*
direction	*0.03*	NA	*0.07*	NA	*NA*	*0.05*
information	*0.12*	**0.35**	*0.06*	*-0.04*	**0.36**	*-0.05*
repetition	NA					
comparison	*-0.08*	*-0.03*	*-0.1*	*-0.13*	*-0.1*	*-0.1*

PARTN PERF	Total Cor	SSMA	SSFM	MX	MA	FEM
clarification	*-0.04*	*-0.11*	*-0.09*	*0.09*	*0.03*	*-0.07*
challenge	*-0.01*	**-0.32*	NA	*0.19*	*-0.09*	*0.06*
direction	*-0.01*	NA	*0.01*	NA	*NA*	*0.01*
information	*0.1*	*0.12*	*0.08*	*0.06*	*0.12*	*0.08*
repetition	NA	NA	NA	NA	NA	NA
comparison	*-0.1*	*0.04*	*-0.04*	*-0.23*	*-0.17*	*-0.04*

This table shows correlations between Map 2 performance (own-OWN PERF or partner-PARTN PERF) and the proportion of each question type (clarification, challenge, information, direction, repetition, other) based on pair sex (same sex male-SSMA, same sex female-SSFM, and mixed sex-MX), talker sex (male-MA or female-FEM), or overall- Total Cor. Italic indicates not significant, * indicates marginally significant, **bold** indicates $p < .05$, NA indicates there were not enough observations to run a successful correlation.