How Beliefs about the Presence of Machine Translation Impact Multilingual Collaborations

Ge Gao[1], Bin Xu[2], Dan Cosley[2], Susan R. Fussell[1,2]
[1]Department of Communication
[2]Department of Information Science
Cornell University
Ithaca NY 14850 USA
[gg365, bx55, drc44, sfussell] @cornell.edu

ABSTRACT

Traditional communication tools tend to make their presence known, e.g., "when my collaborators and I are using IM to discuss our work, how could we *not* realize the actual presence of IM?" In the case of machine translation (MT) mediated collaborations, however, the absence or presence of MT is not obvious. English sentences with poor grammar can result from both a partner's lack of fluency and errors in the MT process. We hypothesize that partners' attributions about the source of the errors affects their collaboration experience. To test this hypothesis, we conducted a laboratory experiment in which monolingual native English speaking participants collaborated with bilingual native-Mandarin speakers on a map navigation task. Participants were randomly assigned into a 2 (beliefs about MT: absence vs. presence) by 2 (actual mediation of MT: absence vs. presence) experiment design. Beliefs about presence of MT significantly impacted the collaboration experience, opening new opportunities for both research and design around MT-mediated collaborations.

Author Keywords
Machine translation; collaboration; multilingual communication

ACM Classification Keywords
H.5.3 Group and Organization Interface: Computer-supported cooperative work

General Terms
Experimentation; Human Factors

INTRODUCTION
With the emergence of distributed organizations and virtual work, people increasingly interact with others who speak different native languages. Such multilingual collaboration is becoming increasingly common and important in

distributed work, and a variety of communication tools have been developed to help people overcome language barriers and establish collaboration [e.g.,17,26,31].

Figure 1. The use of machine translation (MT) during multilingual collaborations may not be obvious, leading to confusion about where to attribute errors in communication.

In recent years, machine translation (MT) technology has made it possible for people with different language backgrounds to interact via their own native languages [e.g.,17]. The adoption of MT creates new opportunities as well as challenges for multilingual collaboration. In particular, unlike most CMC tools whose presence is obvious in use, when collaborators communicate via MT, they may or may not realize that MT is operating in the background. Figure 1 illustrates the issue: an English speaking person who receives a message in English from a native Mandarin speaking partner may assume either that the partner has typed this message in English or that he/she has typed the message in Chinese and it was subsequently translated via MT.

This leads to a potential design opportunity for overcoming problems around multilingual collaboration that arise in part because of language disfluency, such as establishing trust and relationships between native and non-native speakers of a shared language [6,14]. If some of the attribution for disfluency in conversation can be shifted to

	Actual MT is *absent*	**Actual** MT is *present*
Believe MT is *absent*	Receive English messages that were directly typed by the partner in their second language.	Receiving English messages that were translated from Chinese via MT, **but** believe the messages were directly typed by the partner in their second language.
Believe MT is *present*	Receive English messages that were directly typed by the partner in their second language, **but** believe the messages were translated from Chinese via MT.	Receive English messages that were translated from Chinese via MT.

Table 1. The 2 × 2 experiment design for native English speakers.

the technology, people's opinions of their conversation partners might improve. On the other hand, if the quality of MT is poorer than what a bilingual speaker could achieve without MT, then the additional disfluency might harm the collaboration.

This tricky situation creates a new space for examining the role of beliefs about MT during multilingual collaboration and their effects on people's experience.

The remainder of this paper presents a laboratory study that explores how beliefs about MT impact collaboration experience in multilingual settings. In the experiment, monolingual English speaking participants performed a collaborative task with a native Mandarin speaking partner who spoke English as a second language. We manipulated both participants' beliefs about the use of MT (absence vs. presence) as well as the actual mediation of MT (absence vs. presence), creating the four collaboration conditions shown in Table 1. Participants assessed five aspects of their collaboration experience: reasons for miscommunications, message clarity, impressions of their partners, quality of the collaboration, and willingness to collaborate in the future.

Our data show significant influences of people's beliefs about the presence or absence of MT on the collaboration experience. Some of these influences were qualified by the actual presence of MT. while others were independent of its actual presence. Insights from this study provide better understanding of MT-mediated collaboration in multilingual settings and suggest ideas for the development of MT-based collaboration tools.

RELATED WORK AND HYPOTHESES
We first describe prior work on MT in CMC and collaborative work settings. We then address how beliefs about MT, which we define as a communicator's belief that the messages received from a partner have been translated via machine translation rather than typed directly, might affect attributions of error. Finally, we describe how such beliefs might shape both current and subsequent collaboration experience and set hypotheses for the study.

Machine Translation as a Communication Tool
Machine translation is a new but fast-developing communication tool. Different from most communication

tools such as IM and email, which pass messages more-or-less unchanged, MT takes a more active role in the communication, processing messages via computational translation before sending them. Adding translation capabilities to a communication channel is a promising idea for facilitating multilingual collaborations [e.g., 17].

However, several studies show that MT negatively influences multilingual collaboration. Yamashita and colleagues [28, 29, 30], for example, have shown how the use of MT, as opposed to a shared second language, created confusion about the meaning of referring expressions between multilingual speakers. Wang et al. [27] examined multilingual brainstorming conversations and found that both native and non-native English speakers viewed MT-mediated messages as less comprehensible than English messages. From these studies, it is clear that message understanding plays a fundamental role in multilingual collaboration. When messages with bad translation quality are generated by MT, it hinders the success of communication and collaboration [12].

Machine Translation and Attribution
The studies above show that issues of translation quality affect the usability of MT and the quality of collaboration experience. Theories of attribution [16, 20] paint a more complex picture of how MT might affect collaboration by changing people's beliefs about both the partner and the communication technologies used. Previous studies have shown the importance of attribution, especially the attribution of problems and failures, in shaping both current and subsequent collaboration experience [4].

Much attribution research makes a fundamental distinction between *dispositional attributions* (e.g., attributing problems to a partner's personality or lack of interest) and *situational attributions* (e.g., attributing problems to technological constraints or task difficulty) [e.g., 11, 13, 14, 22, 23]. In general, negative dispositional attributions lead to worse perceptions of partners and collaborations [1,13]. Thus, one plausible strategy for improving people's collaboration experience is to encourage attribution of problems toward the situation rather than toward the partner. For example, cues available during the collaboration, such as the absence or presence of video, can

trigger different attributions for communication and performance problems [e.g., 9].

In the current study, we hypothesize that by making the (supposed) presence of MT salient, participants will reassess where communication problems come from. When people believe the collaboration is mediated by MT, they will expect that all the messages they get are processed beforehand by MT, regardless of the actual mediation of MT. In such a case, MT has some agency in the communication and we expect people to attribute some of the responsibility for the collaboration experience to the technology rather than to their partner. This leads to the following pair of hypotheses:

H1a. People will attribute collaboration problems less often to dispositional factors (e.g., the partner's intelligence) when they believe the conversation is mediated by MT rather than conducted in a common language.

H1b. People will attribute collaboration problems more often to situational factors (e.g., quality of the translated message) when they think the conversation is mediated by MT rather than conducted in a common language.

Machine Translation and Collaboration

These attributions based on beliefs about the presence or absence of MT may in turn change people's collaboration experience. Attributing collaboration problems to a partner rather than to situational factors can lead to more negative impressions of that partner [e.g., 1, 13]. This in turn may lead to more negative opinions of the current collaboration and less motivation for future collaboration. On the other hand, attributing collaboration problems to situational factors may lead to more positive impressions of the partner as well as the collaboration. Based on these arguments, we hypothesized that beliefs about MT would influence social experience in the current collaboration and willingness to collaborate in the future:

H2a. People will judge messages from their partners to be higher in clarity when they believe MT is in use.

H2b. People will judge the collaboration to be higher in quality when they believe MT is in use.

H3a. People will be more positive about their partners when they believe MT is in use.

H3b. People will be more willing to collaborate with the same partner in the future when they believe MT is in use.

The above hypotheses pertain to people's beliefs about whether or not their collaboration is mediated by machine translation, and all posit that these beliefs will have effects independent of whether the collaboration is actually

mediated by MT or not. As we know of no prior work that has examined interactions between beliefs about MT and the actual presence of MT, we pose the following research question:

RQ. What interactions exist between beliefs about and the actual use of MT for each of our dependent measures?

METHOD

Overview

We conducted a laboratory experiment to examine the effect of beliefs about MT on attributions and collaboration experience. We used the HCRC map navigation task developed by Anderson and colleagues [2], which has been used in previous studies to examine computer-mediated communication and collaboration [e.g., 9, 10].

We paired monolingual native English speaking participants with a Chinese/English bilingual native Mandarin speaking partner using an asymmetric design similar to that used by Wang and colleagues [27]. English speaking participants always typed and received messages in English. We manipulated their beliefs about whether the English messages they received were typed by the partner directly or translated from Chinese via MT. Mandarin speaking participants typed in either English or Mandarin depending on whether MT was actually used to support the collaboration or not. Their beliefs about the presence or absence of MT could therefore not be manipulated. Participants answered a post-task survey after finishing the collaboration task on each map, and answered a post-experiment survey after finishing the whole experiment.

Participants

The study involved a total of 64 participants recruited from a U.S. university. Half (N = 32) were native English speaking undergraduate students (23 female), with a mean age of 20.6 years (*SD* = 2.12). All had lived in the U.S. for more than 10 years. They reported having some previous experience using MT (M = 3.28, SD = 1.47 on a 7-point scale ranging from 1 = never to 7 = very often).

The other half of the participants (N = 32) were bilingual native-Mandarin speakers (19 female). Their mean age was 26.8 years (SD = 5.67). All spoke English as a second language and on average reported moderate English fluency (M = 4.38, SD = 1.09 on a 7-point scale ranging from 1 = not fluent at all to 7 = very fluent) and some experience using MT (M = 4.09, SD = 2.13).

Because beliefs in the presence or absence of machine translation during the interaction could only be manipulated for the English speakers, we restrict our analysis in this paper to the 32 English speakers.
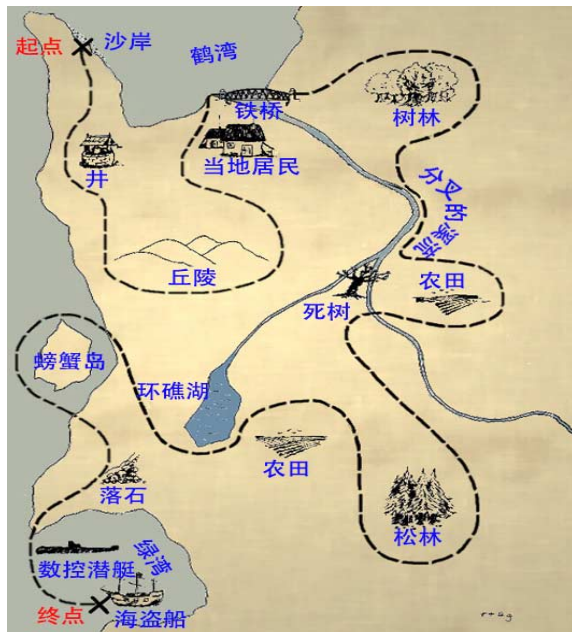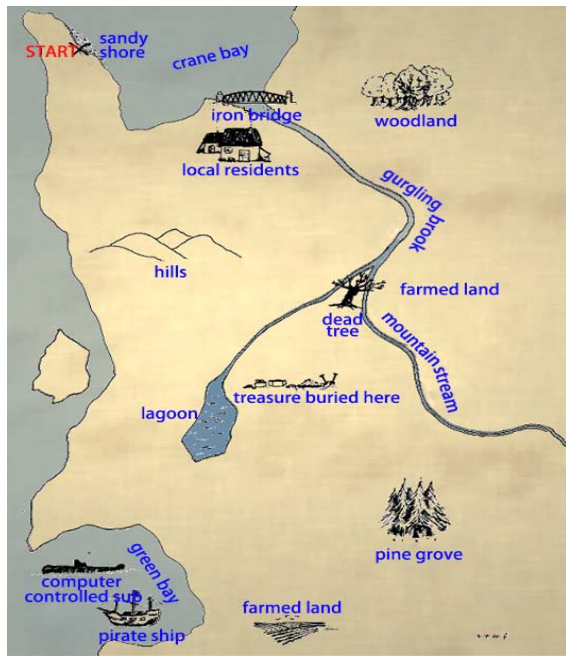
Figure 2. One set of maps used in the navigation task (the top one for the English follower, the bottom one for the Chinese instructor).

Materials

We used two HCRC map navigation tasks of equivalent difficulty [2]. One participant (the instructor) is given a map that shows a prescribed route around a series of landmarks, while a second participant (the follower) is given a map with the landmarks but not the route. The instructor and follower work together to help the follower learn the prescribed route. Figure 2 shows the two versions of one of the HCRC maps used in the study. To further complicate to the task, each pair of HCRC maps has some differences in

the depicted landmarks, e.g., a bridge present in one map might be missing in the other. We manually translated the labels on each map into Mandarin Chinese for use in the MT-mediated communication condition.

Software and Equipment

We developed an online chat tool that displayed the map during the task and provided machine translation when appropriate for a participant's condition. Figure 3 shows the two parts of the tool interface: the IM window and the map window.

IM window. On the left side of the interface was an instant messaging window where participants typed messages and saw their partner's messages. For the MT-mediation condition, participants typed in their native language and received messages translated from their partners' language to their own native language using Google Translate. Translations were automatic and no preview or pre-editing of the translated message were afforded by the IM tool. For the MT-absence condition, this module was not available, and all communication was typed in English.

Map window. On the right side of the interface was the given map. Instructors saw the map with both landmarks and the route; followers started with a map that had only landmarks. Followers could use the cursor to draw and/or correct the route they created under the instructor's guidance. For both the followers and the instructors, only the landmarks and route (if any) shown on their own maps could be seen.

Equipment. All participants were seated at Mac Pro laptops with 13.3 inch monitors in a separate room from their partners. They wore headphones during the study so as not to be distracted by outside noise.

Procedure

Participants were assigned to pairs consisting of one native English speaker and one native Chinese speaker and seated in two separate rooms in the lab. All pairs completed the two map navigation tasks using text-based IM (no video or audio was permitted). Pairs were randomly assigned to one of two actual communication medium conditions: MT-mediation (in which actual MT was present) or English as a common language (in which actual MT was absent).

As noted earlier, the manipulation of beliefs about the presence or absence of MT mediation was done only for the native English speakers in each dyad. Half of the English speakers were told that they would be working with a native Mandarin speaking partner using MT (that is, participants believed MT was present). The other half of the English speakers were told that they would be working with a native Mandarin speaking partner using English as a common language (that is, participants believed MT was absent).

Participants then completed two HCRC tasks. During the first task, participants were randomly assigned as either an

instructor or a follower and one of the two sets of maps. Participants had 15 minutes to complete the task, after

which they answered a post-task survey to assess the perceived clarity of their partner's messages, the quality of the collaboration, and attributions they made for miscommunications during the task. Then, participants switched roles with their partner and completed the second 15-minute task on the other set of maps and answered an identical post-task survey. The order of maps and roles were fully counterbalanced using a Latin Square design.

After completing both tasks, each participant rated his/her impression of the partner and willingness to collaborate with the same partner in the future based on the overall collaboration experience. We then debriefed participants.

Measures
Participants answered a post-task survey after each map task and a post-experiment survey given at the end of the experiment. All measures reflected the communication and collaboration experience of each participant at the individual level.

Post-Task Survey
Clarity of partner's messages. Participants assessed the perceived clarity of their partner's messages using a single item, "I felt my partner always expressed his/her idea clearly". Responses were on a 7-point scale (1 = strongly disagree, 7 = strongly agree).

Quality of collaborations. Participants' evaluation of the quality of their collaboration with their partner was measured using four 7-point Likert scales (e.g., "Generally, I'm satisfied with our collaboration on this task."). The questions formed a reliable scale (Cronbach's α = .92) and were averaged to create a measure of quality of collaboration.

Attribution for miscommunications. Participants' attributions for miscommunication during the collaboration were measured using 18 7-point Likert scales. Factor analysis indicated the presence of five attribution factors (Table 2). Items under each factor were averaged to create a measure for each of the attribution factors, including message quality (Cronbach's α = .91), the partner (Cronbach's α = .62), shared information within pairs (Cronbach's α = .67), oneself (Cronbach's α = .68), and the task (Cronbach's α = .61).

The Post-Experiment Survey
Manipulation checks. Beliefs about the presence or absence of MT were assessed by four single-choice questions. These questions asked which language (English vs. Mandarin) was used to write and receive messages on their partners' side and their own side.

Impression of the partner. Participants' impressions of their partners were measured using ten 7-point Likert scales (e.g., "My partner is extraverted and enthusiastic."). Scores
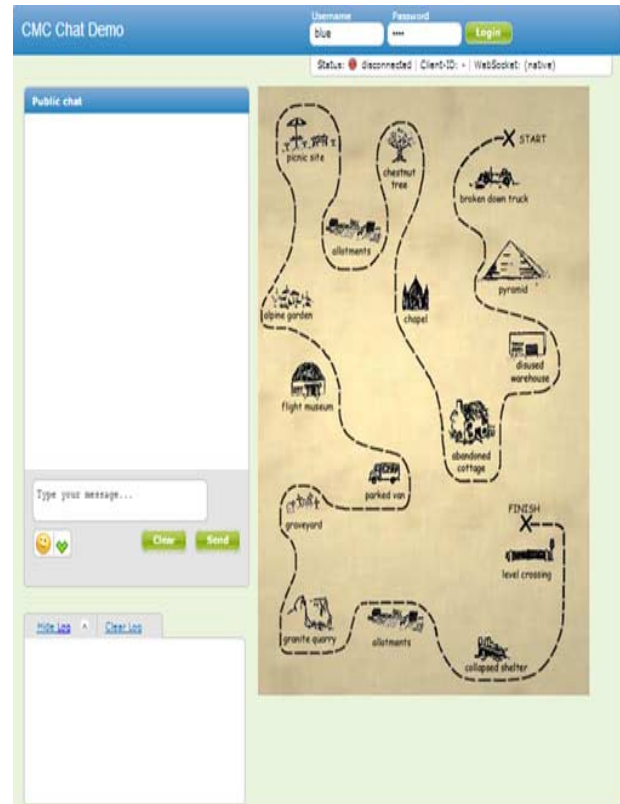

Figure 3. The interface of the chat tool as seen by the instructor.

	I	II	III	IV	V
Problematic sentence structure	.90				
Inaccurate expression	.90				
Inconsistent wording	.90				
Using of MT/English	.81				
Partner's understanding of maps		.75			
Partner's effort		.78			
Partner's intelligence		.58			
Audio cues absence			.79		
Familiarity between partners			.70		
Visual cues absence			.66		
Common background knowledge			.50		
Own intelligence				.78	
Own understanding on maps				.66	
Own effort				.54	
Route complicity					.72
Landmarks					.65
Time limitation					.65
Differences between the maps					.58

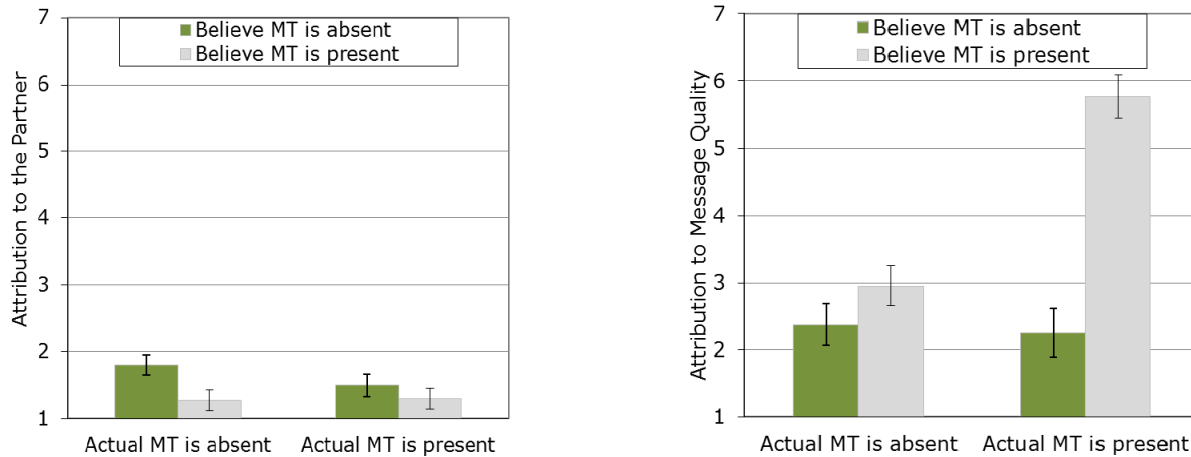**Table 2. Attribution for miscommunications, with factor loadings.**

**Figure 4. Mean attribution to the partner (left) and message quality (right) ratings by condition.
Error bars represent standard errors of the mean.**

were averaged to create a measure reflecting the positivity of their impressions of their partner (Cronbach's α = .78).

Willingness to collaborate in the future. Participants' willingness to collaborate again with the same partner was measured using three 7-point Likert scales (e.g., "I would like to collaborate with my partner in other tasks."). The questions formed a reliable scale (Cronbach's α = .87) and were averaged to create a measure of willingness to collaborate again in the future.

RESULTS

To test our hypotheses, we conducted 2 (beliefs about the MT: absence vs. presence) × 2 (actual mediation of MT: absence vs. presence) Mixed Model ANOVAs that took into account the fact that each participant provided two sets of measures, one after each map. As noted earlier, we look only at the data provided by the native English speaking participants, for whom beliefs about the presence or absence of MT could be manipulated. The demographic backgrounds (e.g., age and gender) and previous MT experience of each participant were set as control variables in all the ANOVA models.

Manipulation Checks

The manipulation checks indicated that all manipulations were successful. Regardless of the actual absence or presence of MT, 100% of English speaking participants who were told MT was absent thought their Chinese partners wrote and received messages in English and 100% of those who were told MT was present thought their Chinese partners wrote and received messages in Chinese.

The manipulation check of the actual absence/presence of MT was done by checking the chat tool settings and language using in each conversation. We found that 100% of participants and their partners in the actual presence of MT conditions generated messages in their native languages and used MT to mediate their communication, and 100% of participants and their partners in the actual

absence of MT conditions used English as a common language to communicate during the experiment.

**Attribution for Miscommunications**

To test H1a and H1b, we examined participants' attributions for miscommunications by running Mixed Model ANOVAs of the form outlined above.

For each participant, there was a significant effect of the attribution type regardless of which experiment condition the participant belonged to ($F_{[4, 25]} = 71.81$, $p < .0001$). Generally, participants attributed miscommunications more to situational factors such as message quality (M = 3.35, SE = 0.23) and the task (M = 4.09, SE = 0.15), and less to dispositional factors such as the partner (M = 1.47, SE = 0.08) and oneself (M = 1.77, SE = 0.11).

There was also an interaction between the actual presence of MT and the beliefs about the presence of MT on attribution ($F_{[4, 25]} = 5.11$, $p < .001$). To get a clearer understanding of this result, we examined each of the five attribution factors separately. No difference was found for the attributions to oneself ($F_{[1, 28]} = 0.05$, $p = .83$), the task ($F_{[1, 27.45]} = 2.60$, $p = .12$), or shared information within pairs ($F_{[1, 28]} = 0.51$, $p = .48$). However, attributions to the partner and message quality both varied with the experiment conditions.

Consistent with H1a, attribution of miscommunication to the partner (left side of Figure 4) was significantly lower when participants believed that MT was present than when they believed MT was absent (M = 1.28, SE = 0.11 vs. M = 1.64, SE = 0.11; $F_{[1, 28.15]} = 4.69$, $p = .04$). There was no main effect of the actual presence of MT ($F_{[1, 28.15]} = 1.52$, $p = .23$), nor an interaction between the actual and believed presence of MT ($F_{[1, 28.15]} = 0.53$, $p = .47$).

H1b, about belief in the presence of MT leading to increased attribution of miscommunication to situational factors, was partially supported. Attributions to message quality (right side of Figure 4) were significantly higher for
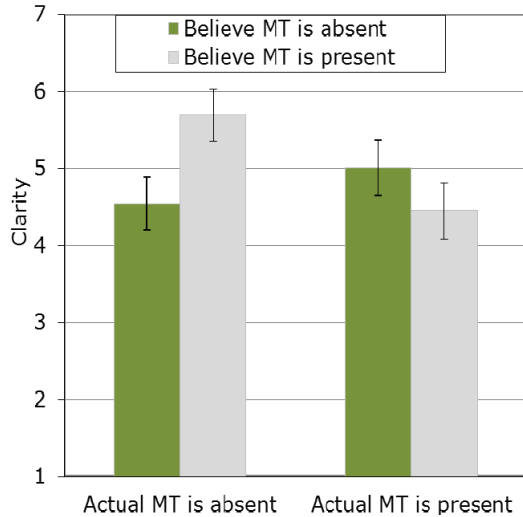
Figure 5. Mean clarity ratings by condition.
Error bars represent standard errors of the mean.
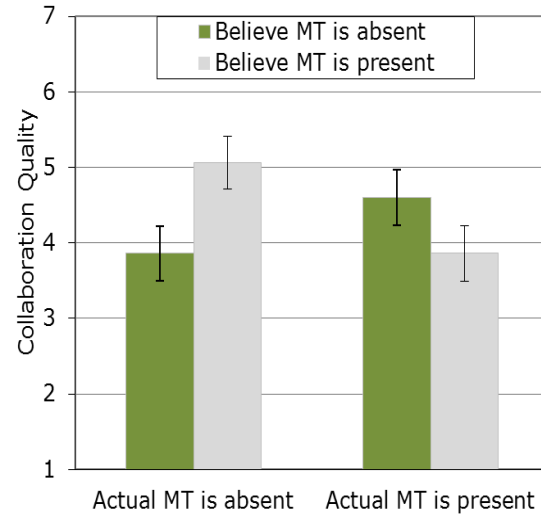
Figure 6. Mean collaboration quality ratings by condition.
Error bars represent standard errors of the mean.

both the actual mediation of MT (F [1, 28] = 17.35, p < .001) and the belief that MT was present (F [1, 28] = 48.45, p < .001) when looking at main effects. There was also a significant interaction between the actual presence of MT and the belief that MT was present (F [1, 28] = 23.53, p < < .001). In particular, the combination of MT being present and participants believing it was present led to much higher attributions of miscommunication to message quality (M = 5.77, SE = 0.32) than in other conditions.

Thus, overall, we found that beliefs about the use of MT did affect attributions, as predicted by H1a and H1b. Participants who believed that MT was present were less likely to attribute miscommunications to dispositional factors such as their partner and more likely to attribute them to situational factors such as quality of the translated message. Attributions of miscommunication to dispositional factors were especially pronounced when the actual use of MT corresponded to people's beliefs about its use.

Collaboration Experience during the Task
H2a and H2b predicted that the perceived clarity of partner's messages and collaboration quality during the task would vary depending on whether people believed MT was being used. This prediction was partially supported by participants' responses on the post-task surveys.

For H2a, about perceived clarity (Figure 5), there is a significant interaction between actual and believed use of MT (F [1, 28] = 8.91, p = .01), though neither the main effect of actual mediation nor the belief that MT was used was significant (F [1, 28] = 2.06, p = .16; F [1, 28] = 0.99, p = .33). Here, the *presence* of MT had no strong interaction with beliefs (M = 4.45, SE = 0.36 vs. M = 5.01, SE = 0.34). Instead, the *absence* of MT mattered: participants perceived

higher clarity when they believed MT was in use (M = 5.69, SE = 0.36 vs. M = 4.54, SE = 0.34).[1]

For H2b, about perceived collaboration quality (Figure 6), we saw a similar pattern to H2a, with a significant interaction effect (F [1, 28] = 7.25, p = .01) but no main effect of either the actual or believed presence of MT (F [1, 28] = 0.41, p = .53; F [1, 28] = 0.36, p = .55). As with perceived clarity, belief about MT use made no difference when it was actually present (M = 3.86, SE = 0.37 vs. M = 4.60, SE = 0.35). When actual MT was absent, however, participants perceived higher collaboration quality when they believed MT was in use versus when they did not (M = 5.06, SE = 0.37 vs. M = 3.86, SE = 0.36).

Thus overall, beliefs about the use of MT also affected people's collaboration experience during the task. Here, the dominant result is that people who believe that their partner is using MT rate both the clarity of communication and the quality of the collaboration higher—but only when there partner is actually using English as a second language, rather than MT.

Collaboration Preference beyond the Task
H3a and H3b predicted that people's overall impression of partners and willingness to collaborate with them in the future would also vary with beliefs about whether MT was

---

[1] The clarity question, "I felt my partner always expressed his/her idea clearly", did not specify whether the message to be rated was the original Chinese message or the MT translation. Thus, those who believed MT was absent based their ratings on the English messages, but those who believed MT was present may have rated either the original or the "translated" message. However, since only the interaction between actual MT and beliefs is significant, this indeterminacy does not affect the pattern of results.
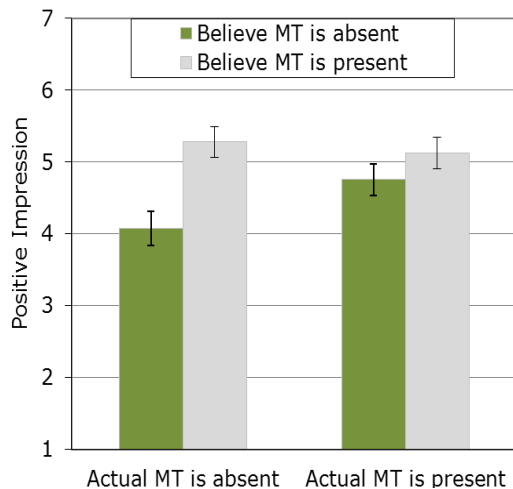
Figure 7. Mean impression of the partner ratings by condition. Error bars represent standard errors of the mean).
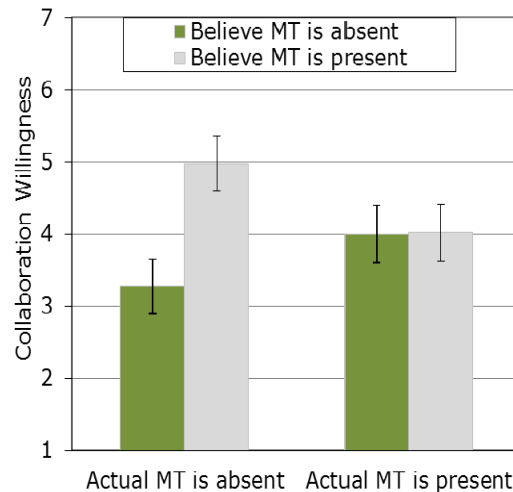

Figure 8. Mean willingness to collaborate in the future by condition. Error bars represent standard errors of the mean.

in use. This prediction was fully supported by participants' responses on the post-experiment surveys.

Consistent with H3a, people's impressions of their partner (Figure 7) were significantly higher when they believed that MT was in use, regardless of the actual presence of MT (M = 5.21, SE = 0.15 vs. M = 4.42, SE = 0.16; F [1, 23] = 13.56, p = .001). Neither the main effect of actual mediation of MT nor the interaction were significant (F [1, 23] = 1.20, p = .29; F [1, 23] = 3.54, p = .07).

We also examined H3b, participants' willingness to The main effect of believing MT was in use led participants to give higher future collaboration ratings (F [1, 25] = 5.16, p = .03) and there was a significant interaction between believed and actual use of MT. As with the ratings of collaboration experience on individual tasks (H2a and H2b), the main effect was driven by what happened when MT was absent. Here, participants who believed MT was in use had higher ratings than those who did not (M = 4.97, SE = 0.40 vs. M = 3.27, SE = 0.38; F [1, 11] = 13.79, p = .003). There were no differences based on belief when MT was actually in use (M = 4.02, SE = 0.40 vs. M = 4.00, SE = 0.38; F [1, 11] = 0.05, p = .83), and no main effect of the actual mediation of MT (F [1, 25] = 0.07, p = .80).

Once again, we see that believing MT is in use has positive effects on both participants' impression of their partner and their willingness to collaborate with them in the future. The effect is stronger when MT is not actually being used.

DISCUSSION

Our results shed new light on the possible effects of MT on collaboration experiences. Previous studies have looked at the impact of MT on collaborations from the perspective of message understanding. Instead, we explore how knowing, or believing, that MT systems are in use might affect how people make attributions about the causes of miscommunication, in turn affecting their perceptions of the partner and of the collaboration experience.

The broadest and most consistent finding is that message receivers who believe their partner's communication is mediated by MT have better communication experiences than those who don't. Belief that MT is in use leads people to make fewer dispositional attributions for miscommunication to their partner, more positive overall ratings of their partner, and a greater willingness to engage in future collaborations.

These beliefs interact with the actual use of MT in interesting ways. In particular, for attributions to situational rather than dispositional factors, the effect of believing that MT is in use is overall enhanced when MT is actually being used. For ratings of communication and partner quality, however, the opposite obtains: on balance this belief primarily affects ratings when MT is *not* being used.

We believe these interaction effects are driven by how the actual quality of messages received matches people's expectations. We suspect that people who believed that MT was in use expected to see lower quality messages on average. When MT was actually in use and people saw the obvious errors that MT sometimes generates, their beliefs about the quality of the technology were confirmed and they blamed situational factors rather than their partner. But the kinds of errors that second language speakers make when generating non-native text are perhaps both less glaring than MT errors and more familiar to native English speakers who have interacted with second language speakers. Thus, incoming non-MT messages were easier than expected to cope with for people who believed MT was being used—leading to better perceptions of their communication and their partner.

Multilingual CMC tools as socio-technical systems

One way to directly address this suspicion would be to marry our analysis of attributions with measures of message quality. Measuring the quality of translated text is non-trivial and typically requires a human gold standard (although using Mechanical Turk to judge translation quality has some promise [5]). The closest measure we have in our study is people's ratings of their partner's message clarity, which was not significantly different whether MT was present or not. On the other hand, Wang et al. [27] found that MT significantly affected perceptions of message comprehensibility.

We think this difference arises because of the nature of the task. HCRC tasks such as those used in the current study are designed to require grounding and repair work by deliberately including mismatches between landmarks on the maps. The confusion induced by these mismatches, along with the relatively constrained domain of the conversation, may have washed out differences in language clarity. In the brainstorming tasks used by Wang et al., the domain is less constrained and the need to build on others' ideas is more central to the task, leading to lower ratings of clarity [27]. Also in that study, participants knew nothing about their partner's background or the use of machine translation; running a version of that study that made clear to participants that MT would be interesting for deepening our understanding of the interplay between beliefs about and actual use of translation.

These results highlight the socio-technical nature of using MT to support collaboration. Most work so far on MT has compared it to second language use and implicated differences in comprehensibility as a problem, but quality is only part of the story. Social and psychological processes, task characteristics, and translation quality all interact to affect how using MT will affect people's experiences and outcomes.

Designing for and with MT

Interface design also can affect those experiences. We suggest several strategies that leverage these results to help people collaborate more effectively via MT by making the fact of MT's use more salient, by helping people improve the quality of messages they generate, and by more directly encouraging attribution of problems to the system rather than to collaborators. We discuss each in turn.

Making MT salient. In our experiment we used explicit instructions to prime people's beliefs that MT was present. However, interface design elements could do this work as well and have been shown in a number of cases to improve both collaboration outcomes and processes. For instance, recent CSCW work demonstrates that priming beliefs about community goals by showing desirable example content in the interface can increase the quality [24], structuring [21], and descriptiveness [18] of content people contribute to a group, Likewise, even small interventions to make group activity [7] and communication [8] more salient can

increase attachment to groups, while subtle cues about desirable collaborative behaviors can increase their frequency [3].

A simple way to increase the salience of translation would be to link to the original source text for translated text, or put an icon or the original text itself alongside the translation. The original text might not be useful in and of itself for a monolingual speaker, but it would raise awareness of the translation process in play. The interface might also provide explicit translation controls for senders, to make them aware that they (and, in turn, their partners) might be using translation. Or, if automatic translation is desired, the interface could allow receivers to specify their preferred language and tell other people in the conversation that all text will be automatically translated to partners' preferred languages. This might be especially valuable in groups with more than two native languages.

Increasing message quality. For many of our measures, the strongest effects came when people believed MT was in use, but it was not. And, although our measure of idea clarity did not reveal significant differences, prior work [27, 29] suggests that a main factor is that text generated by second language speakers is on average more readable than machine translations. Thus, systems should strive to help people generate the highest quality messages possible.

Rather than MT being an all or nothing proposition, systems might support seamless transitions between MT and second language use. The idea of auto-translation would also be useful here. People could choose to write in any language and the system could translate only as needed. The interface should also make users aware that they can choose to communicate in either language and that it's probably best to use whichever language is likely to lead to the best translation for their partners.

Another idea would be for the system to help senders generate higher-quality translations. Seeing translated messages, as suggested earlier for making translation more salient, would have useful side effects. People would both know when a particular translation was confusing and learn which kinds of language constructs an MT system tended to struggle with. The system could also allow notify senders when a message is likely to be mistranslated, and allow them to either edit the translation directly, simplify the message so that it's easier for MT to work with, or switch to a different language. The cognitive costs and the delays imposed by such a system might make it impractical, but it is an idea worth exploring.

Increasing perceived agency. Finally, we suspect that part of the effect around increased dispositional attributions comes from people seeing the system as an active agent in the communication channel. Thus, emphasizing not just the salience of MT but also its agency might increase the power

of the effect[2]. Borrowing ideas from Nass and colleagues [e.g., 19] around anthropomorphic agents, multilingual collaboration tools that include MT might literally include a "Translation Agent". The agent might present with human-like characteristics such as apologizing for errors or looking confused when translation engines report uncertainty in translation quality or parses of translated text show large deviations from normal grammar. More generally, they could exhibit behaviors that encourage users to attribute errors in communication to the tool rather to than the partner. Finally, following Adar, Tan, and Teevan's defense of benevolent deception in HCI [1], they might even lie a little about how much effect the translation has on quality.

Potential costs. Such deception, and the general goal of encouraging people to attribute problems to the system rather than their partners, does sometimes pose tradeoffs. In situations where multilingual teams are assigned and MT must be part of the collaboration, we see only benefit in reducing friction by putting the blame on the system rather than on teammates.

The picture is more complicated when people have more freedom in choosing partners. One argument in favor of using the aforementioned strategies is that non-native language speakers often face an uphill battle in establishing trust [14] and relationships [6] with native speakers in organizations with a common language. These strategies are one way to help second language speakers move past surface aspects of language disfluency to be judged on deeper merits. However, because MT and second language use do in fact impose communication costs, this may impact their partners or the long-term success of collaboration. Further, because the effects are stronger when MT is not actually in use, this may come at the expense of less fluent or mono-lingual collaborators who must make greater use of MT. These issues will need to be explored as MT tools, and designs that leverage them, become more commonly used for cross-language collaboration.

Limitations
There were several limitations to this study. First, the current study examined how MT influences collaboration only at the individual level. The measurements used here all addressed individuals' experience during the collaboration, without attention to group-level evaluations. Both are important, and future work should examine the relationship between the beliefs about and mediation of MT and collaboration from both individual and group levels. Furthermore, we examined beliefs about MT in a group of people who were relatively unfamiliar with it, and we asked them to perform only two tasks. It would be interesting to

see if beliefs had as strong of an effect among a group of heavy MT users, or over a longer time period.

Also, beliefs about presence of MT were only examined from the perspective of monolingual English speakers receiving messages. This was largely a practical consideration that arose because we recruited participants from a U.S. university. Finding monolingual non-English speaking participants is nearly impossible in this setting, and finding enough bilingual native English speakers of Chinese (or pairs of speakers who are bilingual in each other's native language in any language except perhaps Spanish) would also have been a monumental task.

This does leave open the question of how strongly the effect would appear for bilingual receivers, who might better understand the foibles of both manual and machine translation between the languages. The effect may also be less pronounced for language pairs where MT on average works better (such as Chinese-Japanese or translations between many European languages). Finally, we don't know how the effect of beliefs or choices about MT might influence message senders' perception of receivers.

CONCLUSION
Our study both answers and raises questions about the ways that people's beliefs about the use of MT affect multilingual collaboration, and we hope that both the answered and raised questions contribute to our understanding of MT.

At least for message receivers, if they believe that MT is in use, they attribute miscommunication problems to technological constraints (e.g., quality of the translated message) much more than to their partner (e.g., the partner's intelligence). They also give higher ratings of the communication experience and of the partner, particularly when MT is not actually in use, and we suspect this because messages likely seem more natural and/or of higher quality when MT is not used, and thus the level of fluency exceeds their expectations.

Our findings have a number of potential design applications around making the presence and agency of MT in multilingual collaboration systems more salient in order to encourage people to attribute errors to the system rather than to their partners. They also suggest that supporting seamless use of both MT and second language proficiency is likely to increase collaboration quality, especially if bilingual speakers are coached to use tools in ways that maximize translation quality.

From a research perspective, our open questions about monolingual versus bilingual speakers, senders versus receivers, and easy versus hard language pairs provide a set of directions to explore. Finally, our results suggest that the quality of messages and the causes of imperfect messages, along with people's understanding of those causes, all work together to shape individuals' collaboration experience together; none alone can tell the whole story. Research that addresses each chapter of the story will lead to better

---

[2] Interestingly, there is some debate in both the literary community and the professional translation community about how salient the human translator of texts should be as an active agent in the communication [25].

understanding of and systems for using MT to support multilingual collaboration.

REFERENCES

1. Adar, E., Tan, D.S., & Teevan, J. Benevolent Deception in Human Computer Interaction. In *Proc. of CHI 2013*, 1863–1872.

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

3. Balaam, M., Fitzpatrick, G., Good, J., & Harris, E. (2011, May). Enhancing interactional synchrony with an ambient display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 867–876). ACM.

4. Bazarova, N.N. & Hancock, J.T. (2011). Attributions after a group failure: Do they matter? *Communication Research*, *40*, 1–24.

5. Callison-Burch, C. (2009, August). Fast, cheap, and creative: evaluating translation quality using Amazon's Mechanical Turk. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1* (pp. 286–295). Association for Computational Linguistics.

6. Chevrier, S. (2003). Cross-cultural management in multinational project groups. *Journal of World Business*, *38*, 141–149.

7. Dabbish, L., Farzan, R., Kraut, R., & Postmes, T. (2012, February). Fresh faces in the crowd: turnover, identity, and commitment in online groups. In*Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (pp. 245–248). ACM.

8. Dabbish, L., Kraut, R., & Patton, J. (2012, May). Communication and commitment in an online game team. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems* (pp. 879–888). ACM.

9. Diamant, E.I., Fussell, S.R. & Lo, F.I. (2008). Where Did We Turn Wrong? Unpacking the Effects of Culture and Technology on Attributions of Team Performance. In *Proc. of CSCW 2008*, ACM Press, 383–391.

10. Doherty-Sneddon, G., Anderson, A.H., O'Malley, C., Langton, S., Garrod, S., & Bruce, V. (1997). Face-to-face and video mediated communication: A comparison of dialogue structure and task performance. *Journal of Experimental Psychology*, *3*, 105–125.

11. Dweck, C. (1999). Self theories: Their role in motivation, personality and development. Philadelphia, PA: Psychology Press.

12. Gao, G., Wang, H.C., Cosley, D., & Fussell, S.R. (2013). Same translation but different experience: The effects of highlighting on machine-translated conversations. In *Proc. of CHI 2013*, 449–458.

13. Hancock, J.T. & Dunham, P.J. (2001). Impression formation in computer-mediated communication revisited: An analysis of the breadth and intensity of impressions. *Communication Research*, *28*, 325–347.

14. Henderson, J. K. (2005). Language diversity in international management teams. *International Studies of Management and Organization*, *35*, 66–82.

15. Hinds, P. (1999). The cognitive and interpersonal costs of video. *Media Psychology*, *1*, 283–312.

16. Kelley, H.H. (1967). Attribution theory in social psychology. *Nebraska Symposium on Motivation*. University of Nebraska Press.

17. Lim, J. & Yang, Y.P. (2008). Exploring computer-based multilingual negotiation support for English-Chinese dyads: Can we negotiate in our native languages? *Behaviour and Information Technology*, *27*, 139–151.

18. Malu, M., Jethi, N., & Cosley, D. (2012, February). Encouraging personal storytelling by example. In *Proceedings of the 2012 iConference* (pp. 611–612). ACM.

19. Nass, C. & Lee, K.M. (2000). Does computer-generated speech manifest personality? An experimental test of similarity-attraction. *Proceedings of CHI 2000* (pp. 329–336). NY: ACM Press.

20. Nisbett, R.E. & Eugene, B. Attribution and the psychology of prediction. *Journal of Personality and Social Psychology*, *35*, 932–950.

21. Solomon, J. & Wash, R. (2012, February). Bootstrapping wikis: developing critical mass in a fledgling community by seeding content. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (pp. 261–264). ACM.

22. Storck, J. & Sproull, L. (1995). Through a glass darkly: What do people learn in videoconferences? *Human Communication Research*, *22*, 197–219.

23. Straus, S.G. & McGrath, J.E. (1994). Does the medium really matter? The interaction of task type and

technology on group performance and member reactions. *Journal of Applied Psychology*, *79*, 87–97.

24. Sukumaran, A., Vezich, S., McHugh, M., & Nass, C. (2011, May). Normative influences on thoughtful online participation. In *PART 5 of Proceedings of the 2011 annual conference on Human factors in computing systems* (pp. 3401–3410). ACM.

25. Tymoczko, M. (2007). Enlarging Translation, Empower Translators. Manchester: St. Jerome Press.

26. Wang, H.C., Cosley, D., & Fussell, S.R. (2010). Idea expander: supporting group brainstorming with conversationally triggered visual thinking stimuli. In *Proc. of CSCW 2010*, 103–106.

27. Wang, H.C., Fussell, S.R., & Cosley, D. (2013). Machine translation vs. common language: Effects on idea exchange in cross-lingual groups. In *Proc. of CSCW 2013*, 935–944.

28. Yamashita, N. & Ishida, T. (2006). Automatic prediction of misconceptions in multilingual computer-mediated communication. In *Proc. of IUI 2006*, 62–69.

29. Yamashita, N. & Ishida, T. (2006). Effects of machine translation on collaborative work. In *Proc. CSCW 2006*, 515–523.

30. Yamashita, N., Inaba, R., Kuzuoka, H., & Ishida, T. (2009). Difficulties in establishing common ground in multiparty groups using machine translation. In *Proc. of CHI 2009,* 679–688.

31. Yamashita, N., Kuzuoka, K., Hirata, K., Aoyagi, S., & Shirai, Y. (2011). Supporting fluid tabletop collaboration across distances. In *Proc. of CHI 2011*, ACM Press, 2827–2836