What Did You Say She Was Like? Features of Gossip Associated with Hearsay Accuracy and Consensus

Chenle Xu*, Psychology

ABSTRACT

When gossiping about a person, how does one's verbal responses affect the impression formed by the gossipers? A previous study on hearsay reputation find that when two gossipers gossiping about a target person, the hearsay accuracy about the target is low but the consensus between the two gossipers is high (Costello & Srivastava, 2017). As there has not been any research, to our knowledge, examining the factors that affect hearsay accuracy and consensus, this study intends to fill in the gap by exploring verbal responses as the potential factors. This study investigated the impacts of nine types of verbal responses on the accuracy and consensus. We coded 114 gossip conversations from the previous study (Costello & Srivastava, 2017) for nine specific responses and eight global features of the conversations. We explored the extent to which these specific and global features of gossip relate to consensus and accuracy in interpersonal impressions, operationalized as profile agreements. By using profile correlations with interpersonal ratings, the result suggests that certain types of responses may be associated with accuracy and consensus; for example, offering evaluation as a type of response seems to lead to a higher consensus between gossipers but cannot lead to an accurate impression of the target. Although few of the correlations were statistically significant, this project should still help to shed light on the conversational features associated with accuracy and consensus of impressions formed through gossip.

INTRODUCTION

Gossip is a very common activity in people's daily lives. Gossip may be associated with the intimacy or familiarity within groups or relationship (McDonald, Putallaz, Grimes, Kupersmidt, & Coie, 2007). Sharing information in gossip conversations is also important as it can help to stop the damages that may happen in social contexts. Dunbar (2004) explains that gossip may have evolved to help prevent people from exploiting the community by letting honest community members alert each other about any potential free rider. Moreover, through gossip, people can collect different kinds of information and form impressions about the target person (i.e. the one who gets discussed in the gossip episode). Forming an accurate impression of people is important for different occasions in life; for example, since an accurate personality

*Chenle Xu graduated from the University of Oregon in June of 2017. She majored in psychology and minored in theatre art. Chenle was a research assistant at the Personality and Social Dynamic lab led by Dr. Sanjay Srivastava while at the University of Oregon, where she also conducted her honors thesis. Chenle has also worked at Dr. Maureen Zalewski's The Science and Treatment of Affect Regulation Team lab as research assistant. Chenle is now pursuing a master's in psychology at Villanova University. Her plan after this master program is to apply for a PhD program in clinical psychology. Direct correspondence to: chenlex@uoregon.edu

judgment can help to predict the potential employee or student's future behaviors, a recommendation letter reader would need to gather accurate information from the letter to predict if the target person is worth hiring or admitting (Funder, 1995). The same issue may also exist in other social contexts, like in close relationships, where accurately perceiving personality can provide important information about a potential romantic partner (Buss, 2011). Therefore, gossip and the interpersonal judgments that accompany it are very important for people's daily lives.

COMPONENTS OF ACCURACY AND CONSENSUS

The importance of gossip or the interpersonal perception has intrigued researchers to ask what are the underline components behind this action? Among several different ways of conceptualizing agreement in interpersonal perception, one conceptualization stood out - *Profile Correlations*, which correlate ratings made across a set of items (Furr, 2009). In this conceptualization, Furr (2009) noted that the overall profile (i.e. raw personality profile) has two components: normativeness and distinctiveness. Normativeness is the average person's personality profile. Distinctiveness is what make each person different from the average person's personality profile. As different components reflect different valuable information, profile correlations would be the ideal way to look at interpersonal perception.

For example, say we have a dyad: the eleventh doctor and his companion Amy Pond, and we ask Amy to rate the doctor's personality on the Big Five Inventory. We would also ask the doctor to rate himself on the same questionnaire. Then we are left with two personality profiles for the doctor: one is the doctor's self-rating and the other is Amy's rating of the doctor. In order to know how accurate Amy (i.e. the informant) is at rating the doctor (i.e. the target), we can assess the overall profile accuracy by looking at the correlation between these two profiles. A high correlation between these two profiles would mean that Amy rated the doctor accurately. However, high overall profile accuracy does not tell us if Amy really knows the doctor well or if Amy just rated the doctor based on her impression of the average person. That is, we are not sure how much of the accuracy is driven by distinctiveness or normativeness.

As everyone is likely to be similar to average person (by definition), it is possible that a high level of profile accuracy in Amy's rating of the doctor could be due to using the impression of an average person. The correlation between Amy's rating about the doctor and the average person's personality profile is the normative accuracy. However, the doctor is somewhat different from the average person, and the differences between the doctor's personality characteristics and the average person's personality characteristics are the distinctive aspects of the doctor's personality. After meeting the doctor, Amy may be able to pick up some distinctive features in the doctor's personality, and reflect this information in her rating of the doctor. The correlation between the distinctive aspects of Amy's rating and the distinctive aspects of doctor's self-rating is the distinctive accuracy, which we can get at by subtracting the normative profile from Amy's and the doctor's ratings respectively.

Suppose then that Amy gossips about the doctor with another person, Rory Williams, giving Rory an impression of the doctor. Again, we ask the doctor to rate himself and Amy to rate the

doctor. Without meeting the doctor in person, Rory can also rate the doctor after hearing about him from Amy. We would then have three profiles about the doctor: one is doctor's self-rating, one is Amy's rating, and one is Rory's rating. In this case, Rory is the receiver as he only receives information about the target from the informant, Amy. The correlation between Rory's rating on the doctor and the doctor's self-rating in the triad is called the *Hearsay Profile Accuracy*. If more than one triad exists in a population pool, averaging all the potential target's self-ratings creates a new profile called the *normative profile*, the rating of the average person in this sample. Then when comparing Rory's rating on the doctor with the normative profile, the correlation between these two profile ratings is the *Hearsay Normative Accuracy*. However, the doctor's self-rating could be a little different from the normative profile, since everyone is a little different from the average person. This difference between the doctor's self-rating and the normative profile is the doctor's self-rated distinctive features, which is the doctor's *distinctive profile*. The difference between Rory's rating on the doctor and the averaged receivers' ratings on their targets is Rory's perception about the doctor's distinctive features. Thus, the correlation between the Rory's perception about the doctor's distinctive features and the doctor's distinctive profile is called the *Hearsay Distinctive Accuracy*.

Moreover, hearsay consensus between the gossipers (i.e. the informant and the receiver) in our example could be explained as follows: *Hearsay Profile Consensus* is the correlation between Amy's rating and Rory's rating. The distinctive features in doctor's personality could be captured in Amy's rating, so it is also likely that, after talking with Amy about the doctor, Rory can learn some distinctive features about the doctor's personality as well. As mentioned above, Rory's perception about the doctor's distinctive features is the difference between Rory's rating on the doctor and the average across all the receivers' ratings on their targets. Similarly, the differences between Amy's rating on the doctor and the averaged all informants' ratings on the target is Amy's perception about the doctor's distinctive features. Thus, the correlation between Amy's perception about the doctor's distinctive features and Rory's perception about the doctor's distinctive features is the *Hearsay Distinctive Consensus*. Additionally, *Hearsay Normative Consensus* is the correlation between Rory's rating and the average across all informants' ratings on their targets (see table below for all profile correlations' definitions). By separating out normative and distinctive accuracy, it allows for a more nuanced and interpretable analysis. This paper uses some terms interchangeably, such as overall profile accuracy and hearsay accuracy. These are summarized in the Table 1 below.

FACTORS THAT INFLUENCE ACCURACY AND POSSIBLE ACTS IN GOSSIP EPISODES

Normative and distinctive accuracy can be affected differently. Biesanz and Human (2010) investigated the effect of accuracy goals on interpersonal perception. Before having participants view a standard set of targets, they asked half of the participants to be as accurate as possible when rating the target's personality, and gave no specific instructions to the remaining participants. Biesanz and Human found that distinctive accuracy was higher for the participants they told to be more accurate, but normative accuracy was lower for those participants.

Table 1: Profile Correlations' Meaning/Definitions.

Profile Correlations	Meanings/Definitions
Hearsay Profile Consensus/ Overall Profile Consensus	The correlation between the informant's rating and the receiver's rating of the target
(Hearsay) Normative Consensus	The correlation between the receiver's rating and the average across all informants' ratings of the target
(Hearsay) Distinctive Consensus	The correlation between the informant's rating on target's distinctive features (i.e., the informant's rating compared with the average across all informants' rating) and the receiver's rating on target's distinctive features (i.e., the receiver's rating compared with the average across all receivers' ratings)
Hearsay Profile Accuracy/ Overall Profile Accuracy	The correlation between the receiver's rating and the target's self-rating
(Hearsay) Normative Accuracy	The correlation between the receiver's rating and the normative profile (i.e., average across all targets' self-ratings)
(Hearsay) Distinctive Accuracy	The correlation between the receiver's rating on target's distinctive features (i.e., the receiver's rating compared with the average across all receivers' ratings) and the distinctive profile (i.e., target's self-rating compared with the average across all targets' self-ratings)

Although this study provided some insight on what it may take to form an accurate distinctive accuracy, it did not dig deeper and capture what factors lead the participants to have more accurate judgment on the distinctive aspects of the target. The result may also be limited as it ignored triads and only studied interpersonal impression among the dyads (i.e. participants and the target from the video). Another study explored the structure of gossip in adolescents. Eder and Enke (1991) collected data from middle school students. They observed and recorded naturally occurring gossip conversation at one middle school. After collecting and analyzing all these audio and video files, Eder and Enke identified some basic structure of these gossip episodes. They found that after one middle school student identified a target at the beginning of a gossip episode, the others tended to respond with one of six possible "acts:"

- Request clarification of evaluation
- Explanation
- Support
- Expansion of evaluation
- Exaggerated affect
- Challenge

Each type of "act" could either restrain or support further discussion in the gossip conversation. As these responses had functions in gossip conversations and may elicit information that people use in their judgments, just like the objects in a person's room can serve as cues to help people forming an accurate impression of others (Gosling, Ko, Mannarelli, &

Morris, 2002), it is possible that certain types of verbal responses can be used as cues when forming accurate interpersonal impressions.

THE PRESENT STUDY

To our knowledge, there is no available research on factors in gossip episodes that affect the accuracy and consensus of gossipers' impressions. This study, as an exploratory project, looks for these factors using the data from a previous study on hearsay reputations (Costello & Srivastava, 2017). In their study, Costello and Srivastava had college undergraduate students come into the lab. One dyad was assigned to play the get-to-know-you game, where both people learned about each other. After the game, the researchers brought two new people to meet with one of original dyad member. The original dyad member (P1) told their partner (P2) about the person they played the game with (Target). All the participants then filled out the questionnaires to assess each other's personalities. The primary analysis focused on hearsay accuracy and consensus. Hearsay accuracy is the correlation between P2's personality rating on the target and the target's self-rating; hearsay consensus is the correlation between P1's personality rating on the target and the P2's rating on the target. They found that the hearsay accuracy is fairly low, but hearsay consensus is relatively high. However, as they did not look at the impression formation process or the potential reasons that lead to the high hearsay accuracy and low hearsay consensus, the current paper intends to use the possible structures of each gossip episode to find the potential cues that affect the hearsay accuracy and hearsay consensus. The specific exploratory question here is to see if the nine types of verbal responses, which were modified from Eder and Enker's study (1991) about the structures of gossip, have any impact on the hearsay accuracy and hearsay consensus of the target impression. The nine types of verbal response are:

- Offer explanation
- Request clarification
- Request evaluation
- Summarize
- Request expansion
- Offer evaluation
- Ask new questions
- Validation
- Inferences

METHOD

PARTICIPANTS

In Costello and Srivastava's original study (2017), there were 288 college undergraduate students who participated. In this study, pre-registration exclusion criteria led to the final 264 people remaining in the study. Participants' age range is from 18 to 30 years old (M= 19.59, SD=1.82). Most of the participants were female (57.7% female, 27.7% male, 14.6% no report).

The majority identified themselves as white (3.5% Black/African American, 17.7% Asian, 8.8% Hispanic, 0.4% Native American, 60.4% White, 3.8% other, 14.6% did not report race). Participants could choose more than one ethnicity. All students were from the same university that is located on the west coast of United States.

The current project had three coders to code the audiotapes. Two of coders were white female and one was Asian female. All coders were aged from 21 to 22 and were from the same university as all the participants who participated in the previous study.

MATERIALS/STIMULI

The previous study (Costello & Srivastava, 2017) produced audio data by recording the gossip conversation between P1 (i.e., the informant) and P2 (i.e., the receiver), and the only eligibility criterion for the recording was the subjects not knowing each other. The present study codes audio recordings.

There was a total of 134 audio recordings from this previous study. Of these, we excluded 20 recordings that had eligibility issues from being coded. There were two coders for each conversation recording, except for one that was coded by all three coders by mistake.

We randomized the orders of the total 114 audio recordings, and embedded them in a Qualtrics survey (see Appendix A). Coders used the survey to code the frequencies (scale range from zero to positive infinity) of each type of verbal response (i.e. the verbal response of P2).

MEASURES

BIG FIVE INVENTORY AND THE QUESTIONNAIRE BIG SIX

Costello and Srivastava's study (2017) used the modified Big Five Inventory-2 (BFI-2; Soto & John, 2016) and the Questionnaire Big Six (QB6) as the instruments to measure personality impressions. After the gossip conversation, participants filled out the questionnaires that contains questions from BFI-2 and QB6 to assess each other's personality on the six personality traits (openness, conscientiousness, extraversion, agreeableness, neuroticism, honesty and propriety). Among the six personality traits, honesty and propriety were measured with eight items (five of which appear in the QB6, three of which do not; Thalmayer et al., 2011). All the personality questions had a score range from one (disagree strongly) to five (agree strongly) with a neutral point of three (neutral; no opinion).

In the personality questionnaires, there were three types of ratings: the self-rating is when the target rates their own personality traits; the informant-rating is when P1 rates the target's personality; and the receiver-rating is when P2 rates the target's personality.

CODING

This study used nine types of verbal responses to code the frequency of P2 (i.e., the receiver)'s verbal response (M = 1.27, SD = 0.95) in their gossip conversation with P1 (i.e., the informant). The nine types of codes included: *offer explanation*, *request expansion*, *ask new questions*, *inferences*, *validation*, *request clarification*, *request evaluation*, *offer evaluation*, and *summarize*. Table 2 provides definitions and examples for each coding categories.

Table 2: Coding Categories.

Coding Category	a priori Definition	Example
Offer Explanation	P2 offers an explanation of something the target did or said that P1 mentions.	P2: "maybe [target] did it for…"
Request clarification	P2 directly asks P1 to clarify something about the target.	P2: "So you said [target] was unhappy about that? / So when you mentioned that [target] is… do you mean it is…?"
Request Expansion	P2 directly asks P1 to expand upon something said about target.	P2: "Can you say a little more about what [target] said about X?"
Ask New Questions	P2 asks a new question about the target	P2:" Do you know if [Target]…? / Did you guys talk about…?"
Inferences	P2 responds to P1 with an inference about what the target is like.	P2: "it sounds like [Target] is shy and happy"
Validation	P2 responds with other related comments (even side stories) to P1 showing that P2 had understood/heard what P1 said.	P2: "right, I understand that. I would totally do the same thing."
Request Evaluation	P2 asks P1 to offer an evaluation of the target (evaluative, in this sense, means valenced, or a positive or negative statement)	P2: "Is [Target] a nice person?" / "Is [Target] easy to talk to?"
Offer Evaluation	P2 offers an evaluation of the target (evaluation = valenced)	P2: "What the [Target] did is bad"
Summarize	P2 summarized what P1 has said about the target.	P2: " So you said [Target] is like…"

There were also eight questions about the global features of the conversation. These general questions were coded based on the coders' general impression of each gossip conversation. For example, the first question asked what percentage of the conversation was related to the target (range from 0 to 100), and General Question 1 asks if both people were interested and engaged in the conversation. All the general questions were rated on a Likert scale and ranged from one (strongly disagree) to five (strongly agree), with a neutral point of three (neither agree nor disagree). See all eight global features questions in Table 3.

Table 3: Global Features Questions.

Global Features Question	Question Content
First Question/ Percentage On-Topic	What percentage of the conversation was 'on topic' (i.e., about the target)?
General Question 1	Both people were interested and engaged in the conversation.
General Question 2	There was rapport between the two people in the conversation.
General Question 3	The content of the conversation about the target was mostly positive.
General Question 4	The content of the conversation about the target was mostly negative.
General Question 5	The conversation was detail-oriented, contained a lot of specific details about the target.
General Question 6	The conversation was abstract, contained a lot of generalities about the target.
General Question 7	The two people seemed to agree a great deal about the personality of the target.

The ICC score range from zero to one with one indicating a perfect agreement between coders and zero indicating no agreement between coders. In this study, the nine verbal behavior coding categories, the score ranges from 0.36 to 0.77; for the global featured questions, the ICC score ranges from 0.02 to 0.44. ICC and descriptive statistics is provided in Table 4.

PROCEDURE

Two of the coders were recruited by the lab as research assistants, and the third coder was the author of this paper. During the coders' training, all the coders received the coding orders for their specific coding team and the *Instructions for the Response Coding* (these materials can be obtained upon request from the author) to familiarize themselves with the coding categories. Due to the large amount of coding work, coders were divided into two coding team (Team A and Team B), and each team coded half of the audio recordings. Thus, each coding team had coded 57 recordings in 10 weeks. On average, each team coded five or six audio recordings every week.

Table 4: Interclass Correlation Coefficients and Descriptive Statistics

Code	Mean	SD	ICC
Offer Explanation	0.19	0.52	0.44
Request clarification	0.94	1.49	0.51
Request Evaluation	0.66	1.08	0.36
Summarize	0.98	1.65	0.54
Request Expansion	1.13	1.87	0.59
Offer Evaluation	0.57	0.98	0.40
Ask New Questions	2.98	3.47	0.77
Validation	2.88	3.74	0.54
Inferences	1.07	1.58	0.65
Mean of the Codes	1.27	0.95	

General Questions	Mean	SD	ICC
Percentage on Topic	70.58	27.115	0.77
General Question 1	3.64	0.99	0.26
General Question 2	3.68	1.07	0.24
General Question 3	3.87	0.81	0.44
General Question 4	1.79	0.78	0.22
General Question 5	2.8	1.01	0.22
General Question 6	3.07	0.80	0.02
General Question 7	3.71	0.68	0.35

RESULTS

AVERAGE PROFILE CORRELATION

The profile accuracy for each dyad is the correlation between the target's self-ratings with the receiver's ratings on the target's personality. Across groups, the average profile accuracy was $r=.31$. Profile consensus is the correlation between the informant's rating on the target with the receiver's rating on the target in each dyad. The average across all groups was $r=.55$.

Normative hearsay accuracy was obtained by correlating the receiver's rating with the mean of all the targets' self-ratings (across dyads). The average profile correlation for hearsay normative accuracy was $r=.42$. Normative consensus was obtained by correlating the receiver's rating with the mean of all informants' ratings. The average hearsay normative consensus was $r=.47$.

Hearsay distinctive accuracy is obtained by correlating the differences between the receiver's rating and the mean of all receivers' ratings with the differences between the target's self-rating and the mean of all targets' self-ratings. The average distinctive accuracy was $r=.10$. Distinctive consensus in each dyad is the correlation of the differences between the receiver's rating and the mean of all receivers' ratings with the differences between the informant's rating and the mean of all informants' ratings; the average hearsay distinctive consensus score was $r=.31$. Table 5 shows the average profile correlation scores.

Table 5: Average Profile Correlations.

Profile Correlation	Mean
Profile Consensus	0.55
Profile Accuracy	0.31
Normative Consensus	0.47
Normative Accuracy	0.42
Distinctive Consensus	0.31
Distinctive Accuracy	0.10

ANALYTIC APPROACH

As the current paper is an exploratory project, we wanted to see if the features of gossip episode (i.e. the verbal responses) have impacts on the hearsay accuracy and consensus by focusing on the effect size of the correlation rather than the test statistics and p values. Additionally, the data in this study have a nested structure (i.e. the triads are nested in groups), so the assumption of independence of observations does not apply and test statistics must take this into consideration.[1]

Thus, we focus on the effect sizes of these correlations. The two cutoff points we used here are the correlations with an absolute value above 0.1 as minimally of interest and correlations with an absolute value above 0.2 as worth considering for future confirmatory work.

PROFILE CORRELATION

To study the relations between these nine types of verbal responses and hearsay accuracy and consensus, we conducted a Pearson's correlation. The results are summarized in Table 6. For hearsay profile consensus, two codes are worth mentioning as they had a small to moderate positive correlation: General question 3 (positive gossip content), $r(101)= 0.20$, and General Question 4 (negative gossip content), $r(101)= -0.20$. Other codes like request evaluation, $r(101)= -0.13$, offer evaluation, $r(101)= 0.10$; General Question 2 (rapport between gossipers) $r(101)= 0.12$, and General Question 7 (seemly agreement between gossipers), $r(101)= 0.10$, all had small correlations with profile consensus.

For normative consensus, there were small correlations with request expansion, $r(104)= -0.18$, request clarification, $r(104)= -0.15$, request evaluation, $r(104)= -0.10$, the percentage on

topic, $r(104)$= -0.11, General Question 3 (positive gossip content), $r(104)$= 0.19, and General Question 4 (negative gossip content), $r(104)$= -0.14.

Distinctive consensus had some small correlations with several codes: offer explanation, $r(101)$= 0.12, request expansion, $r(101)$= 0.16, ask new questions, $r(101)$= 0.18, validation, $r(101)$= 0.18, the percentage on topic, $r(101)$= 0.10, General Question 3 (positive gossip content), $r(101)$= 0.11, General Question 4 (negative gossip content), $r(101)$= -0.15, and General Question 5 (detailed conversation), $r(101)$= 0.17. As an index of overall interactivity in the gossip conversation, we computed the average of all nine verbal response codes. This interactivity index (mean of the codes) had a small correlation with distinctive consensus, $r(101)$= 0.18.

Moreover, hearsay profile accuracy had small to moderate correlations with two of the nine codes: mean of codes, $r(100)$= -0.20, and General Question 6 (abstract gossip conversation), $r(100)$= 0.25. Codes that had small correlations with profile accuracy are: request clarification, r(100)= -0.16, request expansion, r(100)= -0.15, ask new questions, r(100)= -0.18, validation, r(100)= -0.12, inferences, r(100)= -0.14, the percentage of conversation is on topic, r(100)= -0.15, General Question 3 (positive gossip content), r(100)= 0.11, General Question 4 (negative gossip content), r(100)= -0.10, and the General Question 5 (detailed gossip conversation), r(100)= -0.17.

Normative accuracy also had a small to moderate correlation with request expansion, $r(104)$= -0.26. There were also small correlations with the following codes: offer explanation, $r(104)$= -0.10, request clarification, $r(104)$= -0.17, request evaluation, $r(104)$= -0.12, ask new questions, $r(104)$= -0.15, the mean of codes, $r(104)$= -0.12, the percentage of conversation on topic, $r(104)$= -0.16, General Question 2 (rapport between gossipers), r(104)= 0.16, General Question 3 (positive gossip content), r(104)= 0.16, General Question 4 (negative gossip content), r(104)= -0.11, and General Question 6 (abstract gossip conversation), r(104)= 0.12.

Finally, distinctive accuracy had small correlations with validation $r(100)$= -0.13, inferences, $r(100)$= -0.13, mean of the code, $r(100)$= -0.10, and General Question 6 (abstract gossip conversation), $r(100)$= 0.14.

Table 6: Profile Correlation Effect Size & P Values Matrix

Codes	Hearsay Profile Consensus	Hearsay Profile Accuracy	Hearsay Normative Consensus	Hearsay Normative Accuracy	Hearsay Distinctive Consensus	Hearsay Distinctive Accuracy
Offer Explanation	.04	-.04	-.06	**-.10**	**.12**	-.07
Request Clarification	-.02	**-.16**	**-.15**	**-.17**	.08	.03
Request Evaluation	**-.13**	-.08	**-.10**	**-.12**	.01	-.01
Summarize	-.05	-.03	-.02	.04	-.01	-.04
Request Expansion	.00	**-.15**	**-.18**	<u>**-.26** *</u>	**.16**	-.02
Offer Evaluation	**.10**	-.04	.07	.04	-.07	-.00
Ask New Questions	.06	**-.18**	-.08	**-.15**	**.18**	-.04
Validation	.08	**-.12**	.02	.02	**.18**	**-.13**
Inferences	.05	**-.14**	.06	.05	-.03	**-.13**
Mean of Codes	.05	<u>**-.20** *</u>	-.08	**-.12**	**.18**	**-.10**
Percentage on Topic	-.02	**-.15**	**-.11**	**-.16**	**.10**	.06
General Question 1	.01	-.06	-.01	.04	.02	.01
General Question 2	**.12**	.02	.09	**.16**	.04	-.03
General Question 3	<u>**.20** *</u>	**.11**	**.19**	**.16**	**.11**	.03
General Question 4	<u>**-.20** *</u>	-.10	**-.14**	-.11	**-.15**	-.04
General Question 5	.08	**-.17**	.03	-.03	**.17**	-.07
General Question 6	.02	<u>**.25** *</u>	.08	**.12**	-.05	**.14**
General Question 7	**.10**	.07	.03	.05	.06	.09

DISCUSSION

This study explores the impact of the nine types of verbal responses on hearsay accuracy and the consensus between two gossipers. Focusing on the effect size of the correlation between different codes and the components of accuracy and consensus, there are some findings that indicate certain verbal responses may be potential factors that have impacts on accuracy and consensus. The types of verbal response that could be worthwhile to examine further for the hearsay accuracy and consensus are the ones that had a small to moderate effect size.

Taking the request expansion as an example, the negatively small to moderate correlation with normative accuracy suggest, in our example, the more Rory asks Amy to expand on what she said about the doctor, the less likely Rory would think the doctor is like the average person in the population. This may be due to Rory learning more about the doctor by asking Amy to expand more. He may then know more—or at least, thinks he knows more—unique information about the doctor, which would make him to be less likely to rate the doctor based on what an average person is like.

FACTORS ASSOCIATED WITH CONSENSUS

OVERALL PROFILE CONSENSUS

Both positive and negative gossip content seems to have impact on the hearsay profile consensus. The correlation between positive gossip content and the overall profile consensus suggests that the more the two people talk about the target positively the more they both agree with each other. This could be due to the positive impression that the informant formed for the target, as the face-to-face interaction can lead to more positive impressions through providing more interpersonal cues (Okdie, Guadagno, Bernieri, Geers, & Mclarney-Vesotski, 2011). If the target was discussed in a positive way, then both the receiver and informant would be more likely to agree with each other as they know the informant had met the target for the first time. A negative correlation between the negative gossip content and the profile consensus would then not be surprising. It indicates that the receiver has lower agreement with the informant on the ratings of the target when they talk about the target negatively. This may be because that the disconfirmation of the receiver's expectation leads them to disagree with the informant on the impression of the target. People usually form expectation from their past experiences before the social interaction with someone (Snyder & Stukas, 1999). The receiver, in this case, may expect that the informant to have a positive impression of the target, since people tend to use positive terms rather than negative ones to describe others (Zimmermann et al. 2017). However, when the informant starts to discuss the target in a negative manner, which contradict with the receiver's expectation, the receiver is more likely to still retain their expectation (Snyder & Stukas, 1999) and keep the positive impression of the target. It would then lead to a discrepancy between the receiver and the informant's impressions of the target when doing the rating.

What's more, some of the other conversation features also seem to affect the agreement between two gossipers. Request evaluation negatively affects profile consensus, which suggests

that the more the receiver asks about the informant to give evaluative information about the target, the less likely that the informant and the receiver agree with each other. It could be that when the informant does not seem to be comfortable enough to give out more evaluative information spontaneously, and the receiver must constantly ask the informant to evaluate the target, the receiver would start to question how well the informant knows the target. This could then result in a difference in rating. However, offering evaluation seems to promote the consensus between the two gossipers. We found that it is common for some informants to notify the receiver that they did not get enough information about the target. Perhaps when the informant has limited information about the target, they would be easily influenced by other sources of information, even if these were merely new ideas from the receiver. Thus, the more evaluation that the receiver gives about the target, the more likely the informant would follow the receiver's thoughts and agree with the receiver's interpretation about the target. Moreover, having rapport and agreement also help to increase the agreement between the two gossipers. This could be because the more rapport has been established between two people, the more likely they agree with each other. Assuming rapport is associated with interpersonal coordination (Miles, Nind, & Macrae, 2009), the informant and receiver could have been acting as if they were in a coordinative situation where they had to work together in forming their impressions of the target. Also, when the two people appear to agree a lot about the target's personality, they would be more likely to have the similar opinions about the target when doing the personality ratings.

NORMATIVE CONSENSUS

There are several conversation features that show impacts on the normative consensus. The more the receiver asks the informant to make clarifications or expansions on what the informant had said about the target, the more likely the receiver tends to disagree with all the informants' perceptions about their targets. When the receiver asks for detailed information from their informant, the receiver would form an impression specifically about his/her own target. This impression could then be different from what all the informants think their targets are like on average. High percentage of the conversation about the target is another illustration of the high likelihood for the receiver spend more time in asking for more information about the target, and this eventually result in seeing their specific target to be less like the average of all the informants think their targets are like. The same idea may apply when two gossipers talk more about the target in their conversation through expanding information about the target. The receiver may know their specific target too well to perfectly match their rating on the target with all informants' impression of their targets.

Additionally, General Question 4 suggests that when the two gossipers talk about the target in a negative way, the receiver is less likely to agree with the impression that all informants had for their targets. Just as the rationale discussed above, the receiver may expect the informant to describe the target positively (Zimmermann et al., 2017). But, when the informant talks negatively about the target, which conflicts with the receiver's expectation, the receiver may then decide to ignore the informant's judgment and believe their own conceptualization of the target

(Snyder & Stukas, 1999). This could make the receiver's impression of the target to diverge from all informants' impression of their targets. However, the correlation between General Question 3 and the normative consensus also indicates that, when talking about the target positively, the receiver would be more likely to agree with what all informants think their targets are like. This may also make sense in that, as the average person usually has some socially desirable traits, talking positively about their specific target may lead both the informant and the receiver to think that the target is very similar to the average person. By agreeing with their own informant's impression of the target and thinking the target is like the average person, the receiver's idea of what the target is like could be similar to all informants' idea about what their targets are like. Thus, the receiver's rating is more likely to match with all informants' impression of their targets.

DISTINCTIVE CONSENSUS

Two gossipers seem to agree more on the distinctive characteristics of the target with the influences of certain conversation features. When the receivers offer more explanation about the target, request more expansion on the target's behavior, validate what the informant says about the target, ask more questions about the target, or simply spend more time in discussing the target, both the informant and the receiver seem to agree more with each other on their specific impression of the target. By making sense about what the informant says about the target, learning specific details about the target, validating on the informant's explanation about the target, or spend more time talking about the target, the receiver can also form a more precise impression of the target after motivating the informants to provide more detail information about the target. This way, both the receiver and the informant would share the same precise impression of the target, which may lead them to have similar opinion in rating target. Also, asking new questions, especially follow-up questions, may help people to be more likable by their conversation partners (Huang, Yeomans, Brooks, Minson & Gino, 2017). Hence, when the receiver asks any type of questions, the agreement between the two gossipers would increase as the informant may like the receiver better.

Moreover, the positive correlation between the mean of the nine responses and the distinctive consensus suggests that all nine verbal responses seem to be helpful for the receiver to agree more with the informant on the target's distinctive features. When the receiver utilizes the nine types of responses more often, the informant could have higher chance to tell the receivers more about the unique features of the target. This could then result in an increase in agreement between the receiver and the informant on rating the target's distinctive features. The results for General Question 3 and 5 also show that, when talking about the target positively or in great detail, the two gossipers are more likely to have consensus on their rating of the target's distinctive features. It is possible that talking about the target in a more positive way can lead the receiver to pay more attention on any distinctive features of the target. Since people tend to describe others in a positive manner (Zimmermann et al., 2017), the receiver may feel the need to remember any other distinctive features that make the target different since it would help him/her to recall the impression easier when doing rating. As the result of a smooth

discussion on the target's uniqueness, a higher agreement between the receiver and the informant on rating the target's unique features would then become possible.

Also, similar idea may apply when talking about the target in great detail. To the receiver, the distinctive characteristics of the target may "pop-out" more, as it can help him/her to remember the target better. After receiving the unique information about the target, the receiver would then have a higher chance to agree with the informant on rating the target's distinctive features. In contrast, General Question 4 suggests that when talking the target in a more negative way, the receiver is less likely to agree with the informant on the rating of the target's uniqueness. Similar idea that has discussed in normative consensus section may help to explain this relation. With the expectation that the informant would form a positive impression of the target and discuss the target positively, the receiver may choose to retain his/her own expectation even when the informant disconfirmed the expectation by discussing the target negatively (Snyder & Stukas, 1999). The receiver would be less likely to agree with the informant's impression of the target, let alone to agree with the informant on rating the target's unique characteristics.

FACTORS ASSOCIATED WITH ACCURACY

OVERALL PROFILE ACCURACY

The average of the nine verbal responses had a negative but small to moderate correlation with profile accuracy, which suggests that the more the receiver use all nine types of responses, the less accurate s/he is when rating the target's personality. Since half of the verbal codes can help the receiver learn more information about the target (e.g. request clarification, request expansion etc.) and more than half of the verbal responses happen when the receiver is trying to interpret the information that s/he learn from the informant (e.g. offer explanation, summarize etc.), it may actually hurt the receiver's understanding about the target when using all nine types of responses. The receiver could over-interpret the information about the target. Hence, this means that the receiver would be less accurate at rating the target. Another possible explanation for this correlation could be that, although utilizing all the responses may show the high motivation of the receiver, which could lead to a higher distinctive accuracy (Biesanz & Human, 2010), the overall profile accuracy is still not increased as it has more than just this one component (i.e., it also contains the normative component). Nonetheless, these verbal responses that had a small to moderate effect should provide some implication for future researchers in studying hearsay accuracy and consensus.

Some other conversation features can also have influence on the hearsay accuracy. General Question 6 has a positive correlation with all three components of accuracy, so it indicates that the more two gossipers talk about the target in an abstract way, the better the receiver is to form an accurate impression about the target. This can be easily understood, as knowing more generalities about the target may help the receiver to have a general impression of the target. With this general impression that provides more comprehensive information about the target, the receiver could achieve a higher overall profile accuracy. If there is any missing information

about the target on certain personality traits, the receiver may be more likely to make a judgment based on the average person's profile, which would then increase their normative accuracy. On the other hand, talking about the target in generalities do not mean that the informant fails to deliver any unique information about the target, as the target could be simply like the average person who has less special characteristics. Thus, receiver should still be able to score high on distinctive accuracy as the target only has few distinctive features and acts very much like the average person.

Furthermore, the more the receiver ask for clarification, expansion, new question, or spending more time in discussing the target, the less likely they would be accurate with their rating of the target. This may be due to the high likelihood of making assumptions when discussing the target. As more information or longer exposure to the information could lead to a more accurate judgement (Wall, Taylor, Dixon, Conchie, & Ellis, 2013), the informant may not have a fully accurate impression of the target after meeting the target for only 10 minutes. When the receiver utilizes verbal responses like ask for clarification, request expansion, or asks new question to acquire more information that is not known by the informant, there is a higher chance for both the informant and the receiver to start making assumptions. These assumptions are not facts about the target. Thus, it could result in the receiver forming a less factually accurate impression of the target. In addition, it seems that the more the receiver use validation or offer inference, the less accurate s/he is at rating the target. This may be because, as we found from the recordings, when the receiver is not fully interested in the discussion about the target, s/he tends to validate the informant by giving out simple verbal response, like "yes", "OK", or "that makes sense." If the interest in the conversation is low, then it is likely that the receiver would be less attentive to information about the target such that, any accurate information about the target would not be received by the receiver. But the interest about the conversation could also be high to the extent that the receiver is actively making inferences. The receiver then runs the risk of making inaccurate assumptions as the informant has supplied insufficient information, leading to a low accuracy in rating.

Likewise, the profile accuracy was lower when talking more about the target in the gossip conversation. Although having more information about the target can lead to more accurate judgment of the target (Letzring Wells & Funder, 2006), this study found that talking more about target with the receiver could produce a risk for the two gossipers to make distant assumptions when the informant only has limited information about the target. Thus, the receiver's rating would have a low accuracy. What's more, when talking about the target generally or positively, the receiver seems to be more accurate with their rating. Since talking about the target in generalities can help the receiver form a more general and comprehensive impression of the target, the receiver may be more likely to accurately rate the target. Receivers also have high accuracy on rating target after gossiping about the target positively since positive attitudes towards unacquainted others let people minimize cognitive effort and are largely correct when judging others' personalities (Zimmermann et al., 2017).

Contrastingly, we found that talking about the target negatively or in detail leads to a lower profile accuracy. This could be that talking negatively about the target contradict with the

receiver's expectation, so the receiver decides to retain his/her own expectation (i.e., a positive impression of the target; Snyder & Stukas, 1999). By neglecting some accurate information from the informant about the target and relying on the expectation, the receiver is less likely to rate the target correctly. Regarding the low accuracy when talking the target in detail, using limited information to discuss the target in detail may indicate a higher likelihood of guessing or assumption making, like in the reasoning above. The receiver would then have a low accuracy in rating the target.

NORMATIVE ACCURACY

Hearsay normative accuracy appears to be affected by some features of the gossip conversation. The more the receiver explain the target's characteristics, asks for clarification, evaluation, new questions, or spend more time in talking about the target, the less accurate the receiver's rating is when matching the personality profile of target with the average person. This may be because the target is somewhat different from the average person and the more the receiver knows about their specific target and tries to make sense of target's personality, the more likely the receiver's rating would match with this specific target's self-rating and be less accurate with the average person's profile. This could also help to explain why using all nine verbal responses and talking more about the target would lead the receiver to think the target is less like the average person. Utilizing the nine verbal responses and talking more about the target can bring more information about their specific target to the receiver, which leads the ratings to diverge from the average person's profile. Another explanation could be that using more of these verbal responses shows a higher accuracy motivation of the receiver, which would then lower the receiver's normative accuracy (Biesanz & Human, 2010).

General Question 4's result suggests that talking about the target negatively can lead the receiver to have lower normative accuracy. This may be because perceiving the target to have less socially desirable traits could make the receiver to believe the target is different from the average person, who tends to have socially desirable traits. A discrepancy between the receiver's rating and the profile of the average person would then emerge. Nevertheless, the results about General Question 2 and 3 indicate they each can help to boost the normative accuracy. Having a harmonious conversation (i.e., building rapport) with the receiver, the informant may feel more comfortable to admit or reveal any limitation in their impression of the target. By understanding these limitations, the receiver would then, rather than rely on the informant's information entirely, be more likely to rate the target based on what the average person is like. In addition, talking about the target in a positive manner may allow the receiver to see the similarity between the target and the average person, as they both would have socially desirable personality traits.

DISTINCTIVE ACCURACY

Validation, inferences and utilize all nine verbal responses seem to be related with the decrease in hearsay distinctive accuracy. This implies that when the receiver is validating what the informant says, making inference about the informant's opinions, or using all verbal responses, the receiver is less accurate at rating the target's distinctive features. These findings

should not be surprising, since the receivers, like what we have discussed in profile accuracy section, usually give out a simple validating response when the informants describe the common or ordinary features of the target. Those features may not trigger any special impression of the target. Because of this, the receiver may feel the need to utilize more verbal responses in hoping to gain more unique information. Thus, doing more validation or using all nine types of verbal responses could simply suggest the informant fails to provide any unique or interesting information about the target. The receiver, therefore, would not know much more special features of the target, let alone to reflect this information on the rating of the target. However, General Question 6 suggests that talking about the target abstractly is related to increase in distinctive accuracy. When the target is very much like the average person without much unique features, then talking generalities about him could still lead to a high distinctive accuracy. But if the target is truly different than the average person, a general conversation about him should still carry out his unique characteristics. Thus, the receiver would still be likely in accurately rating target's distinctive features. Also, just as people tend to better accept information that they feel connected with (Niemiec, & Ryan, 2009), making more inferences may help the receiver to better remember the assumptions that s/he comes up with. Only remembering these assumptions but not other distinctive information about the target may lead the receiver to have a lower accuracy on rating the target's distinctive features.

LIMITATIONS

The current study is an exploratory project without much empirical evidence, and some conclusions may need to be interpreted with caution as the effect size for each coding categories is small. But because so little is known about the relationship between the features of gossip and the hearsay accuracy and consensus, this study helps shed light on this topic. For example, teachers have linguistic patterns, like the utility of evaluative language and the communal "we", in their professorial talk (Neal, 2008). In the context of gossip, it's likely that people would have different type of verbal responses that may lead to various result in forming accurate impression of others. With profile correlations, a more systematic method in studying hearsay accuracy and consensus, the current study looked at this link and was able to provide some complementary implications to better understand the feature of gossip.

Another limitation of this project could be the lack of enough coders. Since there were only three coders in two coding teams (one coder working in both coding teams), it is likely that the data would be weighted more towards the way that one coder who worked in both teams was doing coding. The use of the ICC to measure the reliability between coders showed that the coders did reliable coding work in general with the ICC for verbal responses range from 0.36 to 0.77.

IMPLICATIONS

Since there are not many studies examining the power of response in gossip, our current project offers complementary information on studying gossip through the angle of gossip's features. Just as the study on the room with cues (Gosling et al., 2002), which suggests that

observers can use some valid cues to form an accurate impression of occupants, the current paper can help to offer a new way of thinking gossip accuracy and consensus by looking for certain verbal cues in the conversation. In daily life, although people do gossip quite often, they do not realize that there may be some patterns in their ways of reacting during these gossip conversations. This project could help people to understand how the way that they respond in gossip may lead them to have different results on forming accurate impressions of others. It may also be applied to the field of counseling psychology, especially in therapeutic contexts: when the clients talk about other people in their lives with the counselor, the counselor could utilize certain types of verbal responses to form a more accurate impression of that target person. This way, the counselor may be able to know if the client is being objective or subjective.

FUTURE DIRECTIONS

Future research could consider specific types of verbal response and their correlations with hearsay accuracy and consensus, as it would offer more precise insight into the mechanism on how different verbal responses affect gossip accuracy and consensus. Specifically, any correlations that is above .1 in the current study would be meaningful to explore further. For example, the correlation between *request expansion* and *normative accuracy* is deserving of further study. When people ask for more information about a person, it can lead to lower accuracy by creating the perception that target is less like the average person. Maybe knowing more about someone would make this someone more distinguishable (i.e., less like the average person), or maybe there is other third variable accounts for this correlation. Exploring this relationship further could raise more directions for further study.

Future studies could also investigate the correlation between the types of verbal responses in naturally-occurring gossip and the accuracy of interpersonal perception, as it would have more validity in terms of the effects of gossip features on impression accuracy. As it is possible that the current project missed some cues in gossip conversations, such as facial expressions or body gestures, future studies should use video recordings to detect more cues that are worth noticing in gossip episodes. This would help further understanding of the features of gossip better by including more detailed information on cues or other relevant factors.

ACKNOWLEDGEMENTS

REFERENCES

Biesanz, J. C., & Human, L. J. (2010). The Cost of Forming More Accurate Impressions: Accuracy-Motivated Perceivers See the Personality of Others More Distinctively but Less Normatively Than Perceivers Without an Explicit Goal. *Psychological Science*, 21(4), 589-594.
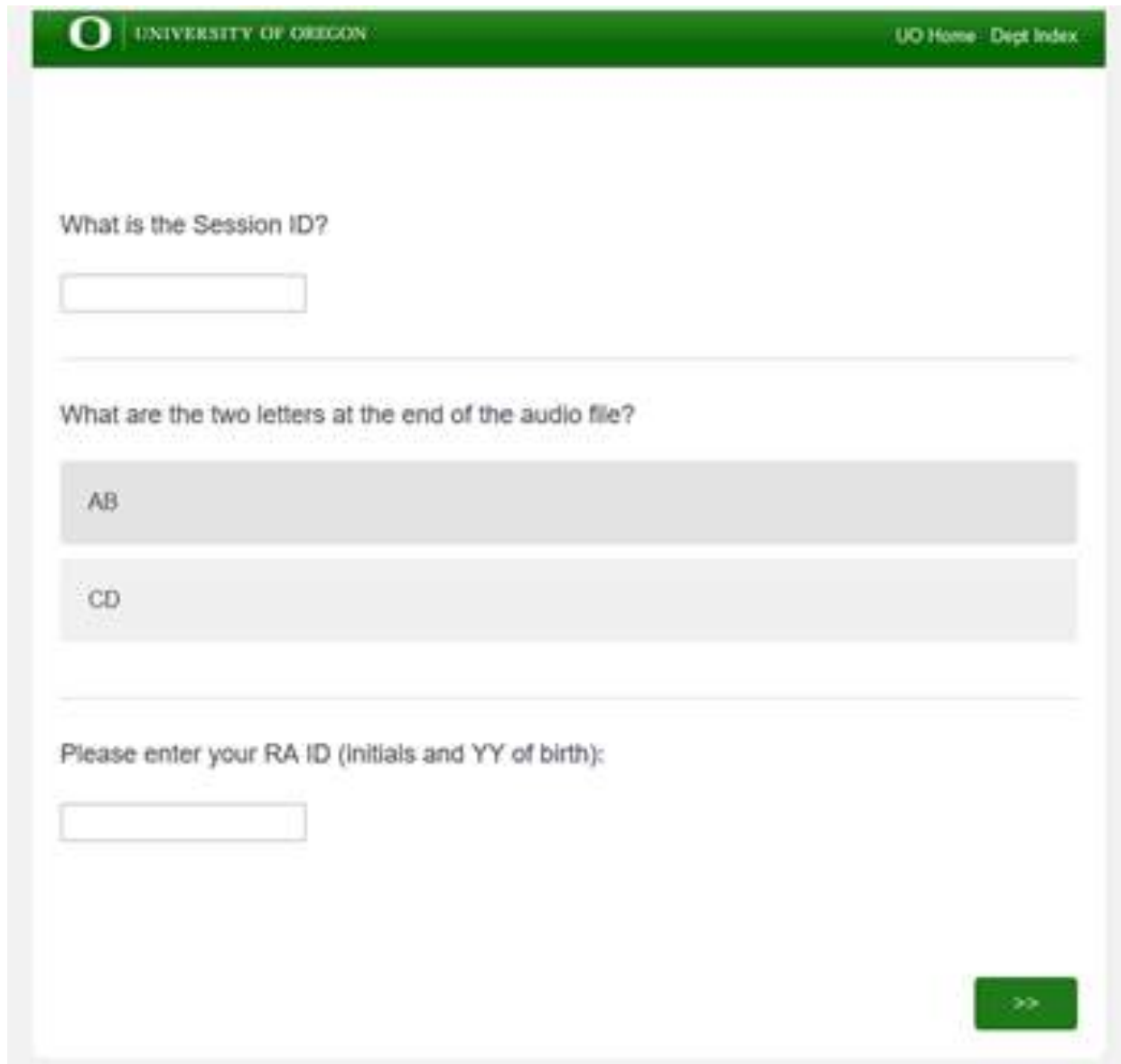
Buss, D. M. (2011). Personality and the Adaptive Landscape: The Role of Individual Differences in Creating and Solving Social Adaptive Problems. D.M. Buss & P.H. Hawley (Eds.), *The Evolution of Personality and Individual Differences* (29-57). New York: Oxford University Press.

Costello, C. K. & Sarivastava, S. (2017). *Perceiving through the Grapevine: Consensus and Accuracy of Hearsay Reputations*. Poster section presented at the 18th annual meeting of the Society for Personality and Social Psychology (San Antonio, TX).

Dunbar, R. I. (2004). Gossip in Evolutionary Perspective. *Review of General Psychology,* 8(2), 100-110. doi:10.1037/1089-2680.8.2.100

Eder, D., & Enke, J. L. (1991). The Structure of Gossip: Opportunities and Constraints on Collective Expression among Adolescents. *American Sociological Review*, 56(4), 494-508.

Furr, R. M. (2008). A Framework for Profile Similarity: Integrating Similarity, Normativeness, and Distinctiveness. *Journal of Personality*, 76(5), 1267-1316.

Funder, D. C. (1995). On the Accuracy of Personality Judgment: A Realistic Approach. *Psychological Review*, 102(4), 652-670.

Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A Room with a Cue: Personality Judgments Based on Offices and Bedrooms. *Journal of Personality and Social Psychology*, *82*(3), 379-398.

Huang, K., Yeomans, M., Brooks, A. W., Minson, J., & Gino, F. (2017). It Doesn't Hurt to Ask: Question-Asking Increases Liking. *Journal of Personality and Social Psychology*, 113(3), 430-452.

Letzring, T. D., Wells, S. M., & Funder, D. C. (2006). Information Quantity and Quality Affect the Realistic Accuracy of Personality Judgment. *Journal of Personality and Social Psychology*, 91(1), 111-123.

McDonald, K. L., Putallaz, M., Grimes, C. L., Kupersmidt, J. B., & Coie, J. D. (2007). Girl Talk: Gossip, Friendship, and Sociometric Status. *Merrill-Palmer Quarterly*, 381-411.

Miles, L. K., Nind, L. K., & Macrae, C. N. (2009). The Rhythm of Rapport: Interpersonal Synchrony and Social Perception. *Journal of experimental social psychology*, *45*(3), 585-589.

Neal, M. (2008). Look Who's Talking: Discourse Analysis, Discussion, and Initiation-Response-Evaluation Patterns in the College Classroom. *Teaching English in the Two Year College*, 35(3), 272-281.

Niemiec, C. P., & Ryan, R. M. (2009). Autonomy, Competence, and Relatedness in the Classroom: Applying Self-Determination Theory to Educational Practice. *School Field*, 7(2), 133-144.

Okdie, B. M., Guadagno, R. E., Bernieri, F. J., Geers, A. L., & Mclarney-Vesotski, A. R. (2011). Getting to Know You: Face-to-Face versus Online Interactions. *Computers in Human Behavior*, 27(1), 153-159.

Snyder, M., & Stukas Jr, A. A. (1999). Interpersonal Processes: The Interplay of Cognitive, Motivational, and Behavioral Activities in Social Interaction. *Annual Review of Psychology*, 50(1), 273-303.

Thalmayer, A. G., Saucier, G., & Eigenhuis, A. (2011). Comparative Validity of Brief to Medium-Length Big Five and Big Six Personality Questionnaires. *Psychological assessment*, 23(4), 995-1009.

Wall, H. J., Taylor, P. J., Dixon, J., Conchie, S. M., & Ellis, D. A. (2013). Rich Contexts Do Not Always Enrich the Accuracy of Personality Judgments. *Journal of Experimental Social Psychology*, *49*(6), 1190-1195.

Zimmermann, J., Schindler, S., Klaus, G., & Leising, D. (2017). The Effect of Dislike on Accuracy and Bias in Person Perception. *Social Psychological and Personality Science*, 9(1), 80-88.

NOTES

[1] For results that accounts for the nested structure, please contact author to obtain the update. Current result is very similar to the updated result.

APPENDIX A

▶ ● ————— 0:00 / 8:43 🔊 ————●   Session ID: 06031601-AB

Based on the recording, how many times did P2 give out the following responses? (Indicate the number of times in the box.)

                                                              Frequency

Offer Explaination   [   ]

Request Clarificaiton   [   ]

Request Evaluation   [   ]

Summarize   [   ]

Request Expansion   [   ]

Offer Evaluation   [   ]

Ask New Questions   [   ]

Validation   [   ]

Inferences   [   ]

What percentage of the conversation was 'on topic' (i.e., about the target)?

What percentage of the conversation was 'on topic' (i.e., about the target)?

0   10   20   30   40   50   60   70   80   90   100

Pick a number in percentage

Please indicate the extent to which you agree and disagree with the following statements.

Both people were interested and engaged in the conversation.

| Strongly disagree | Somewhat disagree | Neither agree nor disagree | Somewhat agree | Strongly agree |

There was rapport between the two people in the conversation.

| Strongly disagree | Somewhat disagree | Neither agree nor disagree | Somewhat agree | Strongly agree |

The content of the conversation about the target was mostly positive.

| Strongly disagree | Somewhat disagree | Neither agree nor disagree | Somewhat agree | Strongly agree |

The content of the conversation about the target was mostly negative.

Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree

The conversation was detail-oriented, contained a lot of specific details about the target.

Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree

The conversation was abstract, contained a lot of generalities about the target.

Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree

The two people seemed to agree a great deal about the personality of the target.

Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree